Article:
Ma, M., Chen, J., Liu, W. et al. (2018) Ship classification and detection based on CNN using GF-3 SAR images. Remote Sensing, 10 (12). 2043. ISSN: 2072-4292

https://doi.org/10.3390/rs10122043

remote sensing

MDPI

Article

Ship Classification and Detection Based on CNN Using GF-3 SAR Images

Mengyuan Ma [1], Jie Chen [1,2], Wei Liu [3] and Wei Yang [1,*]

1 School of Electronic and Information Engineering, Beihang University, Beijing 100191, China; mamengyuan@buaa.edu.cn (M.M.); chenjie@buaa.edu.cn (J.C.)
2 Collaborative Innovation Center for Geospatial Technology, Wuhan 430079, China
3 Electronic and Electrical Engineering Department, University of Sheffield, Sheffield S1 3JD, UK; w.liu@sheffield.ac.uk
* Correspondence: yangweigigi@sina.com; Tel.: +86-010-8233-8670

Received: 22 October 2018; Accepted: 12 December 2018; Published: 14 December 2018

Abstract: Ocean surveillance via high-resolution Synthetic Aperture Radar (SAR) imageries has been a hot issue because SAR is able to work in all-day and all-weather conditions. The launch of Chinese Gaofen-3 (GF-3) satellite has provided a large number of SAR imageries, making it possible to marine targets monitoring. However, it is difficult for traditional methods to extract effective features to classify and detect different types of marine targets in SAR images. This paper proposes a convolutional neutral network (CNN) model for marine target classification at patch level and an overall scheme for marine target detection in large-scale SAR images. First, eight types of marine targets in GF-3 SAR images are labelled based on feature analysis, building the datasets for further experiments. As for the classification task at patch level, a novel CNN model with six convolutional layers, three pooling layers, and two fully connected layers has been designed. With respect to the detection part, a Single Shot Multi-box Detector with a multi-resolution input (MR-SSD) is developed, which can extract more features at different resolution versions. In order to detect different targets in large-scale SAR images, a whole workflow including sea-land segmentation, cropping with overlapping, detection with MR-SSD model, coordinates mapping, and predicted boxes consolidation is developed. Experiments based on the GF-3 dataset demonstrate the merits of the proposed methods for marine target classification and detection.

Keywords: Synthetic aperture radar (SAR); marine target classification; marine target detection; convolutional neutral network (CNN)

1. Introduction

With continuous development of Synthetic Aperture Radar (SAR) technology, an increasing number of very high resolution (VHR) SAR images have been obtained, providing a new way to strengthen marine monitoring. Different from optical sensors, SAR is capable of working in all-day and all-weather conditions, and it is receiving more and more attention. However, it is very time-consuming to interpret SAR images manually because of speckle noise, false targets, etc. With increasing demand for ocean surveillance in shipping and military sectors, marine target classification and detection has been an important research area in remote sensing with great application prospects. In this work, we will focus on marine target classification on patch level and marine target detection in large-scale SAR images.

Earlier studies on marine target classification were carried out on simulated SAR images due to a lack of real image samples [1]. In recent years, with the deployment of several spaceborne SAR satellites such as TerraSAR-X, RadarSat-2, and GF-3, a wide variety of SAR images with different

resolutions and covering different regions in the world have been obtained. Up to now, researches on marine target classification in SAR images are mainly focused on large ships with distinctive features such as oil tankers, container ships, and cargo ships [2–6]. The scattering characteristics of the three kinds of ships have been fully exploited by some initial works [2,6]. However, classification of other marine targets such as platform, windmills, and iron towers were not considered. Some studies went further to explore deeper features of different targets combined with classifiers such as sparse representation [3,5] and support vector machine (SVM) [7]. The work in Reference [5] employed histogram of oriented gradients (HOG) features and dictionary learning to performing classification with an accuracy of 97.5% for the three kinds of ships. While these feature-based classifiers can achieve high performance, the features have to be carefully designed especially when dealing with a wide variety of targets. There are also some works focused on combining the complementary benefits of traditional machine learning classifiers [4,8]. However, the classifier-combination strategy increases the computational complexity as it applies the classifiers one by one.

Different from the carefully designed feature-based methods mentioned above, the CNN based methods can extract the deep features of targets automatically, which have made great progress in object classification and recognition in recent years. Some CNN models such as Alexnet [9], GoogleNet [10] and ResNet [11] are capable of working on ImageNet dataset including 1000 classes of images with high accuracy, showing great potential in object classification and recognition. Motivated by previous works in target classification, CNN models have been used in SAR target classification in some earlier studies [12–16], most of which used the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset containing 10 classes of military ground targets. Chen et al proposed the ConvNets which consist of five convolutional layers and three pooling layers, without fully connected layers being used [13]. Its classification accuracy among the ten classes reached 99.13%, which was a great contribution to target classification. Driven by the demand of ocean surveillance, an increasing number of works employed CNN for marine target classification and recognition [17–20]. The work in Reference [19] proposed a simple CNN model with two convolutional layers, two pooling layers and two hidden layers and it was the initial work to perform object classification in oceanographic SAR images. Bentes et al. [18] built a larger dataset consisting of not only ships, but also manmade platforms and harbors, after which a CNN model with four convolutional layers, four pooling layers, and a fully-connected layers was introduced. It adopts multi-looking images in different channels and achieves 94% accuracy among the five types of marine targets, which is far superior in performance to those of other CNN models and traditional machine learning methods. However, previous studies on marine target classification using CNN algorithms only focus on four or five types of marine targets, limiting their practical applications and the networks adopted over simple structures or inappropriate layer arrangements, failing to fully realize the CNN's potentials in classifying marine targets in SAR images. To solve this problem, in this work we propose a novel CNN structure to classify eight types of marine targets in SAR images with higher accuracy than the existing methods.

As for the detection task, many algorithms including Constant False Alarm Rate (CFAR) based methods [21,22], feature based methods [23] and CNN based methods [24,25] have been developed to detect marine targets in SAR images. Among them, the CFAR based methods are the most widely used ones due to their simplicity. Traditional CFAR based methods determine the detection threshold by estimating the statistical models of the sea clutters, including Rayleigh distribution [26], Gamma distribution [27] and K-distribution [28], etc. Usually, the CFAR based methods are applied after sea-land segmentation to rule out false alarms on land, such as buildings and roads. However, the performance of CFAR based methods is not satisfactory under low-contrast conditions. In addition, they fail to give the labels of different targets because of its lack of classification layers. For feature-based methods, in Reference [23], an effective and efficient feature extraction strategy based on Haar-like gradient information and a Radon transform is proposed, however, the features have to be designed carefully to achieve a good performance.

In recent years, the region-based CNN networks such as Faster-RCNN [29], YOLO [30], and SSD [31], which can not only generate the coordinates, but also predict the labels of the targets, have shown a great success on the PASCAL VOC dataset [32]. Faster-RCNN uses a deep convolutional network to extract features and then proposes candidates with different sizes by the Region Proposal Network (RPN) at the last feature map. The candidate regions are normalized through RoI Pooling layer before they are fed into fully connected layers for classification and coordinates regression. This algorithm can detect objects accurately but cannot realize real time detection. YOLO processes images at a faster speed but with lower accuracy than Faster-RCNN. An end-to-end model called SSD was proposed in Reference [31], which can detect the target at real time with high accuracy. It generates region proposals on several feature maps of different scales while Faster-RCNN proposes region candidates with different sizes on the last feature map provided by the deep convolutional network.

The CNN based methods have been used for target detection in SAR images, e.g., ship detection [24] and land target detection [33], and has shown a better performance than the traditional methods. One method splits the images into small patches and then uses the pre-trained CNN model to classify the patches, after which the classification results are mapped onto the original images [34]. However, this method has a low target location precision because it does not take the edges of target into consideration. Some other works apply the region-based CNN networks to detect ships in SAR images. The study in Reference [25] adopts the structure of Faster-RCNN and fuses the deep semantic and shallow high-resolution features in both RPN and Region of Interest (RoI) layers, improving the detection performance for small-sized ships. Kang et al. used the Faster-RCNN to carry out the detection task and employed the CFAR method to pick up small targets [35]. While the modifications to Faster-RCNN could help detect small ships, they introduce false alarms to the detection results. Furthermore, researchers in Reference [36] applied SSD algorithms to ship detection in SAR images. Apart from comparing the performance of different SSD models, almost no changes are introduced to the SSD structure to improve the performance in terms of marine target detection. To sum up, the previous studies demonstrate that the CNN-based methods can detect the marine targets more accurately than the CFAR methods and feature-based methods. However, among all the detection methods introduced above, they only focus on ship detection in SAR images and are unable to detect other marine targets with classifying the targets at the same time. Moreover, the existing CNN-based methods generate some false alarms and miss the targets due to the complex sea background. To solve the existing problems, a multi-resolution SSD model is proposed to detect marine targets with classification in this work.

Overall, this paper builds a novel CNN structure to recognize eight types of marine targets at patch level and propose an end-to-end algorithm using a modified SSD to realize marine target detection with classification in large-scale SAR images. The main contributions of the work are as follows:

- The Marine Target Classification Dataset (MTCD) including eight types of marine targets and the Marine Target Detection Dataset (MTDD) containing six kinds of targets are built on GF-3 SAR images, which provide a benchmark for future study. The features of various targets are analyzed based on their scattering characteristics to generate the ground truths.
- A novel CNN structure with six convolutional layers and three max pooling layers is developed to classify different marine targets in SAR images, whose performance is superior to the existing methods.
- A modified SSD with multi-resolution input is proposed to detect different targets. This is the first study for detection of different types of marine targets instead of only detecting ships in SAR images, to the best of our knowledge. Then, the framework for detecting marine objects in large-scale SAR images is introduced.

The remainder of this paper is organized as follows: In Section 2, feature analysis for different targets and the proposed methods are introduced. Experimental results for target classification and

detection in comparison with different existing methods are provided in Section 3. Section 4 discusses the results of the proposed methods. Finally, Section 5 concludes this paper.

2. Methods

2.1. Preprocessing of GF-3 Images

The original GF-3 images used in this paper are large-scale single look complex (SLC) images containing many targets, which means it is impossible to use them directly. In this subsection, the preprocessing method is proposed to extract the targets patches automatically and efficiently.

Firstly, the SLC images are transformed into amplitude images using the following formula:

$$O = |S| \tag{1}$$

where S is the SLC image while O represents the amplitude image.

Then, a non-linear normalization is applied to the images using Equation (2)

$$X(i,j) = \begin{cases} 1 & O(i,j) > T \\ \frac{O(i,j)}{T} & O(i,j) \leq T \end{cases} \tag{2}$$

where T is a constant, $O(i,j)$ is the value of the normalized image at (i,j). The constant T works as a threshold and its value depends on the image. Usually we set $T = 10\overline{O}$, where $\overline{O}$ represents the average value of the image.

As it is time-consuming to select the target patches manually, an image segmentation method is proposed here to extract the target patches in the large-scale SAR images. The Otsu method is an effective algorithm to image segmentation, which searches for a threshold that minimizes the intra-class variance [37].

The proposed method uses the Otsu method to binarize the SAR images, after which the target candidates form the isolated points in the binary images because their pixel values are higher than the Otsu threshold, while the pixel values of sea clutter is lower than the threshold. Then, the algorithm searches for the isolated points and extract the coordinates. Finally, the fixed-size slices are collected according to the coordinates.

2.2. Feature Analysis

In this paper, eight types of maritime targets: Boat, cargo ship, container ship, tanker ship, cage, iron tower, platform, and windmill are selected and studied. Due to the lack of ground truths of the targets in SAR images, the scattering characteristics of each kind of targets are analyzed to get the label for each target. Figure 1 presents the eight types of targets in both optical and SAR images. This is the first such explicit analysis on eight types of marine targets to the best of our knowledge.

Boat: Boats have the simplest and smallest structures among the eight targets. As for boats, the hull edges and the engines at the tail generate strong backscattering, which leads to a closed ellipse in SAR images.

Cargo: Due to the existence of several warehouses, there is a strong secondary reflection on the walls of each warehouse, which is where the rectangular shapes come from in the SAR images.

Container ship: Container ships possess the largest hull. When the ships are fully loaded with containers, the container exteriors would produce strong secondary reflections, resulting in the effect of a washboard in the radar images. In addition, the strong reflections at the tail of the ship come from the complex structure of the ship tower.

Oil tanker: In order to transport oil, an oil pipeline is installed in the middle of the tanker. This causes a bright line in the middle of the tanker in the SAR images. Besides, the closed ellipse in the SAR image comes from the hull edges of the ship.

Cage: Cages used in marine aquaculture are concentrated in square grids in offshore areas. The edges of cages can provide strong backscattering, which forms dotted rectangular distribution in the radar image.

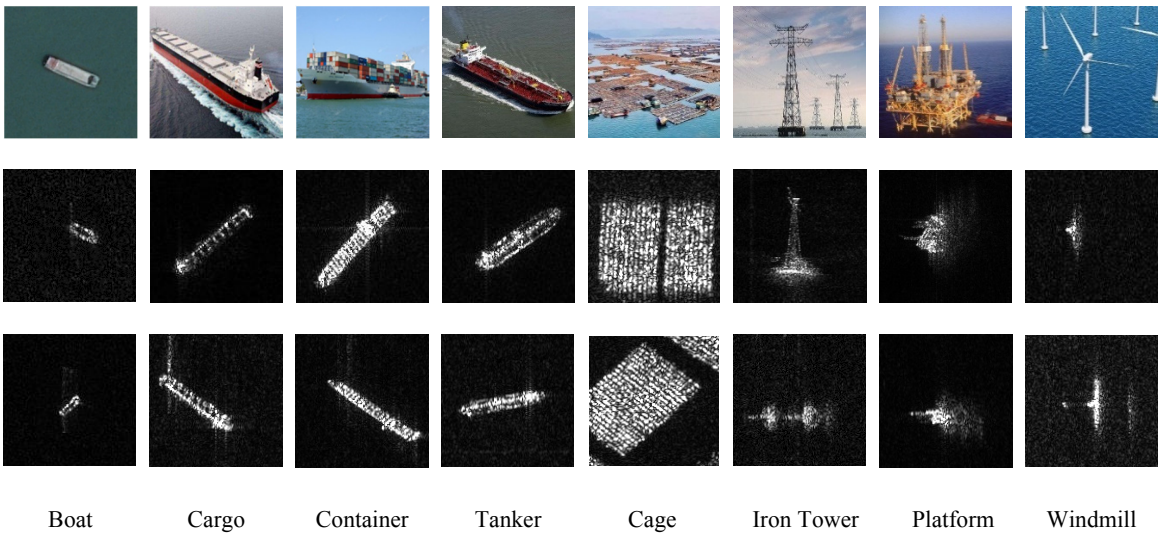


Figure 1. Samples of different maritime targets in optical and SAR images.

Iron tower: When the incident angle is small, the complex structures lead to a strong scattering point. However, when the incident angle is large, the tower target appears as a cone-shaped structure in the radar image. In addition, the transmission lines on the tower produce relatively weak reflection, like a gloomy strip in the image.

Platform: Many offshore countries have built drilling platforms to exploit oil and gas. Usually, they contain support structures, pipelines, and additional combustion towers. The pipelines and combustion towers of the platform result in bright lines, while the support structures produce massive bright spots in SAR images.

Windmill: The strong scatterings of turbines in windmills result in a bright spot. In addition, as the fans rotate, they also produce a bright line that gradually fades toward both ends.

2.3. Marine Target Classification Model Based on CNN

While earlier studies have proposed some CNN models with different structures to classify marine targets, they are employed over simple structures or inappropriate layer arrangements when dealing with small datasets, making it hard to extract distinctive features for marine targets.

In order to solve existing problems, we proposed a CNN structure with six convolutional layers (Conv.1–Conv.6), three max-pooling layers (Pooling1–Pooling3) and two full connection layers, which is shown in Figure 2. It can be seen that a pyramid structure is adopted, and as the CNN goes deeper, the outputs of each layer are down-sampled by pooling layers and the channel of the feature maps increases at the same time. This kind of structure can extract both low level and high level features.

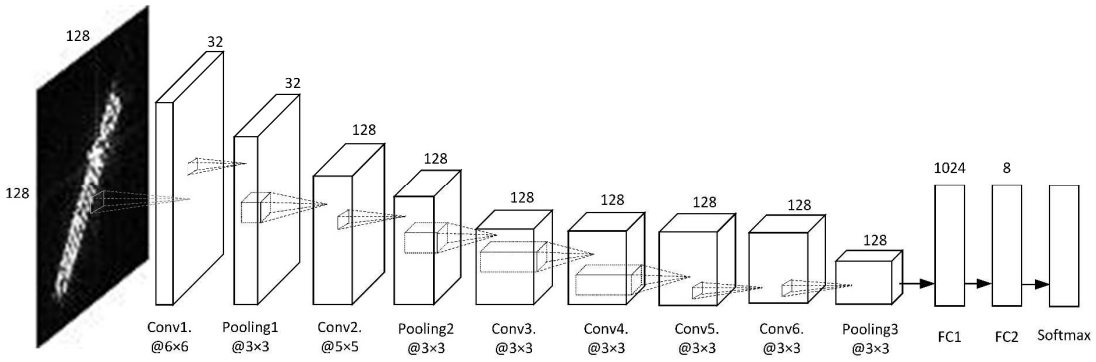


Figure 2. The structure of the proposed classification model.

As the length and width of the marine targets in this study is smaller than 100 pixels, the size of the input patches is set to 128 × 128 pixels to accommodate the objects. At the beginning, 32 convolutional kernels of size 6 × 6 work on the input images to extract features, after which the outputs are down sampled by max-pooling kernels with a size of 3 × 3. Then, the second convolutional layer filters the outputs of the first pooling layer with 128 kernels of size 5 × 5. After that, the convolutional layers are down sampled by the second pooling layer to shrink the feature maps. Then, the CNN network goes deeper with four convolutional layers employing 128 convolutional kernels of size 3 × 3, to generate high-level features, which are transmitted to the third max pooling layer. Finally, two fully-connected layers (FC1 with 1024 output neurons and FC2 with 8 output neurons) take the outputs of the third pooling layers as input and then output the vector to the softmax function to predict the labels of the targets. The strides of all the convolutional layers and all the pooling layers are set to one and two, respectively. Furthermore, the Rectified Linear Units (ReLU) are used for every convolutional layers and full-connected layers to prevent vanishing gradient or exploding gradient.

The training objective is to minimize the cross entropy loss function by forward propagation algorithm and error backpropagation algorithm, which can be written as follows:

$$L(w) = -\frac{1}{m}\sum_{i=1}^{m} \log P(y^{(i)} \Big| x^{(i)}; w) \tag{3}$$

where m represents the total number to training examples and $y^{(i)}$ and $x^{(i)}$ refer to the true label and predicted label of the ith example, respectively. w is the trainable parameter and a regularization term $\lambda \|w\|^2$ is added to the loss function to prevent overfitting, where λ is the regularization factor.

2.4. A Modified SSD Network for Marine Target Detection

Figure 3 shows the proposed Multi-Resolution Single-Shot Multibox Detector (MR-SSD), which has three parts: The first part is multi-resolution image generation, the second part is a standard CNN architecture used for image classification, and the last part is the auxiliary structure containing multi-scale feature maps, convolutional predictors, and default boxes with different aspect ratios. The MR-SSD is capable of extracting features from different resolution images at the same time, which helps to increase the detection precision.

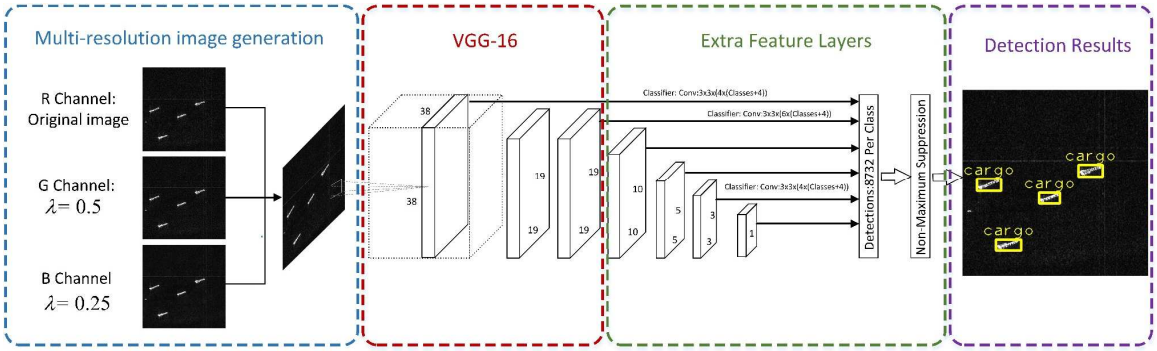


Figure 3. Structure of the MR-SSD.

The input images of traditional SSD have three channels: R, G, and B channels, while the original SAR images only have one channel. Previous practices usually put the same image into the three channels, which causes redundancy in computation and ignore the effects of resolution versions. In this part, a multi-resolution input procedure is designed by adopting images with different resolutions in different channels to extract more features than the traditional SSD. The size of the input images of the MR-SSD is set to 300 × 300. As for the multi-resolution generation part, the images are transformed to the frequency domain using the 2-D Fourier Transformation, and then low-pass filter is used to lower the ground resolution while keeping the image size fixed, described as follows:

$$H(u,v) = \begin{cases} 1 & \left|u - \frac{M}{2}\right| \le \lambda B_a \text{ and } \left|v - \frac{N}{2}\right| \le \lambda B_r \\ 0 & Otherwise \end{cases} \tag{4}$$

where B_a and B_r represent the $M \times N$ image's bandwidth in azimuth direction and range direction, respectively. It can be seen that λ is the factor determining the cutoff bandwidth of the filtered images, which is set to 0–1. After that the filtered images in the frequency domain are transformed to the time domain via inverse Fourier Transformation. Finally, the image ground resolution is reduced because of the linear relationship between ground resolution and SAR image bandwidth. For the proposed MR-SSD, M and N are set to 500. Filters with $\lambda = 0.5$ and $\lambda = 0.25$ are used to reduce the image resolution, and the images are transmitted to the G channel and B channel, respectively.

The second part of the MR-SSD is a standard CNN architecture, i.e., VGG-16 [38], including five groups of convolutional layers combined with ReLU and pooling layers. Different from the VGG-16, the last two fully connected layers are replaced with two convolutional layers to extract features.

The extra feature layers allow the detection at multiple-scales. In this part, we adopt the corresponding parameters used in SSD [31], which proves to be effective in object detection challenges. The extra features layers generate default boxes on each feature map cells with different aspect ratios and then many convolutional filters are used to filter the default boxes to get the class score and offsets. Suppose there are r feature maps in the MR-SSD, the scale of default boxes on different feature maps is defined as follows:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{r-1}(k-1) \quad k \in [1, r] \tag{5}$$

where $s_{\min} = 0.2$, $s_{\max} = 0.9$, S_k is the scale of kth feature map. The aspect ratios for default boxes are denoted as $a_r \in \{1, 2, 3, 1/2, 1/3\}$. Then, the width ($w_k^a$) and height ($h_k^a$) can be calculated by:

$$w_k^a = s_k\sqrt{a_r} \quad h_k^a = s_k/\sqrt{a_r} \tag{6}$$

As for $a_r = 1$, a default box with the scale of $s'_k = \sqrt{s_k s_{k+1}}$ is added. As a result, 6 default boxes on each feature map cell are generated and the number of filters for a $m \times n$ feature map is $6 \times m \times n \times (c+4)$, in which c is the number of class categories and 4 corresponds to the four offsets.

After that, the total number of the default boxes per class is 8732, and non-maximum suppression (NMS) is used to improve the performance of MR-SSD.

When the MR-SSD is trained, it is necessary to determine whether the default box corresponds to a ground truth box or not. For every ground truth box, the default boxes with overlapping rate higher than a threshold (0.5) are selected to match the ground truth boxes. We minimize the loss function as SSD, which is written in Equation (7),

$$L(x,c,l,g) = -\frac{1}{N}(L_{conf}(x,c) + L_{loc}(x,l,g)) \tag{7}$$

where N is the number of matched default boxes, $L_{conf}(x,c)$ and $L_{loc}(x,l,g)$ are the confidence loss and the localization loss, respectively. The confidence loss function employs softmax loss over multiple classes confidences, which is:

$$L_{conf}(x,c) = -\sum_{p}\sum_{i\in Pos}^{N} x_{i,j}^{p}\log(\hat{c}_i^p) - \sum_{i\in Neg}\log(\hat{c}_i^0) \quad \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \tag{8}$$

where $x_{i,j}^{p}$ is an indicator for matching the ith default box to the jth ground truth box of category p. If the two boxes are matched, the indictor will be set to 1, otherwise it will be set to 0. c_i^p represents the confidence of the ith default box of category p. The localization loss uses the Smooth L1 loss between the proposed box (l) and the ground truth box (g) parameters, defined as:

$$L_{loc}(x,l,g) = -\sum_{i\in Pos}^{N}\sum_{m\in\{cx,cy,w,h\}} x_{i,j}^{k} smooth_{L1}(l_i^m - \hat{g}_j^m) \tag{9}$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & if|x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \tag{10}$$

The offsets for the center (cx;cy), width (w) and height (h) of the default box (d) are regressed by the following formulas:

$$\begin{array}{cc} \hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx})/d_i^w & \hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy})/d_i^h \\ \hat{g}_j^{w} = \log\left(\frac{g_j^w}{d_i^w}\right) & \hat{g}_j^{h} = \log\left(\frac{g_j^h}{d_i^h}\right) \end{array} \tag{11}$$

2.5. The Whole Workflow for Marine Target Detection in Large-scale SAR Images

However, target detection in large-scale SAR images (larger than 10,000 × 10,000 pixels) is difficult because some images cover buildings and islands, which would lead to false alarms. Moreover, the CNN based methods can only detect targets at patch level due to the fixed input size. If the large-scale images are resized to a patch size for target detection, it will lose many detail features, making it hard to detect small targets. In order to solve the existing problems, this paper proposes a whole workflow consisting of sea-land segmentation, cropping with overlapping, detection with pre-trained MR-SSD, coordinates mapping and predicted boxes consolidation for marine target detection in large-scale SAR images. It is able to rule out the false alarms on lands, reduce overlapping predicted boxes, and generate accurate coordinates for each marine target. Figure 4 illustrates the whole procedure in detail.

The whole workflow is divided into two processes: Training process and detection process. As for training process, the patches including marine targets are extracted from SAR images to build the training set and then train the MR-SSD model.

The other one is the detection process for large-scale SAR images. In order to reduce the false alarms on lands, the level-set method [39] is used to remove land parts, which proves to be effective

in image segmentation. Due to the high computation complexity of the level-set method, the images are down-sampled and then the level–set method is employed to generate the land masks, which will be resized to the original scale by interpolation later. After that, the land mask removes all the land objectives.

Usually, the large-scale images cannot feed the MR-SSD model directly because it resizes the large-scale images into 300×300, which means that a large number of small targets are hard to be detected. In order to solve this problem, the large-scale images are cropped into overlapping small patches and then the patches are sent to MR-SSD. The purpose of overlapping is to keep the target intact in at least one patch. Given a large-scale SAR image of size $Lw \times Lh$, the total number of patches is $m \times n$, which can be calculated as follows:

$$m = \left\lceil \frac{Lw}{Pw - Overlap} \right\rceil \tag{12}$$

$$n = \left\lceil \frac{Lh}{Ph - Overlap} \right\rceil \tag{13}$$

where Pw and Ph denotes the width and length of the patches, respectively. In addition, $Overlap$ is the overlap distance between the patches, which can be adjusted according to the image ground resolution.

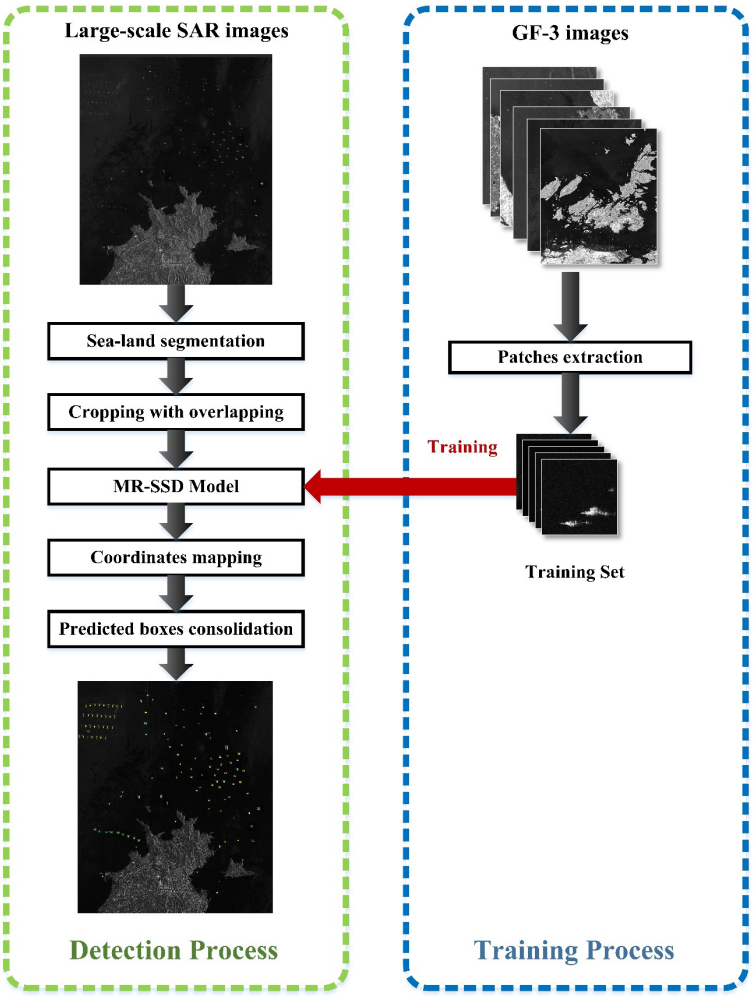


Figure 4. The whole workflow of marine target detection in large-scale images.

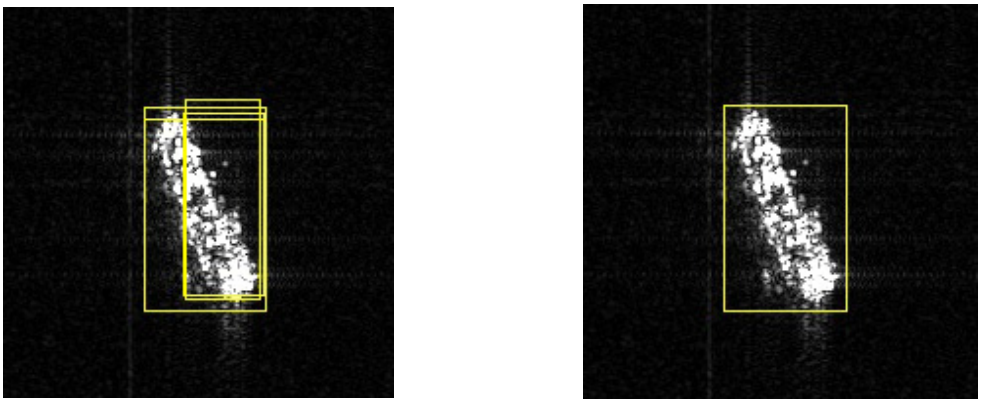
Then, the patches have to be resized to 300 × 300 to meet the input requirements of MR-SSD. The pre-trained MR-SSD model extract deep features of the objectives to generate targets labels and coordinates for each patch later.

With the preliminary detection results, the coordinates on small patches are projected onto the large-scale images and the final detection results are obtained. For a patch whose index in width is ith and index in height is jth, the coordinate of its kth target can be written as $(x_{i,j}^{(k)}, y_{i,j}^{(k)})$. The mapping relationship can be calculated as follows:

$$\begin{aligned} X^{(l)} &= (i-1) * (Lw - Overlap) + x_{i,j}^{(k)} \\ Y^{(l)} &= (j-1) * (Lh - Overlap) + y_{i,j}^{(k)} \end{aligned} \tag{14}$$

where $X^{(l)}$ and $Y^{(l)}$ are the coordinates of the lth target in two directions in the large-scale SAR images.

However, cropping the SAR images would split the targets in two or more pieces, leading to fragmentary predicted boxes and the overlapping operation could cause overlapped predicted boxes, which can be seen in Figure 5a. In order to solve the problems, we consolidate the overlapping and fragmentary predicted boxes by searching the box coordinates to find a coordinates group forming the largest box. As a result, the consolidated box is considered as the final predicted box shown in Figure 5b.



(a) (b)

Figure 5. Procedure for consolidating the predicted boxes. (**a**) predicted boxes before consolidation; and (**b**) predicted box after consolidation.

3. Experimental Results

3.1. Materials

In this paper, a total of 111 VHR spaceborne SAR images generated by the Chinese GF-3 satellite are used, which carries a Band C radar sensor working at 12 imaging modes with a wide variety of ground resolutions. In order to perform target classification and detection, two datasets: Marine Target Classification Dataset (MTCD) and Marine Target Detection Dataset (MTDD),which are built, respectively, using the preprocessing method given in Section 2.1. In the following, we first present the details of the 111 large-scale SAR SLC images and then describe the compositions of the MTCD and MTDD.

3.1.1. GF-3 SLC Dataset

111 GF-3 SAR images covering the offshore areas of Eastern Asia, Western Asia, Western Europe, and Northern Africa are selected. All of them are images of Band C, acquired from December 2016 to May 2018. There are four polarization mode images (51% for HH mode, 27% for HV mode, 9% for VH

mode, and 13% for VV mode), with ground resolution from 0.5 m to 5 m. Table 1 shows the details of the SAR images used in this paper.

Table 1. Composition of the GF-3 SLC dataset.

	HH	HV	VH	VV	Total
0.5 m	2	0	0	4	6
1.7 m	24	0	0	5	29
2.5 m	0	0	7	3	10
3.0 m	11	11	2	2	26
5.0 m	20	19	1	0	40
Total	57	30	10	14	111

3.1.2. Marine Target Classification Dataset (MTCD)

The MTCD is built on the patches captured from the GF-3 SAR SLC dataset and it consists of eight types of maritime targets: Boat, cargo, container ship, tanker, tower, platform, cage, and windmill. Each target chip includes only one type of target and the ground truth is acquired by feature analysis introduced in Section 2.2. The MTCD contains 2522 training samples and 688 testing samples, whose size is fixed to 128 $\times$ 128 pixels. Table 2 lists the numbers of patches per class available for training and testing. In our experiments, the training patches are flipped up-to down to achieve data augmentation, which means a total of 5044 patches are used as training sets.

Table 2. Composition of the marine target classification dataset.

	Boat	Cage	Cargo	Container	Tower	Platform	Tanker	Windmill	Total
Training	390	295	400	200	290	280	312	355	2522
Testing	104	94	154	54	72	55	76	94	688

3.1.3. Marine Target Detection Dataset (MTDD)

The MTDD dataset is built following the PASCAL VOC format [32], containing the slices with corresponding xml files providing the label as well as the location of the target. In this task, six types of targets, i.e., cargo, container ship, tower, platform, tanker, and windmill, are studied because they are more common and valuable than the other targets such as boat and cage. The slices consisting of more than one targets are set to 500 $\times$ 500 and Table 3. presents the composition of the MTDD, including 1727 patches in total.

Table 3. Composition of the marine target detection dataset.

Training Set	Validation Set	Testing Set	Total
549	525	653	1727

3.2. Classification Results of MT-CNN

The experiments are performed on the Caffe [40] framework on Ubuntu 16.04 system, using NVIDIA GeForce GTX 1060 with Max-Q Design acceleration graphics. The training and testing batches are set to 48 and 24, respectively. The network is trained for 60,000 iterations using SGD random gradient descending method with initial learning rate of 0.002 and momentum of 0.9. In addition, the training process takes 1451.22 s, while the testing process takes 153.48 s.

Table 4 gives the confusion matrix of the classification result on the test dataset consisting of eight marine classes. Each row in the table denotes the actual target class, while each column represents the class predicted by MT-CNN. It can be seen that the overall accuracy (OA) achieves 95.20%. Due to the distinctive features of cage and tower, their accuracies reach 100%. However, tanker possesses

the lowest accuracy (88.16%) among the eight classes. As for the low resolution images, the bright lines caused by the pipelines in tankers would be merged by the reflections of hulls, making it hard to discriminate the tankers from other kinds of targets. Interestingly, two cargos are predicted as tankers, while six tankers are predicted as cargos, implying that the two classes share the similar features.

Table 4. Confusion matrix of 8-class classification results of MT-CNN.

	Boat	**Cargo**	**Container**	**Tanker**	**Cage**	**Tower**	**Platform**	**Windmill**	**Total**	**OA**
Boat	98	3	0	1	1	1	0	0	104	94.23
Cargo	3	145	3	2	1	0	0	0	154	94.16
Container	0	3	51	1	0	0	0	0	54	94.44
Tanker	1	6	1	67	1	0	0	0	76	88.16
Cage	0	0	0	0	79	0	0	0	79	100
Tower	0	0	0	0	0	72	0	0	72	100
Platform	1	0	0	2	2	0	50	0	55	90.91
Windmill	1	0	0	0	0	0	0	93	94	98.94
Total	104	157	55	73	84	73	50	93	688	95.20

3.3. Effectiveness of MT-CNN

This subsection compares the proposed MT-CNN with previous methods including CNN based methods and traditional machine learning methods such as SVM and KNN. As for the CNN based methods, three typical CNN networks, i.e., CNN-CB [19], ConvNet [13], and CNN-ML [18], are selected. The CNN-CB is simple constructed with two convolutional layers, two max-pooling layers as well as fully connected layers, while the ConvNet is unique for its lack of fully connected layers. Furthermore, the CNN- ML using a multi-looks input proves to be effective. In addition, We compute the Gist feature of each slices following the procedure in Reference [41] and then train the SVM(RBF-kernel, gamma = 0.5, C = 50) and KNN to classify them. In this subsection, the KNN algorithm employs KD Trees and the number of neighbors and leaf size are set to five and 30, respectively.

Table 5 illustrates the classification accuracies of different methods among the eight categories. It can be noticed that the proposed method outperforms other methods in every category except platform, with the average accuracy achieving 95.20%. CNN-CB and ConvNet can only classify the targets with the overall accuracy of 80.96% and 82.27%, respectively, due to their lack of enough convolutional layers and insufficient convolutional kernels to extract high-level features. While the CNN-ML is able to classify platform more accurately than the proposed MT-CNN does, its performance on other categories is poorer than MT-CNN. As for the traditional machine learning methods such as SVM and KNN, the accuracy of different classes varies a lot. They expert on classifying the targets with distinct characteristics, i.e., boat and windmill, while their performance on other classes are much poorer. Overall, the performance of the proposed method is superior to other methods and the predicted results are more reliable than others.

Table 5. Classification accuracies (%) of different existing methods.

	MT-CNN	**CNN-CB [19]**	**ConvNet [13]**	**CNN-ML [18]**	**SVM**	**KNN**
Boat	94.23	88.46	83.65	91.35	79.81	89.45
Cage	100.0	98.73	97.47	98.73	68.09	68.06
Cargo	94.16	77.92	69.48	92.21	70.13	54.26
Container	94.44	75.93	77.78	79.83	38.89	60.61
Tower	100.0	86.11	94.44	95.83	61.11	61.69
Platform	90.91	83.64	87.27	92.73	45.45	62.76
Tanker	88.16	63.16	65.79	73.68	75.00	65.37
Windmill	98.94	74.47	92.55	93.62	97.87	95.39
Average	95.20	80.96	82.27	90.41	71.80	70.58

3.4. Effectiveness of MR-SSD

In this section, the detection experiments are carried out on the Caffe [40] framework via Ubuntu 16.04 systems. MR-SSD is trained on the MTDD training set including six types of marine targets, i.e., cargo, container ship, tanker, tower, platform, and windmill. The MR-SSD network is trained with learning rate of 0.0001 and a weight decay parameter of 0.005 for 160,000 iterations. After that, the trained model is used to detect the marine targets in the testing set. In addition, the confidence threshold is set to 0.5.

The detection results of MR-SSD on testing samples with different backgrounds are shown in Figure 6. For Figure 6a–g, the objects surrounded by sea clutters are detected with accurate coordinates. The proposed model recognizes all the three tanker against the distractions from the small ships and ambiguities in Figure 6h. Moreover, the defocused container ship in Figure 6i is detected, which demonstrates the robustness of the proposed method. In Figure 6i,j that cover the offshore areas, the trained model is capable of extracting the targets coordinates and predicting the labels precisely.

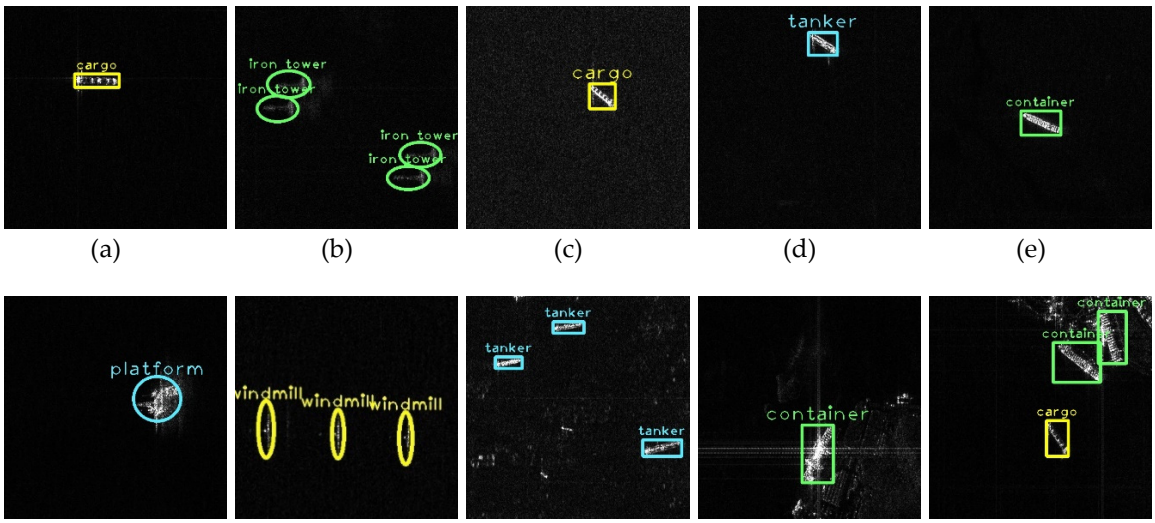


Figure 6. Detection results of the proposed MR-SSD. (**a**) Cargo; (**b**) tower; (**c**) cargo; (**d**) tanker; (**e**) container; (**f**) platform; (**g**) windmill; (**h**) tanker; (**i**) container; (**j**) cargo and container.

As PASCAL VOC challenges, this paper uses average precisions (AP), which is the average of the maximum precisions at different recall values, to access the performance. Recall, precision, and F1 score are defined as follows:

$$\text{Recall} = \frac{T_d}{T_g} \tag{15}$$

$$\text{Precision} = \frac{T_d}{T_d + T_f} \tag{16}$$

$$\text{F1} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{17}$$

where T_d denotes the number of the correctly detected targets, T_g represents the number of ground truths, and T_f indicates the number of false alarms. F1 is the harmonic mean of precision and recall. Besides, the mean Average Precisions (mAP) is used to access the model's ability in detecting all types of targets.

To prove the advantages of the proposed MR-SSD model, existing algorithms (i.e., Faster-RCNN [29] and SSD [31]) are selected for contrast experiments.

Table 6 compares the AP and mAP of different methods. It can be seen that the proposed method achieves 87.38% mAP, which is 5.29% and 1.76% higher than Faster-RCNN and SSD, respectively.

It is evident that MR-SSD has the best AP for every individual category on MTDD. The proposed MR-SSD improves the accuracy for tower significantly, surpassing SSD by 5.52% mAP. While the improvements for other classes are slight, there is less than 2% mAP. Though Faster-RCNN can detect cargo, platform and tanker with more than 85% mAP, its performance in terms of container, tower, and windmill are much worse than that of MR-SSD. The experiments demonstrate that the proposed method can extract more features and detect targets more precisely than the traditional one and can achieve higher performance.

Table 6. Average Precisions (AP) of different algorithms among different targets (%).

Method	Cargo	Container	Tower	Platform	Tanker	Windmill	mAP
Faster-RCNN	89.47	79.78	68.79	89.61	86.70	78.19	82.09
SSD	89.37	87.08	74.55	89.96	86.46	86.34	85.62
MR-SSD	89.77	88.69	80.07	90.43	87.28	88.04	87.38

3.5. Detection Results of Large-Scale SAR Iimages—Case Study

A large-scale SAR image can hardly contain all kinds of marine targets because of the variance of locations of targets, e.g., the windmills are mainly located in open sea while a large number of cargos settle in offshore areas. In order to demonstrate the performance of the proposed method, some types of targets: Windmills, platforms, and towers, are transplanted to the large-scale SAR images (12,000 × 14,000 pixels) covering Weihai City, Shandong Province, China. The imaging mode is Ultlra Fine Strip (UFS), polarization mode is HH and the ground resolution is 1.7 m. The image and the numbers of the ground truths are depicted in Figure 7a and Table 7, respectively.

(a) (b) (c)

Figure 7. Sea-land segmentation results of a large-scale GF-3 image. (**a**) The original image; (**b**) The land mask; and (**c**) The image after land masking.

Table 7. Compositions of the large-scale SAR image.

Category	Cargo	Container	Tower	Platform	Tanker	Windmill	Total
Number	37	8	19	10	23	31	128

3.5.1. Sea-Land Segmentation

In this subsection, the level-set method [39] is employed to sea-land segmentation. The downsampling rate is set to 10 to accelerate computing and the segmentation contour is iterated for 10 times. It takes 43.21 s to generate the land mask and the segmentation results are shown in Figure 7.

It is evident that the land mask wipes out all the land areas precisely, while all of the marine targets remain in the images, which contributes to reducing false alarms and increasing detection precision.

3.5.2. Detection Results of the Whole Workflow

After removing the lands from the images, the image is cropped into 500 × 500 sub-images with overlapping of 200 pixels. Then, the 500 × 500 sub-images are resized to 300 × 300 to match with the input size of MR-SSD. After that, the coordinates in the sub-images are mapped onto the large-scale images.

Figure 8 shows the detection results of the proposed methods. It can be seen that most of the six types of targets can be detected with accurate coordinates. In the large-scale image, a windmill, a tanker, and a platform are missed. The missing windmill and tanker have week intensity, while the missing platform is overlapped by another platform, which diminishes the performance of MR-SSD. Besides, two tankers are misrecognized as container ships because they share similar features: Large hulls and multiple components leading strong reflections. In practice, as the land mask can hardly rule out small reefs near the coastline, some small reefs are transmitted into the MR-SSD. As a result, five reefs are recognized as cargos.

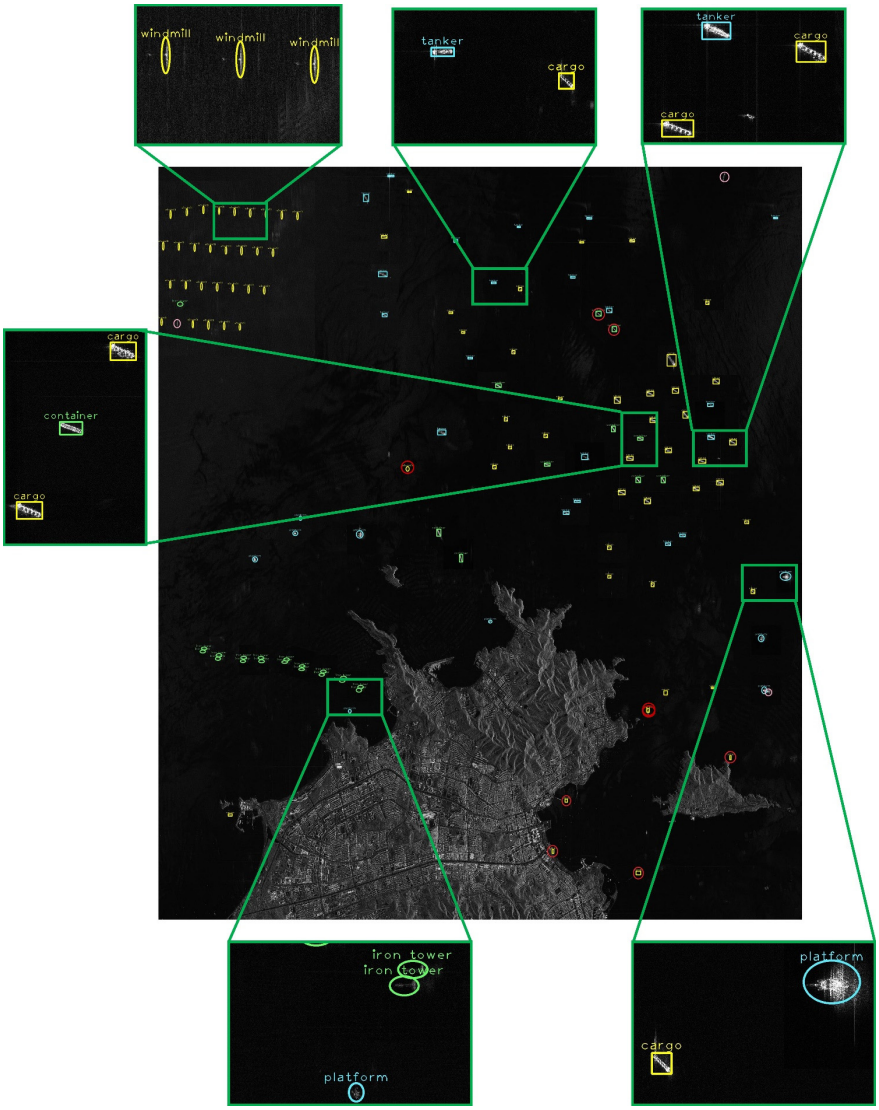


Figure 8. Detection results of a large-scale SAR image. The red circles and pink circles denote the false alarms and missed targets, respectively. Cargo, container ship, and tanker are labelled by yellow, green, and blue rectangles, respectively. Yellow eclipses, green eclipses, and blue eclipses indicate windmills, iron towers, and platforms.

Also, Faster-RCNN [29] and SSD [31] models are employed in the overall scheme to demonstrate the advance of the proposed method. The detection results are recorded in Table 8. The recall, precision, and F1 score are calculated according to the Equations (15)–(17).

Table 8. Detection results of different CNN models on the large-scale SAR image.

Method	T_g	T_d	T_f	Recall (%)	Precision (%)	F1 (%)
Faster-RCNN	128	119	8	92.97	93.70	93.33
SSD	128	121	22	94.53	84.62	89.30
MR-SSD	128	122	8	95.31	93.85	94.57

It can be seen that the MR-SSD gets the highest recall, precision and F1 score among the three methods. Compared with SSD that generates 22 false alarms, the proposed method reduces the number of false alarms, only 8 false alarms exist. Though Faster-RCNN produces the same number of false alarms as t/6he proposed method, its number of the correctly detected targets is less than that of the proposed method, which leads to a lower F1 score. In summary, the proposed method outperforms other methods in detecting different marine targets in large-scale SAR images.

4. Discussion

By comparing and analyzing the results of experiments conducted in our work, the merits of the proposed methods are demonstrated. In this section, we discuss impacts of some parameters on performance of the proposed methods and analyze characteristics of false alarms and missing targets, which helps to improve the performance in the near feature.

4.1. Performance of MT-CNN Trained with Different Data Augmentation Methods

In order to analyze the impacts of data augmentation on the MT-CNN's performance, we use four datasets: Training sets without flipping (TS1), training sets with up-to-down flipping (TS2), training sets with left-to-right flipping (TS3), and training set with up-to-down and left to right flipping (TS4). The experiments are carried out under the same conditions. Table 9 shows the classification results of MT-CNN trained with different augmentation methods. It can be seen that flipping could help to improve the models' performance. However, more flips can hardly improve their performance and this is because this operation cannot provide more information that the models need.

Table 9. Classification accuracies (%) of MT-CNN trained with different augmentation methods.

	Boat	Cage	Cargo	Container	Tower	Platform	Tanker	Windmill	OA
TS1	89.42	98.73	87.66	87.04	97.22	90.91	81.58	97.87	91.13
TS2	94.23	100.0	94.16	94.44	100.0	90.91	88.16	98.94	95.20
TS3	93.27	98.73	92.86	88.89	100.0	98.18	84.21	97.87	94.19
TS4	91.35	98.73	91.56	92.59	100.0	96.36	88.16	97.87	94.19

4.2. Comparison of Performance of Different CNN Structures

In this subsection, we propose four CNN models with different layer arrangements and perform classification experiments on MTCD to demonstrate the merits of MT-CNN. The structures of the CNN models are shown in Table 10 and the parameters of the layers are the same with those of corresponding layers in MT-CNN.

Table 10. Structures of five different CNN models.

Layer	MT-CNN	CNN-A	CNN-B	CNN-C	CNN-D
1	Conv1	Conv1	Conv1	Conv1	Conv1
2	Pooling1	Pooling1	Pooling1	Pooling1	Pooling1
3	Conv2	Conv2	Conv2	Conv2	Conv2
4	Pooling2	Pooling2	Pooling2	Conv3	Pooling2
5	Conv3	Conv3	Conv3	Conv4	Conv3
6	Conv4	Conv4	Conv4	Conv5	Conv4
7	Conv5	Conv5	Conv5	Pooling3	Conv5
8	Conv6	Pooling3	FC1	FC1	Conv6
9	Pooling3	FC1	FC2	FC2	Pooling3
10	FC1	FC2	-	-	FC2
11	FC2	-	-	-	-

Table 11 shows the classification accuracies of different CNN models. While the accuracy of platform of MT-CNN is lower to that of CNN-A, MT-CNN can obtain higher accuracies than the four CNN models in other categories and its overall accuracy achieves 95.20%. In addition, the overall accuracies of the fiver models are all over 90% and increasing or decreasing network layers would have slight impacts on their performance.

Table 11. Classification accuracies (%) of different CNN models.

	Boat	Cage	Cargo	Container	Tower	Platform	Tanker	Windmill	OA
MT-CNN	94.23	100.0	94.16	94.44	100.0	90.91	88.16	98.94	95.20
CNN-A	91.35	100.0	91.56	85.19	100.0	94.55	82.89	98.94	93.16
CNN-B	93.26	98.73	90.26	88.89	100.0	92.73	80.26	97.87	92.73
CNN-C	83.68	98.73	92.86	79.63	95.83	94.55	82.89	95.74	90.84
CNN-D	93.27	100.0	92.21	79.63	98.61	94.55	77.63	97.87	92.30

4.3. Class Imbalance Effect

Among MTCD, there are a few big classes (i.e., cargo and boat) and small classes (i.e., container ship and platform). In order to discuss the class imbalance effect on the MT-CNN's performance, we use two balancing methods to build two balanced dataset: BAL1 and BAL2. As the smallest class (container ship) in MTCD has 200 patches, we reduce the number of slices to 200 in other classes to form BAL1. BAL2 augments small classes in MTCD by left to right flipping and each class contains 400 slices. In the experiment, all of the slices in the three datasets are flipped up-to down to realize data augmentation.

Table 12 compares classification accuracies of MT-CNN trained in the three datasets. As for cargo, which accounts for the largest proportion in MTCD, its accuracy drops when the dataset is balanced. One possible reason is that the MT-CNN tends to extract specific features in other categories as cargo's proportion in the datasets declines. However, platform shows the opposite trend, with the accuracy rising by 2% and 4% in BAL1 and BAL2, respectively. This is because its proportions in BAL1 and BAL2 are higher than that in MTCD. Among other classes, there is not a significant imbalance effect because MTCD is not a serious imbalanced dataset.

Table 12. Comparison of classification accuracies (%) of MT-CNN trained by MTCD, BAL1, and BAL2.

	Boat	Cage	Cargo	Container	Tower	Platform	Tanker	Windmill	OA
MTCD	94.23	100.0	94.16	94.44	100.0	90.19	88.16	98.94	95.20
BAL1	85.58	100.0	85.06	90.74	100.0	92.73	81.57	100.0	91.13
BAL2	96.15	100.0	91.56	92.59	100.0	94.55	82.89	98.94	94.48

4.4. Performance of MT-CNN against Ground Resolution Variance

To evaluate the performance brought by resolution variance in the proposed MT-CNN, extensive experiments using different resolutions images are further conducted. Test dataset including eight types of target slices at 1.7 m ground resolution is built. Then, lower pass filters are used to lower image resolution to generate target slices with eight resolution versions. Table 13 shows the composition of the test dataset. In the experiment, the proposed network is fed with target slices with different resolutions.

Table 13. Composition of the 1.7 m ground resolution target classification dataset.

	Boat	Cage	Cargo	Container	Tower	Platform	Tanker	Windmill	Total
Number	28	35	50	30	48	34	30	45	291

Figure 9 illustrates the robustness of the proposed MT-CNN against ground resolutions. We can see that there is a slight increase in average accuracy from 1.7 m to 3.4 m and then it decreases gradually from 97% at 3.4 m to 85% at 13.6 m. Windmill and cargo keeps at 100% when images ground resolution varies from 1.7 m to 13.6 m. Tanker, container ship, and cage are more sensitive to resolution variance than other kinds of targets. Tanker declines dramatically from 95% at 1.7 m to 57% at 6.8 m and then it remains stable from 6.8 m to 13.6 m. One possible reason is that the auxiliary structures such as pipelines and cranes on tankers could be blurred in low resolution images, making tankers lose distinctive features. Additionally, cages drop rapidly from 6.8 m to 11.9 m because they share some rectangle-like shapes with platforms and many cages are misclassified into platforms.

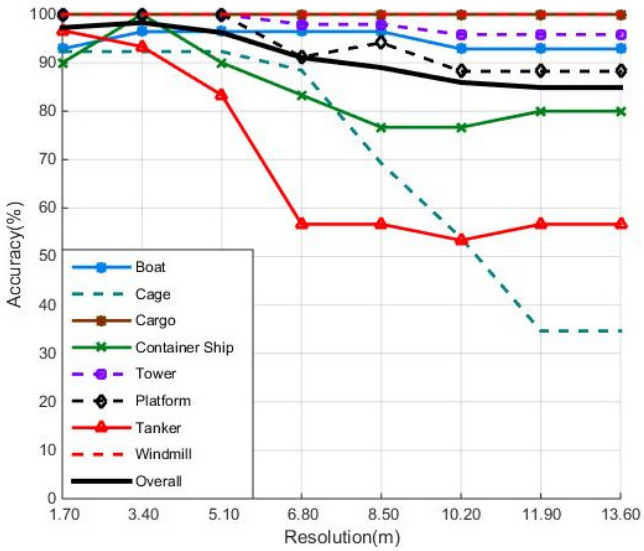


Figure 9. Classification accuracies of MT-CNN under different ground resolutions.

4.5. Comparison of Performance of MR-SSD with Different Low-Pass Filters

In this subsection, we adopt different values for λ in G channel and B channel to analyze its influence on the performance of MR-SSD. All the MR-SSD models are trained on the CAFFE framework and the experimental parameters are the same with those in Section 3.4. Table 14 shows the performance of MR-SSD with different low-pass filters. It can be seen that as λ varies, the mAP of MR-SSD changes slightly, and it achieves the highest mAP (87.38%) when λ is set to 0.5 and 0.25 for G channel and B channel, respectively.

Table 14. Comparison of precisions (%) of MR-SSD with different low-pass filters.

G	B	Cargo	Container	Tower	Platform	Tanker	Windmill	mAP
0.2	0.4	88.11	81.24	71.31	90.27	84.06	86.47	83.58
0.2	0.6	88.22	84.70	71.36	90.27	83.97	87.78	84.39
0.2	0.8	88.41	84.33	74.73	90.29	86.05	88.40	85.37
0.4	0.6	88.69	82.99	70.90	90.11	86.20	84.89	83.96
0.4	0.8	89.03	83.27	78.74	90.75	87.23	69.15	83.03
0.6	0.8	89.69	87.90	79.15	90.29	87.67	87.54	86.79
0.5	0.25	89.77	88.69	80.07	90.43	87.28	88.04	87.38

4.6. Influence of Different Patch Sizes in the Proposed Workflow

As for the proposed workflow, large-scale SAR images are cropped into different slices, which are then sent to the pre-trained MR-SSD and the impacts of slice size are discussed in this subsection. We carry out experiments using the large-scale SAR image provided in Section 3.5 and performance of the proposed method in terms of patch size are compared in Table 15. It can be noticed that the computational time drops dramatically when patch size increases, because the patch size determines the total number of patches. Recall is relatively high when patch size is under 700 × 700 but it drops dramatically from 89.84% at 700 × 700 to 64.06% at 900 × 900, because the resize operation removes many image details, leading to many missing targets. In practice, the cropping size should be carefully considered and keep a balance between computational time cost and F1 score.

Table 15. Results on the large-scale SAR image in terms of patch size.

$Pw \times Ph$	T_g	T_d	T_f	Recall (%)	Precision (%)	F1 (%)	Time (s)
300 × 300	128	112	13	87.50	89.60	88.54	1268.67
500 × 500	128	122	8	95.31	93.85	94.57	636.56
700 × 700	128	115	8	89.84	93.50	91.12	259.94
900 × 900	128	82	2	64.06	97.62	77.36	145.20

4.7. False Alarms and Missing Targets in the Large-Scale Images

Some typical patches containing false alarms are displayed in the blue box, while missing targets are shown in the red box in Figure 10. It can be seen small reefs are easy to be recognized as cargos because they are brighter than the sea clutters and share similar visual features with cargos. For images without geocoding, it is difficult to remove all the reefs precisely. Interestingly, some dams are classified as cargo or container ship. One possible reason is that dams lead to the bright lines similar to that produced by warehouses or containers. Additionally, a cargo is recognized as platform in Figure 10d because it possesses a rectangular contour with high intensity, which looks like a platform visually. As for the platform in Figure 10f, the blurring in image looks like burning towers on the platform, which is the main reason for such misclassification. Besides, some coastlines are classified as cargos because of their bright lines in SAR images.

When it comes to missing targets, some of them have small or weak intensity, which leads to little response in the network, remaining to be undetected. The strong noise and motion blurring in Figure 10k,m exert adverse effects on target detection.

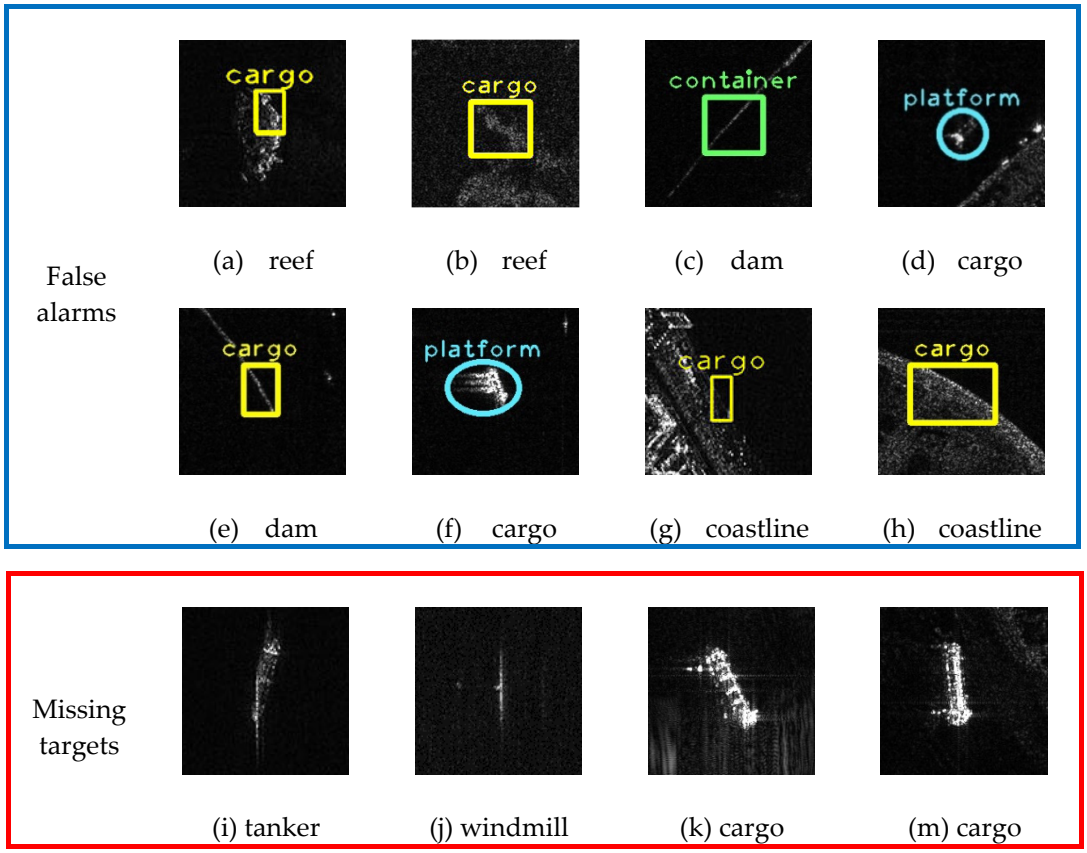


Figure 10. Samples of false alarms and missing targets. The ground truths of the patches:(**a**) Reef; (**b**) reef; (**c**) dam; (**d**) cargo; (**e**) dam; (**f**) cargo; (**g**) coastline; (**h**) coastline; (**i**) tanker; (**j**) windmill; (**k**) cargo; (**m**) cargo.

5. Conclusions

With the labeled SAR images provided by the GF-3 satellites, this paper proposes a convolutional network (MT-CNN) to classify marine targets at patch level and an overall scheme to detect different marine targets in large-scale SAR images. The proposed MT-CNN with six convolutional layers and three pooling layers are capable of extracting features at different levels and achieve higher classification accuracy than existing CNN models. As for the marine target detection task in large-scale SAR images, the proposed MR-SSD with a three-resolution input is able to learn the features on different resolution versions. The proposed framework containing sea-land segmentation, cropping with overlapping, detection with MR-SSD model, and coordinates mapping shows its superiorities to other methods by improving detection accuracy and reducing false alarms. Besides, this is the first such experiments that carries out on such various types of marine targets in SAR images. This paper presents the preliminary results of the proposed methods. Looking ahead, future works can be focused on eliminating false alarms in SAR imageries by image processing methods.

Author Contributions: Conceptualization, J.C. and W.Y.; Data curation, M.M. and W.Y.; Funding acquisition, J.C. and W.Y.; Investigation, M.M. and W.L.; Methodology, M.M. and J.C.; Resources, J.C. and W.Y.; Software, M. M.; Supervision, J.C. and W.Y.; Validation, M.M.; Writing—original draft, M.M.; Writing—review & editing, J.C., W.L. and W.Y.

Funding: This research received no external funding.

Acknowledgments: This work is supported by the National Science Foundation of China (NSFC) under Grant No.61671043, and the NSFC under Grant No. 61701012.

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

References

1. Margarit, G.; Mallorqui, J.J.; Rius, J.M.; Sanz-Marcos, J. On the Usage of GRECOSAR, an Orbital Polarimetric SAR Simulator of Complex Targets, to Vessel Classification Studies. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3517–3526. [CrossRef]
2. Leng, X.; Ji, K.; Zhou, S.; Xing, X.; Zou, H. 2D comb feature for analysis of ship classification in high-resolution SAR imagery. *Electron. Lett.* **2017**, *53*, 500–502. [CrossRef]
3. Xing, X.; Ji, K.; Zou, H.; Chen, W.; Sun, J. Ship Classification in TerraSAR-X Images With Feature Space Based Sparse Representation. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1562–1566. [CrossRef]
4. Ji, K.; Xing, X.; Chen, W.; Zou, H.; Chen, J. Ship classification in TerraSAR-X SAR images based on classifier combination. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, 21–26 July 2013; pp. 2589–2592.
5. Lin, H.; Song, S.; Yang, J. Ship Classification Based on MSHOG Feature and Task-Driven Dictionary Learning with Structured Incoherent Constraints in SAR Images. *Remote Sens.* **2018**, *10*, 190. [CrossRef]
6. Zhang, H.; Tian, X.; Wang, C.; Wu, F.; Zhang, B. Merchant Vessel Classification Based on Scattering Component Analysis for COSMO-SkyMed SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1275–1279. [CrossRef]
7. Lang, H.; Wu, S.; Xu, Y. Ship Classification in SAR Images Improved by AIS Knowledge Transfer. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 439–443. [CrossRef]
8. Srinivas, U.; Monga, V.; Raj, R.G. Meta-classifiers for exploiting feature dependencies in automatic target recognition. In Proceedings of the 2011 IEEE Radar Conference, Kansas City, MO, USA, 23–27 May 2011; pp. 147–151.
9. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Istanbul, Turkey, 9–12 November 2015; pp. 1097–1105.
10. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
12. Chen, S.; Wang, H. SAR target recognition based on deep learning. In Proceedings of the 2014 International Conference on Data Science and Advanced Analytics, Montreal, QC, Canada, 17–19 October 2016; pp. 541–547.
13. Chen, S.; Wang, H.; Xu, F.; Jin, Y.Q. Target Classification Using the Deep Convolutional Networks for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4806–4817. [CrossRef]
14. Gao, F.; Huang, T.; Wang, J.; Sun, J.; Yang, E.; Hussain, A. Combining Deep Convolutional Neural Network and SVM to SAR Image Target Recognition. In Proceedings of the 2017 IEEE International Conference on Internet of Things, Exeter, UK, 21–23 June 2017; pp. 1082–1085.
15. Pei, J.; Huang, Y.; Huo, W.; Zhang, Y.; Yang, J.; Yeo, T.S. SAR Automatic Target Recognition Based on Multiview Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2196–2210. [CrossRef]
16. Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional Neural Network With Data Augmentation for SAR Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 364–368. [CrossRef]
17. Xu, Y.; Scott, K.A. Sea ice and open water classification of sar imagery using cnn-based transfer learning. In Proceedings of the IGARSS 2017—2017 IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017; pp. 3262–3265.
18. Bentes, C.; Velotto, D.; Tings, B. Ship Classification in TerraSAR-X Images With Convolutional Neural Networks. *IEEE J. Ocean. Eng.* **2018**, *43*, 258–266. [CrossRef]
19. Bentes, C.; Velotto, D.; Lehner, S. Target classification in oceanographic SAR images with deep neural networks: Architecture and initial results. In Proceedings of the 2015 IEEE Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 3703–3706.
20. Ødegaard, N.; Knapskog, A.O.; Cochin, C.; Louvigne, J.C. Classification of ships using real and simulated data in a convolutional neural network. In Proceedings of the 2016 Radar Conference, Philadelphia, PA, USA, 2–6 May 2016; pp. 1–6.

21. An, Q.; Pan, Z.; You, H. Ship Detection in Gaofen-3 SAR Images Based on Sea Clutter Distribution Analysis and Deep Convolutional Neural Network. *Sensors* **2018**, *18*, 334. [CrossRef] [PubMed]
22. Wang, C.; Bi, F.; Zhang, W.; Chen, L. An Intensity-Space Domain CFAR Method for Ship Detection in HR SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 529–533. [CrossRef]
23. Shi, H.; Zhang, Q.; Bian, M.; Wang, H.; Wang, Z.; Chen, L.; Yang, J. A Novel Ship Detection Method Based on Gradient and Integral Feature for Single-Polarization Synthetic Aperture Radar Imagery. *Sensors* **2018**, *18*, 563. [CrossRef] [PubMed]
24. Jiao, J.; Zhang, Y.; Sun, H.; Yang, X.; Gao, X.; Hong, W.; Fu, K.; Sun, X. A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection. *IEEE Access* **2018**, *6*, 20881–20892. [CrossRef]
25. Kang, M.; Ji, K.; Leng, X.; Lin, Z. Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection. *Remote Sens.* **2017**, *9*, 860. [CrossRef]
26. Kuruoglu, E.E.; Zerubia, J. Modeling SAR images with a generalization of the Rayleigh distribution. *IEEE Trans. Image Process.* **2004**, *13*, 527–533. [CrossRef] [PubMed]
27. Li, H.C.; Hong, W.; Wu, Y.R.; Fan, P.Z. An Efficient and Flexible Statistical Model Based on Generalized Gamma Distribution for Amplitude SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2711–2722.
28. Armstrong, B.C.; Griffiths, H.D. CFAR detection of fluctuating targets in spatially correlated K-distributed clutter. *IEE Proc. F Radar Signal Process.* **1991**, *138*, 139–152. [CrossRef]
29. Ren, S.; Girshick, R.; 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]
30. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
31. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
32. Everingham, M.; Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
33. Cui, Z.; Dang, S.; Cao, Z.; Wang, S.; Liu, N. SAR Target Recognition in Large Scene Images via Region-Based Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 776. [CrossRef]
34. Cozzolino, D.; Martino, G.D.; Poggi, G.; Verdoliva, L. A fully convolutional neural network for low-complexity single-stage ship detection in Sentinel-1 SAR images. In Proceedings of the Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017; pp. 886–889.
35. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In Proceedings of the International Workshop on Remote Sensing with Intelligent Processing, Shanghai, China, 18–21 May 2017; pp. 1–4.
36. Wang, Y.; Wang, C.; Zhang, H.; Zhang, C.; Fu, Q. Combing Single Shot Multibox Detector with transfer learning for ship detection using Chinese Gaofen-3 images. In Proceedings of the Progress in Electromagnetics Research Symposium-Fall, Singapore, 19–22 November 2017; pp. 712–716.
37. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **2007**, *9*, 62–66. [CrossRef]
38. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. Available online: https://arxiv.org/abs/1409.1556(accessed on 13 December 2018).
39. Li, C.; Xu, C.; Gui, C.; Fox, M.D. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **2010**, *19*, 3243–3254. [PubMed]
40. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 675–678.
41. Oliva, A.; Torralba, A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *Int. J. Comput. Vis.* **2001**, *42*, 145–175. [CrossRef]