applied sciences

MDPI

Article

Real-Time Moving Ship Detection from Low-Resolution Large-Scale Remote Sensing Image Sequence

Jiyang Yu [1], Dan Huang [2,3,*], Xiaolong Shi [2], Wenjie Li [2] and Xianjie Wang [2]

[1] Beijing Institute of Spacecraft System Engineering, China Academy of Space Technology, Beijing 100094, China
[2] College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China
[3] System Design Department, China Research and Development Academy of Machinery Equipment, Beijing 100089, China
* Correspondence: danh314@cqut.edu.cn

Abstract: Optical remote sensing ship target detection has become an essential means of ocean supervision, coastal defense, and frontier defense. Accurate, effective, fast, and real-time remote sensing data processing is the critical technology in this field. This paper proposes a real-time detection algorithm for moving targets in low-resolution wide-area remote sensing images, which includes four steps: pre-screening, simplified HOG feature identification, sequence correlation identification, and facilitated Yolo identification. It can effectively detect and track targets in low-resolution sequence data. Firstly, iterative morphological processing was used to improve the contrast of low-resolution ship target profile edge features compared with the sea surface background. Next, the target area after adaptive segmentation was used to eliminate false alarms. As a result, the invalid background information of extensive comprehensive data was quickly eliminated. Then, support vector machine classification of S-HOG feature was carried out for suspected targets, and interference such as islands and reefs, broken clouds, and waves were eliminated according to the shape characteristics of ship targets. The method of multi-frame data association and searching for adjacent target information between frames was adopted to eliminate the interference of static targets and broken clouds with similar contours. Finally, the sequential marks were further trained and learned, and further false alarm elimination was completed based on the clipped Yolo network. Compared with the traditional Yolo Tiny V2/V3 series network, this method had higher computational speed and better detection performance. The F1 number of detection results was increased by 3%, and the calculation time was reduced by 66%.

Keywords: optical satellite image; ship detection; convolutional neural networks; deep learning

Citation: Yu, J.; Huang, D.; Shi, X.; Li, W.; Wang, X. Real-Time Moving Ship Detection from Low-Resolution Large-Scale Remote Sensing Image Sequence. *Appl. Sci.* **2023**, *13*, 2584. https://doi.org/10.3390/app13042584

Academic Editor: Atsushi Mase

Received: 21 September 2022
Revised: 5 February 2023
Accepted: 15 February 2023
Published: 16 February 2023

1. Introduction

Optical remote sensing image data processing technology, represented by ship target detection and recognition at sea, has become the core key to the current intelligent application of space-earth integration and has been widely used in fishery management, maritime rescue, and regional monitoring [1]. For a long time, the real-time detection of ocean-going ship targets has mainly relied on high-resolution remote sensing satellites. Still, problems include too many mission requirements, significant differences in imaging position and time, and complex satellite mission planning [2]. In addition, the imaging process is interfered with by night and day, clouds, morning and evening light and dark, and some real-time observation requirements are challenging to meet [3].

Several civil remote-sensing small satellites have emerged in the past five years, effectively filling the gap in imaging application tasks [4]. Compared with the large platform high-resolution series satellites, the imaging resolution of small satellites is low, the average definition could be higher, and the platform stability performance could be better, which leads to the low quality of the original image products produced [5]. However,

considering the time–space coverage and application requirements in emergencies, remote sensing images of low-resolution wide-area can effectively display the surrounding state of the target, and sequence images can show the motion information of the target itself, which has attracted wide attention in real-time application scenarios [6].

In a meter-level/10-m-level resolution image, the target may only account for about 10 pixels in the picture. Small and medium-sized ships present a wedge-shaped Gaussian gray distribution, and the details of the target texture disappear, and only the exterior outline of the hull is retained. Therefore, it is difficult to confirm whether the target is a ship from the target's details and surrounding context [7]. The ship targets in the image are tiny and densely distributed, characterized by small size, large number, diverse locations, and extensive background interference, which are pretty different from the large and prominent detection targets in the public data set [8].

In the low-resolution scenario, the visual sea state tends to be flat, and the trawl traces of ships are difficult to identify. However, the similar hull shape formed by broken clouds significantly reduces false alarms [9]. The image sequence can eliminate false alarms according to the motion information. Still, it needs to consider how to deal with the target association between the low-confidence frames under the sparse registration information of the before and after and combine the motion characteristics of the target to form the trajectory information. Although some studies use GAN Network (Generative Adversarial Network) to carry out super-resolution for small remote sensing ships [10], this method could be more practical in low-resolution images with low contrast. Currently, remote sensing ship target detection mainly relies on public data sets based on Google Earth, which are primarily oriented to acceptable recognition applications and difficult to use in low-resolution scene applications [11].

Previous high-resolution remote sensing ship target detection algorithms mainly focused on recognition and classification based on machine learning and deep learning. Low-resolution data are challenging to form accurate feature or texture detail training data sets. At the same time, the statistical characteristics of the target and false alarms are consistent, which makes it difficult to eliminate some false alarms by accumulating data [12]. In reference [13], a multi-frame sequence registration method is designed. The target speed is set to be constant, and the motion information is obtained according to the linear correlation of the target trajectory of each frame. This method sets the target to move linearly, which requires high registration accuracy. However, accurate registration in the ocean area is challenging, so ocean-going ship target detection should refrain from registration processing [14]. In reference [15], based on single frame discrimination based on SOLOv2, an autocorrelation-filtering algorithm was used to form the trajectory association of the moving target and remove the false alarm twice to obtain the target motion information. This method is associated with the problems of a large amount of calculation, strenuous training, and strict requirements on the accuracy of auxiliary image information. In literature [16], constant false alarm detection combined with LeNet network identification was used, and the middle latitude method was used to associate trajectories with global nearest neighbors. Although the calculation amount was small, it was relatively simple to select test data clouds, which was challenging to cover complex scenarios in practical applications. In addition, literature [17] uses the concatenated anchor-assisted detection network (CR2A-NET) to preprocess RESNET34 to remove many false alarms in the air and sea area and the coarse and fine structure network to complete the ship target in any direction. This architecture has become a relatively common method in ship detection and has good adaptability to high-resolution airborne remote sensing. However, it is difficult for RESNET34 to distinguish the blurred object from the interference cloud for low-resolution images. In literature [18], the Multiple Hypothesis Tracking (MHT) algorithm was used to track the target and eliminate the false target to obtain the motion state. However, this method depended on the accuracy of the segmentation stage, and it could not stop the cloud fragmentation target with similar motion information.

More methods are necessary, i.e., than the existing methods, to mine the hidden information in the sequence observation data, or the existing methods focus on the sequence association of the target and need more judgment of the characteristics of the target [19]. Low-resolution wide-area remote sensing target detection mainly removes some interference through target characteristics and visual attention mechanism [20]. In addition, it removes false alarms through association and motion state in big data. Aiming at the problems of the previous design methods, this paper proposes a real-time ship target detection method based on sequence images to apply space-ground integration real-time information acquisition. Firstly, the saliency enhancement calculation was carried out for the low-resolution dim target, and the contrast between the gray level of the target and the surrounding background was improved by iterative morphological reconstruction.

Furthermore, the feature inconsistency caused by minor pixel interference, gray level noise, and edge dispersion was removed to enhance the integrity of the target extraction. Then, the contrast between the target and the surrounding noise is used to realize the adaptive segmentation calculation of the front and rear scenes, and the corresponding position of the suspected target at the pixel level is obtained. Then, according to the simplified HOG feature and support vector machine, most of the false alarm targets were eliminated to get the second-level suspected target results, and the target sequence was obtained according to the local prediction search of sequence images and the non-sequence false alarm targets (including stationary ship targets) were eliminated. Finally, the lightweight CNN network S-YOLO is used to identify the target sequence and finally identify the remaining targets. Because of the multi-level identification, the part with a significant computational load decreases with the gradual reduction of suspected targets, which can effectively meet the accuracy and guarantee the real-time performance of the calculation.

This paper is arranged as follows. Section 2 presents the detection and recognition algorithm of low-resolution sequence targets, and the calculation steps are introduced in detail. And Section 2 also gives the simplified calculation method of the algorithm, as well as the real-time improvement and performance change after simplification. In Section 3, the actual remote sensing data is tested, and the experimental conclusion is given, compared with the previous designs. Finally, Section 4 concludes the proposal.

2. Low-Resolution Sequence Target Detection Algorithm

In low-resolution remote sensing data, the target mainly presents unclear texture and blurred edges, and the gray distribution tends to be consistent. In addition, there are problems in sequence image frames, such as significant frame shaking and complex registration of sea surface scenes. At the same time, there is a considerable amount of data in the wide-field view, which is a great challenge to the real-time performance of data processing [21]. Therefore, the detection of low-resolution wide-area remote sensing targets mainly relies on feature recognition and motion correlation. This section gives a multilevel discriminant detection algorithm to improve accuracy and reduce the false alarm rate. First, the basic algorithm framework is introduced in Section 2.1. Then, Section 2.2 discusses adaptive segmentation preprocessing and SVM false alarm elimination with simplified HOG features. Finally, Section 2.3 discusses the target association of sequence images and the false alarm identification of sequence data based on S-YOLO.

2.1. Basic Framework

The basic framework of the sequence target detection algorithm is shown in Figure 1, including four steps: preprocessing, feature discrimination, association discrimination, and S-YOLO discrimination.

The preprocessing steps include morphology processing, adaptive threshold segmentation, and connected domain area elimination of false alarms. Morphology processing is often used to enhance the edges of breakable objects. This framework adopts multiple iterative morphology processing to improve the adaptability of the enhancement processing to different complex scenes. The adaptive threshold segmentation uses the ratio of the

target region to the mean of the noise region to judge the front and rear scenes, which has good adaptability to low resolution and large fields of view. Finally, the connected domain was labeled for the binary results, the area of each region was calculated, and the larger or smaller parts were eliminated.

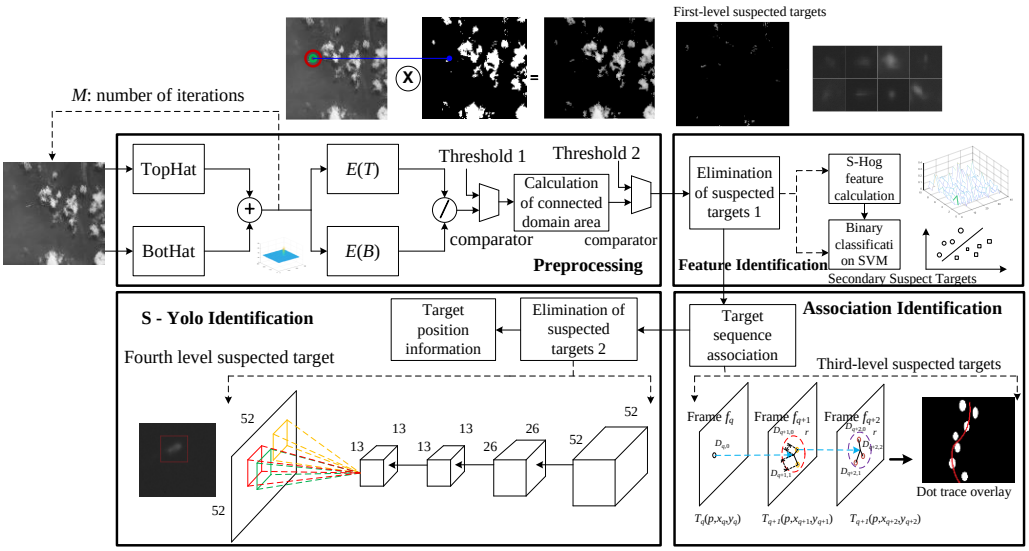


Figure 1. The basic framework of ship target detection algorithm in low-resolution broad area sequence remote sensing images.

For the remaining suspected targets, the features were identified. Considering that the parts are relatively simple and to ensure the real-time performance of the calculation, S-HOG (simplified HOG algorithm [13]) was used for analysis. After removing the gamma correction step, the obtained first-level suspected target image was divided into 16 blocks, and eight angular direction histograms were calculated, respectively. The 9X32 S-HOG feature result was obtained by merging 2 × 2 adjacent blocks. Input S-HOG to binary nonlinear SVM [14] to generate feature discrimination results and eliminate false alarm targets.

Target sequence association, through the position prediction and search of multi-frame sequence images, the suspected target, which cannot form trajectory information in each frame, was eliminated, and the trajectory information of the third-level suspected target was obtained.

In the S-YOLO identification process, the secondary suspected target is further refined through feature labeling and CNN network identification. Considering the small scale of the third-level suspected target slice, the simplified four-layer YOLO network [15] was adopted to identify the associated sequence slice. The series was judged invalid when the number of false alarms in the sequence exceeded the limited threshold.

2.2. Preprocessing

In ship target detection, especially in low resolution and low signal-to-noise ratio image data, it is often difficult to distinguish the gray information of the target from the surrounding background, and it is difficult to eliminate the complex sea state interference information. At the same time, for wide-area and large-width images, the amount of data in a single frame is large, so it is challenging to ensure the real-time performance of the calculation by directly using high-performance and large-parameter CNN networks.

To solve the above problems, morphological enhancement was first carried out, and iterative TopHat transform was used to improve the contrast of the target edge relative to the background and reconstruct the target contour. Secondly, contrast-based binary segmentation was used to remove background noise information. Finally, the connected

domain was labeled, and the regions with larger or smaller related domain areas were removed to obtain the first-level suspected targets.

I. Morphological Reconstruction

Define I as an image matrix of size $M \times N$, $I_0 = I$, then the iterative TopHat transformation process is as follows:

$$I_{m+1} = (OTH_\beta(x,y)\mu_1 + CTH_\beta(x,y)\mu_2) \quad (1)$$

where $OTH_\beta(x,y) = (I_m - (I_m \Theta \beta) \oplus \beta)(x,y)$, $CTH_\beta(x,y) = ((I_m \oplus \beta) \Theta \beta - I_m)(x,y)$, β stands for square full 1 structure elements, $x \in [0, M-1]$, $y \in [0, N-1]$, $m \in [0, R-1]$, R stands for the number of iterations, OTH and CTH, respectively, stand for forward and inverse top-hat transformation, $\oplus$ stands for expansion operation, Θ stands for corrosion operation, μ_1 and μ_2, respectively, stand for weight parameters (the default is 1).

The result of morphological reconstruction is I_R. Figure 2 shows the gray distribution of the ship target before and after reconstruction and the gray distribution of the cloud fragmentation target before and after reconstruction. It can be seen that after reconstruction, the ship target energy is more concentrated, the contour edge is easier to distinguish compared with the background noise, and the background interference noise is effectively suppressed. After reconstruction, the contour of the broken cloud target is also more apparent, and most false alarm targets can be eliminated by contour. In the reconstruction process, the number of reconstruction iterations needs to be set according to the target size. If the number of iterations is too small, the effect cannot be enhanced, and the computation will be increased if the number of iterations is too large.

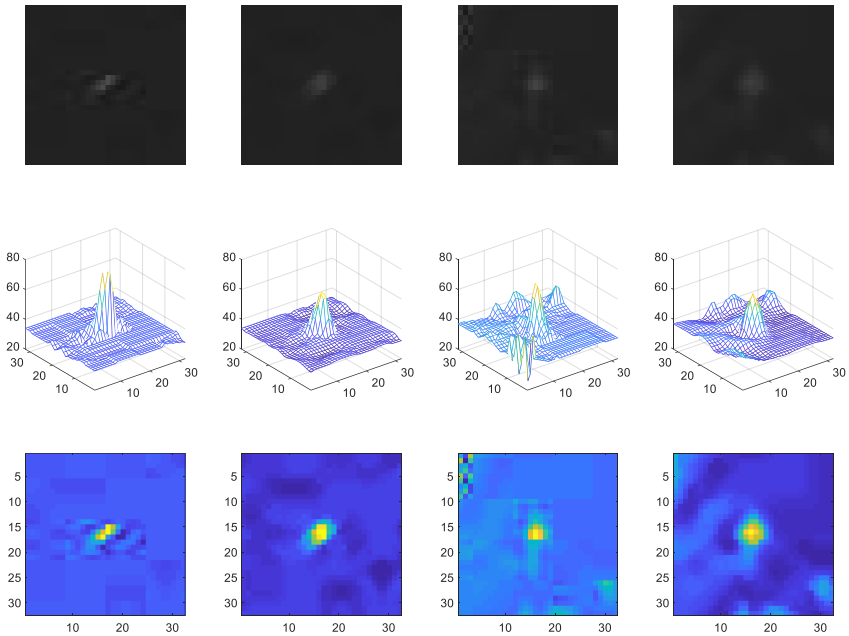


Figure 2. Gray distribution comparison before (**left**) and after (**right**) reconstruction of ship target and broken clouds.

II. Adaptive Threshold Segmentation

The enhanced image data I_R is segmented with an adaptive threshold. The detection process includes the information statistics of three-square Windows: target window T_a, protection window P, and background window B. The length parameters of three types of windows are r_{T_a}, r_P and r_B, respectively. The target window mainly includes the gray information of the target to be detected. The protection window mainly consists of the gray information between the target and the background, which is used to protect the dispersion part of the target from being counted into the background window. The background

window mainly covers the sea surface noise information. The judgment basis for the target detected in the target window is:

$$\delta_e = \frac{\mu_{T_b}}{\mu_B} > T_{Thr} \tag{2}$$

where μ_{T_b} is the mean value of the target window, μ_B is the mean value of the background window, T_{Thr} is the comparison threshold, and δ_e is the mean signal-to-noise ratio of the reconstruction result of the target region.

The result of the binary image obtained by threshold segmentation is:

$$Y_b(x,y) = \begin{cases} 1, when\ \frac{\mu_{T_b}}{\mu_B} > T_{Thr} \\ 0, else \end{cases} \tag{3}$$

After adaptive threshold segmentation, the weak target is extracted according to the contract, which reduces the amount of image data to be processed later and eliminates background noise interference on the suspected weak target. It can be seen in Figure 3 that after adaptive segmentation, the dim target was significantly extracted, and the background noise was isolated.

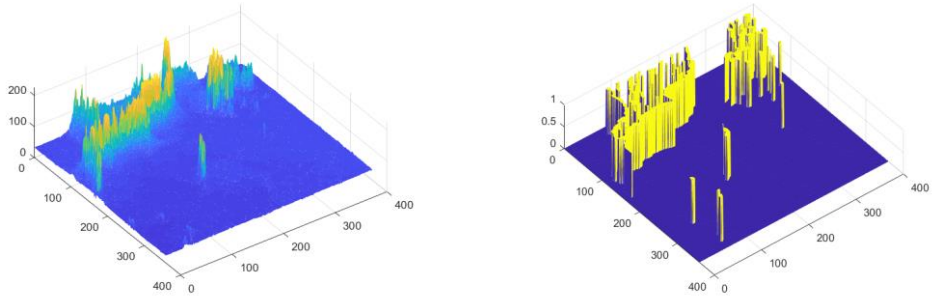


Figure 3. Comparison of remote sensing images before (**left**) and after (**right**) segmentation by adaptive threshold.

III. Connected Domain Labeling

The connected domain set $A_i = \left\{ \vec{x}_i, \vec{y}_i \right\} (i \in [0, P-1])$ is obtained by marking the connected domain of binary image Y_b, P represents the number of regions, $\vec{x}_i$ represents the set of x coordinates of the i-th region, and $\vec{y}_i$ represents the set of y coordinates of the i-th region. In the annotation result, the small area is removed as a false alarm target, the large area is removed as cloud, land, and other targets, and the binary image Y_b' is obtained.

$$A_i' = \begin{cases} A_i, if\ S_{low} < Area_i < S_{high} \\ \phi, else. \end{cases} \tag{4}$$

where ϕ represents null, $S_{low} = 5$ and $S_{high} = 100$ represent the lower and upper limits of the area, respectively, and $Area_i$ represent the area of the A_i region.

After the connected domain labeling and connected domain area elimination, the first-level suspected target set $A_i'(P'$ regions) is obtained.

In the preprocessing process, considering that the target width is about three pixels, the size of the square all '1' structure element β used in morphological reconstruction is 3×3. The number of reconstruction iterations R is related to the mean signal-to-noise ratio (SNR) δ_e of the reconstruction results of the target area. In addition, it is related to illumination conditions, imaging side-swing angle, target sea state, and other conditions. In practical applications, considering the working mode, the same area is often observed in the same period. Therefore, δ_e is observed after processing a large number of targets in the region; as shown in Figure 4, 80 groups of data were counted. When the number of iterations is 3, δ_e reaches the optimum, and the optimal number of reconstruction iterations $R_{optimal} = 3$ is

obtained. The threshold T_{Thr} in adaptive threshold segmentation is mainly related to the mean/variance of the sea state in the imaging area and the gray level of the sea surface background. Considering that the gray level distribution of the sea surface tends to be unified after reconstruction at low resolution, 80 groups of data were statistically analyzed and $T_{Thr} = 1.2$ was obtained. The connected domain was marked, and the suspected target area was calculated. The area eliminated the larger and smaller targets, and the upper and lower limits of the area were $S_{low} = 10$ and $S_{high} = 50$, respectively.

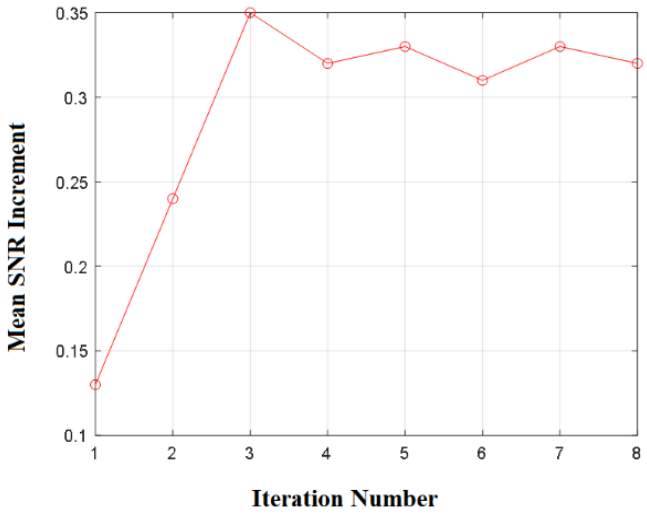


Figure 4. Relationship between the number of reconstruction iterations R and the mean SNR increment after processing.

2.3. Feature Identification

The feature identification process calculates the gray features of all the targets in the first-level suspected target set and classifies whether they are ship targets according to the elements. Among them, the simplified HOG feature (S-HOG) is used for gray feature calculation, and nonlinear binary classification SVM is used for binary classification.

I.S-Hog Feature Calculation

After preprocessing, simplified HOG enhanced the edge and gray distribution of the target, while gamma enhancement will increase the noise interference on the target. Therefore, the calculation process of S-HOG is mainly divided into the following steps (as shown in Figure 5):

Step 1: For any region A_i', take its centroid $C_i = [mean(\vec{x}_i'), mean(\vec{y}_i')]$ as the center, extract slice D_i with a length of 52 pixels from the original image I, and perform S-HOG feature calculation;

Step 2: Slice D_i was segmented. The size of each cell was 13 × 13 pixels. The gradient features calculated for each cell were divided into 8 directions, and each cell had 8 feature values;

Step 3: The features of every four adjacent cells form a block, then there are 9 blocks in total, each block has 32 eigenvalues $Block(i,j) = [Cell(i,j), Cell(i+1,j), Cell(i,j+1), Cell(i+1,j+1)]$, and the eigenvector $\vec{Block}$ has 288 eigenvalues in total.

II. Nonlinear Binary SVM Calculation

The eigenvalue $\vec{Block}_i$ of single slice image D_i is calculated by SVM binary classification. Considering the high feature dimension, nonlinear SVM is adopted. The training and decision calculation are divided into the following steps:

Step 1: First, input the training feature value and category set A, B is the category to be distinguished, '0′ represents the interference target, '1′ represents the ship target;

Step 2: Construct and solve the optimization formula

$$\min 0.5 \sum_{i=0}^{j-1} \sum_{j=0}^{P'-1} y_i y_j a_i a_j K(x_i, x_j) - \sum_{j=0}^{P'-1} a_j \tag{5}$$

The constraint condition is

$$\sum_{i=0}^{P'-1} y a_i = 0, 0 \le a_i \le C \tag{6}$$

where C is the loss parameter, a_i is the Lagrange multiplier, and $K(x_i, x_j)$ is the radial basis kernel function:

$$K(x_i, x_j) = e^{-\gamma ||x_i - x_j||^2}, \gamma > 0 \tag{7}$$

According to Equations (5)–(7), the optimal Lagrange multiplier solution a^* is obtained, and the threshold b is calculated as follows:

$$b = y_i - \sum_{i=0}^{P'-1} y_i a_i^* K(x_i, x_j) \tag{8}$$

Step 3: The final decision function formula is

$$f(x) = \text{sgn}(\sum_{i=0}^{P'-1} a_i^* y_i K(x, x_i) + b) \tag{9}$$

Therefore, after feature identification of the first-level suspected target set A_i', the result of the second-level suspected target set A_i'' (the number of effective targets is P'') is as follows:

$$A_i'' = f(\vec{Block_i}(A_i')) \tag{10}$$

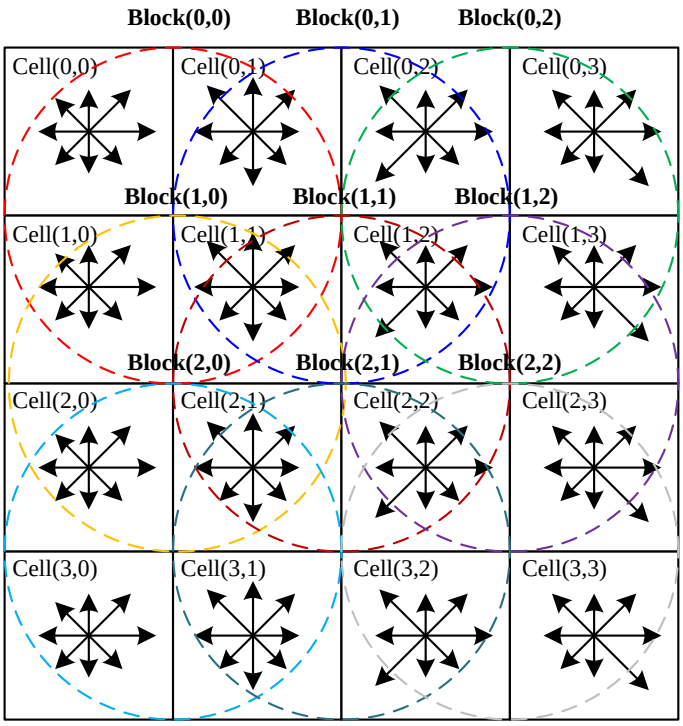


Figure 5. S-HOG eigenvalues calculated by slices and their set.

2.4. Association Identification

The association identification mainly eliminates the interference false alarm target through the correlation of the target in the front and back frames. The assumption is that the target is considered to move at a constant speed in a short observation time, and the target position difference between adjacent frames is equal to the sum of the moving distance and the pose and orbit error between frames. Considering the randomness of pixel-level error between adjacent frames in orbit, a conservative circular search method is adopted in the inter-frame search process. The search radius r is proportional to the sum of the moving distance, inter-frame pose, and orbit error. In the process of association discrimination, the more frames selected, the higher the detection accuracy, but it is also necessary to consider that the real-time performance of the calculation results will be affected in the case of multi-frame association.

The image in frame q is defined as f_q, and the target in frame q is marked as $D_{q,p}$, where $p \in [0, P'' - 1]$, the image in frame $q+1$ is f_{q+1}. The coordinate $T_q(p,x_q,y_q)$ of the target $D_{q,p}$ in frame q is taken as the center of the circle and r as the radius to search for the target in the $q+1$ frame. That is, the three-level suspected target set $A'''_{q,i}$ in frame q (the number of effective targets is P''') can be expressed as follows.

$$A'''_{q,i} = \begin{cases} A''_{q,i}, \ when \ |D_{q,i} - D_{q+1,j}|_m < r \ and \ |D_{q+1,j} - D_{q+2,k}|_m < r \ exist. \\ \phi, else. \end{cases} \tag{11}$$

where $j, k \in [0, P'' - 1]$ and $|\bullet|_m$ represent the Mahalanobis distance of two coordinate points. Therefore, if the target within the radius appears in three consecutive frames, the target is considered to exist; otherwise, the target is considered not to exist. In practical applications, the number of multi-frame association sequences of Formula (11) can be increased according to the real-time requirements to improve the confidence of target discrimination.

After association identification, the false alarm caused by the interference target, especially the broken cloud, can be eliminated well. However, it is also necessary to consider the missing detection of a frame caused by the cloud cover and sea state interference in the sequence association process and increase the tolerance appropriately to improve the adaptability of association identification.

2.5. S-Yolo Identification

After the three-level identification of false alarm elimination, cloud interference with movement association features similar to ship targets will still be broken. In this case, the detailed features of the image should be considered for further false alarm elimination. At the same time, at this time, the data is retained in the form of slices, and the amount of target and false alarm data decreases sharply. Therefore, the CNN method can be considered for target refined feature recognition.

In preprocessing, the segmentation of front and rear scenes results in insufficient target centroid positioning accuracy. The HOG feature cannot effectively describe the refined features such as low-resolution cloud ship edges and trailing traces. Therefore, a simplified version of the Yolo algorithm (S-YOLO) was designed in this section. The lightweight network was designed with Yolo V2 Tiny [10] as the baseline for further feature recognition of the tertiary suspected target set A'''_i.

Table 1 shows the architecture of the S-YOLO network. The input was 52 × 52 image data, which mainly included three convolutional layers, a Maxpool layer, and one Yolo prediction layer, totaling 23,296 parameters. S-yolo outputs the target's position, category, length, and width on the 52 × 52 slice image.

Table 1. S-YOLO network designed in this paper for ship target slice detection and recognition.

	Type	Filter Number	Size/Stride	Output
Backbone	Input		52×52	
	Convolutional-1	16	3×3	$52 \times 52 \times 16$
	Maxpool-1	-	$2 \times 2/2$	$26 \times 26 \times 16$
	Convolutional-2	32	3×3	$13 \times 13 \times 32$
	Maxpool-2	-	$2 \times 2/1$	$13 \times 13 \times 32$
	Convolutional-3	64	3×3	$13 \times 13 \times 64$
	Maxpool-3	-	$2 \times 2/1$	$13 \times 13 \times 64$
Prediction	Yolo-1		-	

For S sequence set $\overrightarrow{A}^{'''}_{q,i} = [A^{'''}_{q,i}, A^{'''}_{q+1,i}, \cdots, A^{'''}_{q+S-1,i}]$ of target i in the sequence image, after S-YOLO identification, if the judgment result of the effective target in the sequence exceeds the threshold D, the sequence is considered as effective sequence and target, that is, the fourth level suspected target set $A^{''''}_{q,i}$ is calculated as follows:

$$A^{''''}_{q,i} = \begin{cases} A^{'''}_{q,i}, \ when \ [\sum_{t=q}^{q+S-1} Y(A^{'''}_{t,i})]/S \geq Thr_Y. \\ \phi, \ else. \end{cases} \tag{12}$$

3. Experiment and Comparison

The experimental environment used Windows 10 operating system, 16G memory, I7-10400F CPU, NVIDIA GTX 3060TI GPU, Pytorch1.8, and Matlab2018 as the development environment for the test and verification of this method.

The algorithm was tested using 100 sets of remote sensing images (10 frames per group) with 10 m to 15 m resolution, each with a pixel of 10,240 × 10,240, of which 80 sets were used as the training dataset, and 20 sets were used as the test dataset. The targets in the test data were all large vessels ranging from 30 m to 150 m, with individual targets ranging from 3 to 10 pixels in length and 1 to 3 pixels in width.

3.1. Parameter Description

To evaluate the method designed in this paper and compare it with other works of literature, four general indexes [11], precision, recall, F1 number, and calculated frame frequency, were used to describe the detection results.

Precision (P_s) indicates the ratio of the number of correctly predicted actual values to the number of all predicted actual values. For example, for remote sensing ship target detection, a higher recall ratio means that the ship target is better detected, which is defined as follows:

$$Ps = \frac{TP}{TP + FP} \tag{13}$$

The recall represents the ratio of correctly predicted truth values to the actual number of truth values. For remote sensing ship target detection, a higher recall rate represents the robustness of ship target detection for real targets, which is defined as follows:

$$Rc = \frac{TP}{TP + FN} \tag{14}$$

where TP represents the number of predicted valid values and actual valid values; FN represents the number of false values indicated and true values actually; FP is the number of predicted valid values and actual false values.

The $F1$ number is the complete result of the recall and recall index, which is defined as follows:

$$F1 = 2\frac{PsRc}{Ps + Rc} \tag{15}$$

The calculated *FPS* (Frames per Second) was used to represent the real-time performance of the algorithm. For example, in large-width remote sensing image detection, the image frame represented an image used for detection in practice.

3.2. Analysis of Experimental Results

During the experiment, the free satellite remote sensing data disclosed by Google Earth was selected. When downloading the data, the options of level 13 (in-space resolution: 38.22 m), Level 14 (resolution: 19.11 m), and Level 15 (resolution: 9.55 m) were selected. Download and make data sets for offshore and ocean-going areas in Asia. According to low-resolution (30 m), offshore, cloud-through and normal targets, the data were divided into four types for separate testing.

The experiment tested 20 groups (denoted as G), a total of 200 images (SNR $\geq$ 3), and there was a total of 2670 targets. A total of 80% of the data was used for training and 20% for testing. Among them, groups 1 to 5 were small ship targets of about 30 m, which were denoted as class G_1; groups 6~10 were the target of sailing ashore, which were denoted as class G_2; groups 11~15 were the ship targets passing through the cloud layer, which were denoted as class G_3; groups 16 to 20 were conventional sea surface ship target scenes, denoted as class G_4.

Table 2 and Figure 6 show the performance comparison between Yolo Tiny V2/V3 in references [7,8] and the results obtained by the proposed algorithm. Yolo Tiny V2/V3 algorithm only depends on the image feature information of the ship target itself, so the detection effect of the docking target (class G_2) and small target (class G_1) was poor. The docking target was caused by the interference caused by artificial construction, buoys, artificial islands, and berthing ships near the port. Figure 7 shows the detection results of remote-sensing ship targets in different scenarios.

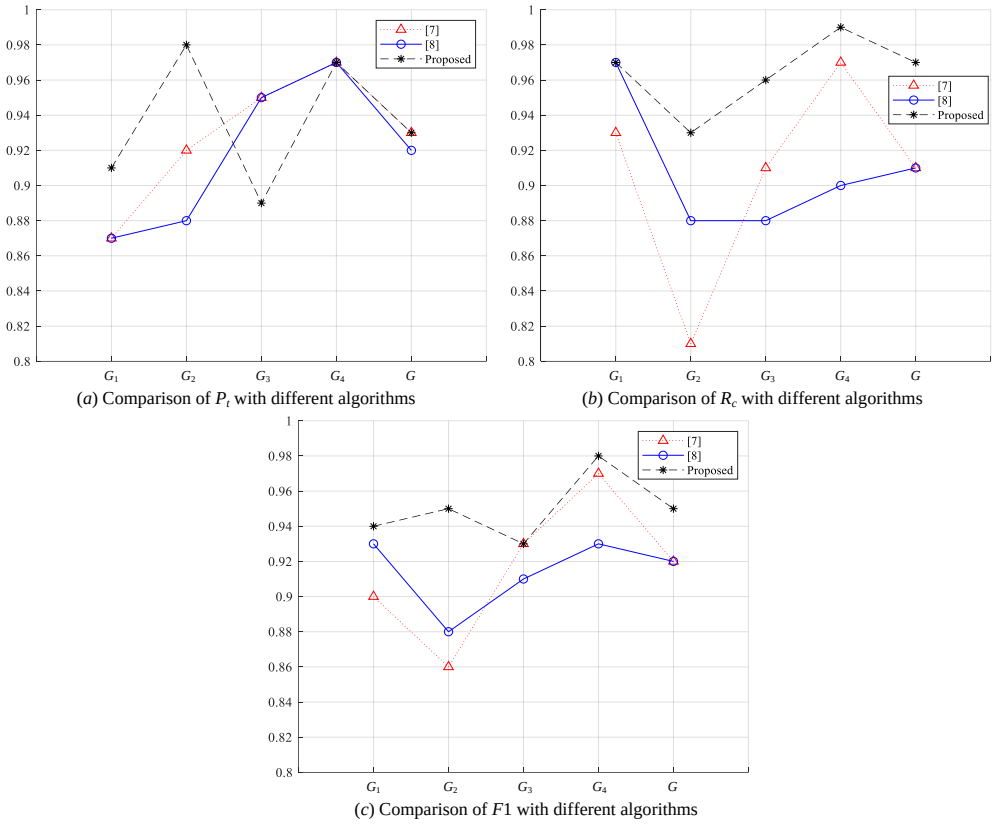
Table 2. Comparison of 20 groups of remote sensing data test results [*Pt*, *Rc*, *F*1].

No.	Model	G_1			G_2			G_3			G_4			G		
1	Yolo Tiny V2 [7]	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1
		0.87	0.93	0.90	0.92	0.81	0.86	0.95	0.91	0.93	0.97	0.97	0.97	0.93	0.91	0.92
2	Yolo Tiny V3 [8]	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1
		0.87	0.97	0.93	0.88	0.88	0.88	0.95	0.88	0.91	0.97	0.90	0.93	0.92	0.91	0.92
3	Proposed	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1
		0.91	0.97	0.94	0.98	0.93	0.95	0.89	0.96	0.93	0.97	0.99	0.98	0.93	0.97	0.95

The proposed algorithm outperformed the traditional lightweight Yolo algorithm in different scenarios. The main reason was the four-level cascaded authentication architecture. In the first stage, false alarms with the large area were eliminated through front and rear scene segmentation; in the second stage, false alarms with significant differences between training targets and scene features were destroyed; in the third stage, non-moving wrong alarm targets were eliminated through multi-frame cascade search; in the fourth stage, false alarms were further eliminated through fine feature discrimination among the remaining suspected targets. Table 3 and Figure 8 compare the recall ratio, recall ratio, and F1 number of different discrimination steps of the design method in this paper. It can be seen that simple preprocessing and S-HOG feature discrimination cannot eliminate false alarms well. On the other hand, association effectively improved detection accuracy by 15%, and S-YOLO discrimination improved each index by 5%.

Table 3. Performance comparison of calculation results of different discrimination steps of the design method in this paper.

Model	Preprocessing			Feature Identification			Association Identification			S-Yolo Identification		
Poposed	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1	*Pt*	*Rc*	*F*1
	0.68	0.59	0.63	0.81	0.68	0.74	0.88	0.93	0.91	0.93	0.97	0.95

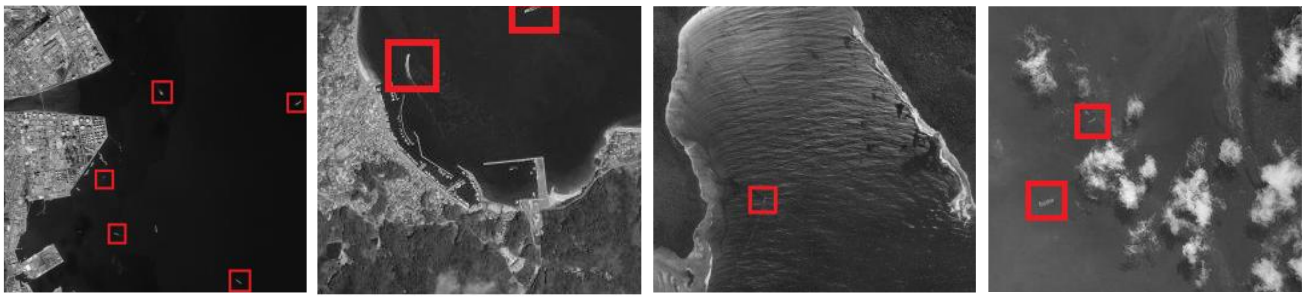


(*a*) Comparison of P_t with different algorithms

(*b*) Comparison of R_c with different algorithms

(*c*) Comparison of $F1$ with different algorithms

Figure 6. Comparison of ship target detection performance between the traditional Yolo algorithm and the proposed algorithm.



(*a*) (*b*) (*c*) (*d*)

Figure 7. Detection results of remote sensing ship targets in different scenarios (**a**) small target G_1 (**b**,**c**) shore target G_2 (**d**) cloud-penetrating target G_3.

Multi-frame association reduced false alarm interference by predicting the trajectory of at least three frames and the nearest target judgment. The more associated frame, the higher the detection accuracy should be. Still, the group delay and cache occupation of the detection result information also increased. Therefore, the appropriate number of associated frames should be selected in the engineering implementation process. For example, after analyzing the number of related frames and the recall rate and F1 number

of discriminant results, as shown in Figure 9, when the number of associated frames was greater than 7, the improvement of the index was less than 0.3%, so 7 can be selected as the number of associated discriminant frames.

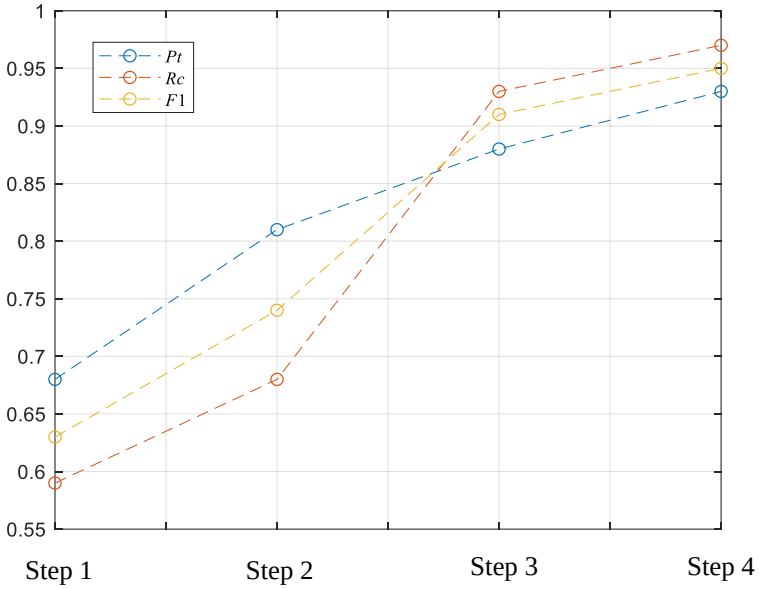


Figure 8. Calculation results in the performance of different discrimination steps of the design method in this paper.

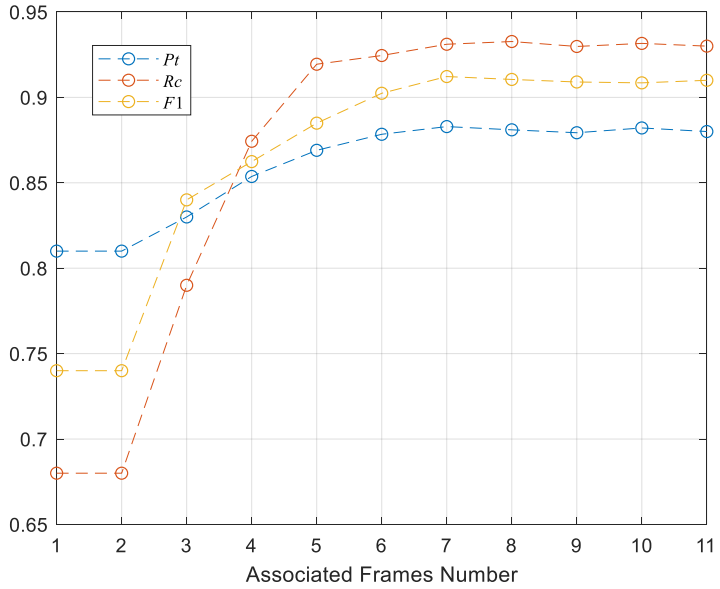


Figure 9. Performance comparison of different numbers of association frames and association discrimination results designed in this paper.

3.3. Real-Time Performance Comparisons

For the input of a single frame of 10,240 × 10,240 pixel image, Yolo Tiny V2 and V3 chose 416 × 416 pixel input. Considering the overlap of 16 pixels between image slices (the target was not more than 150 m), a single image frame needed to be calculated 625 times. In this paper's algorithm design, the central computation time was spent in the preprocessing, and the data volume processed by the subsequent feature discrimination, association discrimination, and S-YOLO discrimination was less than 1% of the preprocessing.

The computing platform of this algorithm was a i7-10700k processor, NVIDIA GeForce RTX 3060 graphics card, with 16 GB DDR4 memory, CUDA 11.1, and MATLAB R2018a.

In the pre-processing calculation process, β adopted the all-1 structure with the size of 3×3, and the iteration number R was selected as 3, then the calculation time of morphological reconstruction was equivalent to completing three image reading and storage operations. The calculation time was proportional to $M \times N \times R / f_{val}$ (f_{val} representing the effective main frequency of calculation, with an average of 2 GHz). The computation time of adaptive threshold segmentation was equivalent to one image reading and storage, and the computation time was proportional to $M \times N / f_{val}$. The calculation time of connected domain label was calculated by fast label, and the time was proportional to $2M \times N / f_{val}$.

In the process of feature identification, the number of first-level suspected targets P' in a single image was between 200 and 1000, and 1000 was selected. Since the input slice size of first-level suspected targets was 52×52, the calculation time of simplified S-HOG algorithm was proportional to $314 \times 52 \times 52 \times P' / f_{val}$, and the calculation time of binary classification SVM was proportional to the number of feature vectors, namely $1044 \times 288 \times P'$. 314 and 288 were statistical values.

In the process of correlation authentication calculation, the input number of second-level suspected targets P''' ranged from 20 to 100, P''' was selected as 100, and the time was mainly spent in searching the circular area with r as the radius of the front and back frames. Considering that the motion of the front and back frames did not exceed 100 pixels, r was selected as 64, then the calculation time of correlation authentication was proportional to $255 \times P'' \times \pi \times r^2 / f_{val}$.

In the calculation process of S-Yolo authentication, the number of second-level suspected targets S was generally no more than 10, and the number of sequence sets in the authentication process was 3. According to the network structure in Table 1, it can be seen that the whole calculation time was proportional to $44 \times 10 \times S \times 23296$, in which 44 was the statistical coefficient.

Table 4 shows the calculation time of four steps and each sub-step in the actual test process, and the total calculation time was 1.7553 s. Table 5 shows the calculation time comparison between this design and previous methods. Considering the literature [7,8], only for less than 512×512 pixels below the calculation of the image, this design input $10{,}240 \times 10{,}240$ pixels. This design on computing time was less than 35% of the reference, and the method of design greatly improved the efficiency of target detection.

Table 4. Time consumption for each step in the actual test process.

No.	Step	Sub-Step	Time Consumption
1	Preprocessing	Morphological reconstruction	0.3389 s
		Adaptive threshold segmentation	0.1207 s
		Connected domain marker	0.2220 s
2	Feature identification	S-HOG	0.4245 s
		Binary SVM	0.1503 s
3	Association identification	Association identification	0.4836 s
4	S-Yolo identification	S-Yolo identification	0.0153 s
5	Total		1.7553 s

Table 5. Comparison of the real-time performance of calculation.

No.	Model	FPS	Single Frame Image Computation Time
1	Yolo Tiny V2 [7]	244	5.12 s
2	Yolo Tiny V3 [8]	220	5.68 s
3	Proposed	–	1.75 s

During the experiment, the design adopted the SNR $\geq$ 3 image data. In practical engineering applications, the SNR has a certain random distribution due to the influence of lighting conditions and certain specific areas at the imaging time. Low SNR will lead to the increase in false alarms and the decrease in accuracy. Figure 10 shows the Receiver Operating Characteristic Curve (ROC) of the proposed method under 6 different signal-to-noise ratios. Noise is added to the test data set to test the robustness of the method. In the ROC curve, the closer the curve is to the top left, the better the method. Any point on the curve represents the choice of different binary thresholds under the current signal-to-noise ratio. The vertical axis R_c represents the percentage of all targets that are correctly classified as true. F_{PR} (False positive rate) is the proportion of real ships that are correctly classified as noise in all noise samples. SNR3.0, SNR2.5, and SNR2.0 curves were very steep and close to each other. The other curves decrease with the decrease in SNR, among which the black curve (SNR = 0.5) had the lowest score.

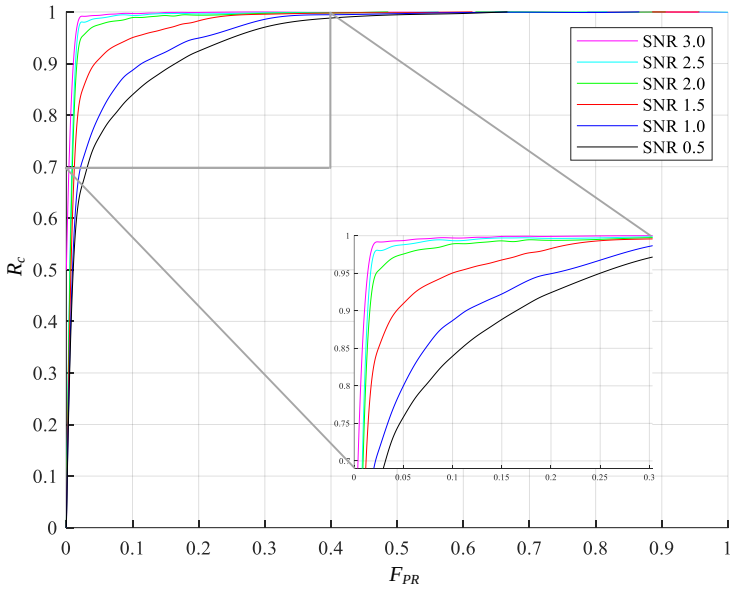


Figure 10. The ROC curves of proposed using different SNRs.

4. Conclusions

The deep convolutional network can effectively complete object detection, semantic segmentation, and completion generation of remote sensing images, but it is also faced with the problems that large-width data cannot be processed in real-time, and low-resolution data is difficult to identify accurately. Especially in the scene with many broken background clouds, the low-resolution data can only rely on the outline and target gray level distribution to distinguish the target. The traditional CNN algorithm often gets many false alarms suspected of being ship targets. To solve these problems, this paper proposes a ship-moving target detection method for low-resolution and large-width remote sensing image sequences, which improves the detection accuracy and real-time computation by combining the multi-level cascade preprocessing method traditional feature detection, sequence association, and CNN features discrimination. In the four-level identification cascade method, false alarms with large target segmentation areas are proposed in the first level, false alarms with large differences between training targets and scene features are eliminated in the second level, non-moving false alarm targets are eliminated by multi-frame cascade search in the third level, and false alarms are further eliminated by fine feature discrimination among the remaining suspected targets in the fourth level. Compared with the traditional Yolo Tiny V2/V3 series network, the proposed method has faster calculation speed, and better detection performance, and the F1 number of detection results increased by 3%. The computation time was reduced by 66%.

Author Contributions: Validation, W.L.; Resources, X.S.; Writing—original draft, J.Y.; Visualization, X.W.; Project administration, D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study aims to improve and optimize the ship targets detection with low-resolution remote sensing images and does not involve human or animal studies. Therefore, ethical review and approval was waived for this study.

Informed Consent Statement: This study is not applicable. This study did not involve human studies.

Data Availability Statement: This research is mainly aimed to improve and optimize the ship targets detection with low-resolution remote sensing images. Experimental data and results are mainly obtained by public data like Google Earth.

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

References

1. Zhang, F.; Wang, X.; Zhou, S.; Wang, Y.; Hou, Y. Arbitrary-oriented ship detection through center-head point extraction. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
2. Xu, Y.; Zhou, J.; Yin, J.; Xie, J.; Wu, B. Review on mission planning strategies and Applications of Earth Observation satellites. *Radio Eng.* **2021**, *51*, 681–690.
3. You, Y.; Ran, B.; Meng, G.; Li, Z.; Liu, F.; Li, Z. OPD-Net: Prow detection based on feature enhancement and improved regression model in optical remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6121–6137. [CrossRef]
4. Qi, Y.; Chen, M.; Wang, M.; Xu, Y.; Gao, F.; Zeng, F.; Niu, J.; Shi, W. Exploration practice and development suggestions of commercial remote sensing satellites in China. *Spacecr. Eng.* **2021**, *30*, 188–194.
5. Tan, Y.; Liang, H.; Guan, Z.; Sun, A. Visual Saliency Based Ship Extraction Using Improved Bing. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1292–1295.
6. Zhang, Y.; Wen, F.; Gao, Z.; Ling, X. A coarse-to-fine framework for cloud removal in remote sensing image sequence. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5963–5974. [CrossRef]
7. Xiong, Y.; Ding, S.; Deng, C.; Fang, G.; Gong, R. Ship detection under complex sea and weather conditions based on deep learning. *J. Comput. Appl.* **2018**, *38*, 3631–3637.
8. Wang, X.; Jiang, H.; Keyu, L. Ship detection in remote sensing images based on improved YOLO algorithm. *J. Beijing Univ. Aeronaut. Astronaut.* **2020**, *46*, 1184–1191.
9. Liu, Y.; Yao, L.; Xiong, W.; Zhou, Z. Fusion detection of ship targets in low resolution multi-spectral images. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 6545–6548.
10. Zhuang, Y.; Li, L.; Chen, H. Small sample set inshore ship detection from VHR optical remote sensing images based on structured sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2145–2160. [CrossRef]
11. Yao, L.; Zhang, X.; Lyu, Y.; Sun, W.; Li, M. FGSC-23: A large-scale dataset of high-resolution optical remote sensing image for deep learning fine grained ship recognition. *J. Image Graph.* **2021**, *26*, 2337–2345.
12. Yu, J.; Peng, X.; Li, S.; Lu, Y.; Ma, W. A Lightweight Ship Detection Method in Optical Remote Sensing Image under Cloud Interference. In Proceedings of the 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Glasgow, UK, 17–20 May 2021; pp. 1–6.
13. Kadyrov, A.; Yu, H.; Liu, H. Ship detection and segmentation using image correlation. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 3119–3126.
14. Li, H.; Man, Y. Moving ship detection based on visual saliency for video satellite. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 1248–1250.
15. Xue, X.; Du, C.; Chen, Y.; Wang, S.; Yu, W.; Zhang, P. Fast Ship Tracking Algorithm for Remote Sensing Video Based on Background Data Mining And Adaptive Selection. In Proceedings of the 2021 IEEE International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 27–29 August 2021; pp. 345–351.
16. Lin, X.; Yao, L.; Sun, W.; Liu, Y.; Chen, J.; Jian, T. GF-4 target tracking of moving ships in cloud fragmentation environment. *Space Return Remote Sens.* **2021**, *42*, 127–139.
17. Yu, Y.; Yang, X.; Li, J.; Gao, X. A cascade rotated anchor-aided detector for ship detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 1–14. [CrossRef]
18. Yao, L.; Liu, Y.; Wu, Y.; Xiong, W.; Zhou, Z. Ship target tracking based on Gaofen-4 satellite. In Proceedings of the 4th Annual High Resolution Earth Observation Conference, Wuhan, China, 17–18 September 2017; pp. 1–16.
19. Tian, Y. Application research and analysis of Gaofen-4 satellite. *Sci. Technol. Innov. Guide* **2020**, *17*, 22–23.

20. Ren, Z.; Tang, Y.; He, Z.; Tian, L.; Yang, Y.; Zhang, W. Ship Detection in High-Resolution Optical Remote Sensing Images Aided by Saliency Information. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
21. Lei, L.; Xu, G.; Li, W.; Song, Q. Ship target detection algorithm and hardware acceleration based on deep learning. *Comput. Appl.* **2021**, *41*, 162–166.