SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2019, Vol. 62 042301:1–042301:16
https://doi.org/10.1007/s11432-017-9405-6

A coupled convolutional neural network for small and densely clustered ship detection in SAR images

Juanping ZHAO, Weiwei GUO, Zenghui ZHANG* & Wenxian YU

Shanghai Key Laboratory of Intelligent Sensing and Recognition, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240*, China*

Received 22 September 2017/Revised 20 December 2017/Accepted 2 April 2018/Published online 19 September 2018

Abstract Ship detection from synthetic aperture radar (SAR) imagery plays a significant role in global marine surveillance. However, a desirable performance is rarely achieved when detecting small and densely clustered ship targets, and this problem is difficult to solve. Recently, convolutional neural networks (CNNs) have shown strong detection power in computer vision and are flexible in complex background conditions, whereas traditional methods have limited ability. However, CNNs struggle to detect small targets and densely clustered ones that exist widely in many SAR images. To address this problem while preserving the good properties for complex background conditions, we develop a coupled CNN for small and densely clustered SAR ship detection. The proposed method mainly consists of two subnetworks: an exhaustive ship proposal network (ESPN) for ship-like region generation from multiple layers with multiple receptive fields, and an accurate ship discrimination network (ASDN) for false alarm elimination by referring to the context information of each proposal generated by ESPN. The motivation in ESPN is to generate as many ship proposals as possible, and in ASDN, the goal is to obtain the final results accurately. Experiments are evaluated on two data sets. One is collected from 60 wide-swath Sentinel-1 images and the other is from 20 GaoFen-3 (GF-3) images. Both data sets contain many ships that are small and densely clustered. The quantitative comparison results illustrate the clear improvements of the new method in terms of average precision (AP) and $F1$ score by 0.4028 and 0.3045 for the Sentinel-1 data set compared with the multi-step constant false alarm rate (CFAR-MS) method. The values are verified as 0.2033 and 0.1522 for the GF-3 data set. In addition, the new method is demonstrated to be more efficient than CFAR-MS.

Keywords SAR image, ship detection, CNN, exhaustive ship proposal network (ESPN), accurate ship discrimination network (ASDN)

Citation Zhao J P, Guo W W, Zhang Z H, et al. A coupled convolutional neural network for small and densely clustered ship detection in SAR images. Sci China Inf Sci, 2019, 62(4): 042301, https://doi.org/10.1007/s11432-017-9405-6

1 Introduction

Ship detection from synthetic aperture radar (SAR) images, which plays a crucial role in global marine surveillance, has been explored in numerous studies [1, 2]. SAR images are highly suitable for ship detection because of their all-weather and day-and-night characteristics [3]. However, SAR ship detection faces numerous challenges. First, ships encounter complex background conditions, such as in inshore, offshore, and inland river locations, and sometimes they are distracted by oil spills and small islands that are visually similar to the ship targets [4]. Even worse, ships are small in size and are usually densely clustered. These problems inevitably lead to false or missing detections.

* Corresponding author (email: zenghui.zhang@sjtu.edu.cn)

© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2018

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

In previous studies, ship detection systems usually consist of four steps: land masking, preprocessing, prescreening, and discrimination [5]. Most ship detection systems require land masking, where the land areas are omitted from the images being processed. Employing registration [6] and coastline detection methods [7], land masking is intended to eliminate the adverse effect caused by land. Many preprocessing methods, such as speckle filtering, aim to improve the facility and accuracy of subsequent stages. The prescreening step is essential for ship detection systems, which attempts to find candidate regions as ship proposals. Constant false alarm rate (CFAR) and generalized-likelihood ratio test are the most widely used prescreening algorithms that serve as the basis of many SAR ship detection methods [8–10]. Finally, the discrimination step aims to eliminate the regions containing false alarms and accept those containing real targets. Notably, salient information is utilized for SAR ship detection [11,12]. Although these systems have been widely used and achieved successful detection results, they still have limitations. First, traditional multi-step systems are time-consuming and not highly intelligent. Second, handcrafted features have limited representation capabilities for discrimination in complex background conditions.

A convolutional neural network (CNN) is a trainable multi-layer architecture that typically consists of one or more pairs of convolutional and pooling layers, followed by several consecutive fully connected layers. Its deep architecture enables extraction of a set of discriminating features at multiple levels [13]. Most of the prevailing detectors are developed based on faster region-based convolutional neural networks (FRCN) [14], such as R-FCN [15], inside-outside net [16], and deep saliency [17], to satisfy various requirements. FRCN is composed of a region proposal network (RPN) for predicting candidate regions and fast R-CNN [18] for classifying object proposals and refining their spatial locations. FRCN is an intelligent end-to-end data-driven detector that takes an image as input and then outputs the location of objects. To detect multi-scale objects, many studies have explored the usage of multi-scale information. Among them, the single-shot detector (SSD) [19] is one of the first attempts to use pyramid feature hierarchy. In this paper, small convolutional filters are used for several intermediate layers and the results are predicted based on each layer. Sequentially, the multiscale CNN (MSCNN) [20] is proposed to predict outputs from several layers, and the feature upsampling by deconvolution technique is explored to reduce the memory and computational costs. The feature pyramid network [21] developed a top-down architecture with lateral connections to build high-level semantic feature maps at all scales. Another work is subcategory-aware CNN [22], which utilizes many different resized images, i.e., image pyramids, as input, and a feature extrapolating layer is introduced to compute features in multiple scales. The feature extrapolating layer generates feature maps for scales that are not covered by the input image pyramid via extrapolation.

To the best of our knowledge, detectors tailored for small and densely clustered ship targets in SAR images are rare. However, ships are extremely small in size, especially when the spatial resolution or pixel spacing is not high, and are spatially distributed in different ways, i.e., isolated or densely clustered. Moreover, ship targets always face complex background conditions, such as inshore or offshore locations, or distractions caused by sea clutter. By the way, in order to maintain the characteristics of electromagnetic scattering information in SAR images, image re-sampling in optical community is often not accepted in SAR community.

To address these problems, we propose a coupled CNN called Coupled-CNN_E_A for detecting ship targets in SAR imagery. This method is particularly effective for small-sized ship targets and densely clustered ones. Inspired by the detection framework of FRCN, the proposed method consists of two significant parts: an exhaustive ship proposal network (ESPN) to generate ship proposals and an accurate ship discrimination network (ASDN) for ruling out false alarms. Throughout the entire process, the following instructions are followed in principle:

- In ESPN, we attempt to propose ship targets exhaustively;
- In ASDN, the false alarms should be discarded as carefully as possible.

Different from RPN in FRCN and other pyramid CNN methods, ESPN extracts feature maps from several representative convolutional layers, i.e., through the convolutional branch, to fully utilize the lower layers' detailed information and the higher layers' abstract information. We further extend each branch to several sub-branches by employing multiple convolutional filters with different sizes to extract

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

feature maps that could fully perceive small ship targets and densely clustered ones without losing other ship targets. For proposal generation, each sub-branch is referred to as a proposal generator. The convolutional filters in ESPN are relatively small in size.

Finding small and densely clustered targets is a fundamentally difficult problem because only a limited signal on the object can be exploited. Considering this condition, one must use image information beyond the object extent, i.e., context region. Moreover, it is visually clear that targets on land are surrounded only by land. Offshore ships are mostly surrounded by water. For inshore ships, approximately half of the surrounding area is water and the rest is land. In addition, researchers have found that the ocean surroundings cannot only help detectors to effectively rule out false alarms on land, but can also help them recognize inshore ship targets well [23, 24]. With all these factors considered, context information plays a vital role in SAR image understanding. Therefore, we combine the ship proposals and their context information together in ASDN to further improve the detection performance. Specifically, the proposal features and context features learned by ASDN are fused by concatenation.

Our method is tested and evaluated on two challenging data sets. One is collected from 60 wide-swath Sentinel-1 images with tens of thousands of annotations. The other data set comes from recently collected GaoFen-3 (GF-3) satellite images, which include 20 images. Both of the data sets are annotated manually by referring to the corresponding automatic identification system (AIS) information and experts' inspections. There are many small and densely clustered ship targets and the ships face various complex background conditions. The quantitative and qualitative comparison results show that our method could perform better than other existing baselines.

The main contributions of this work are the following.

• Two data sets, i.e., Sentinel-1 and GF-3, which contain many small and densely clustered ship targets, are collected and annotated. Based on them, an effective and efficient deep CNN-based method is developed.

• Multiple representative convolutional layers are reused in ESPN, and image features are extracted by extending each branch into sub-branches via several small convolutional filters with different sizes, where ship proposals are generated. Employing this technique, ESPN enables one to exhaustively obtain ship proposals of small and densely clustered ships.

• To eliminate false alarms more accurately in ASDN, we improve the representation capability by taking full advantage of the ship proposals and their context information. The ship features in ASDN are simply extracted by concatenating both the ship proposal and context regions after the RoI pooling operation.

The paper is structured as follows. Section 2 represents a detailed explanation of the proposed method and provides the computational cost. The experimental results and analysis of two data sets are presented in Section 3. Finally, conclusion is drawn in Section 4.

2 Methodology

Figure 1 provides an overview of the processing workflow, which displays the composition of the proposed network. The method consists of an ESPN and an ASDN, both of which share a CNN trunk for feature learning. Due to the limited GPU memory, the original large-scale SAR images are tiled into image blocks for training and testing. The detection results of image blocks from the same original imagery are stitched together. In this section, we describe the shared convolutional layers, the design for ESPN and ASDN, separately. Finally, the computational cost of the proposed method is analyzed.

2.1 Shared convolutional layers

CNN performs local operation on the input image with linear or nonlinear functions, mainly including convolutional, pooling, and activation layers. With a set of fixed-size filters, CNN generates feature maps by employing these filters sliding on local receptive fields after receiving feature maps of the last layer or the input image block. Weights of the filters are shared between convolutional layers. Pooling

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

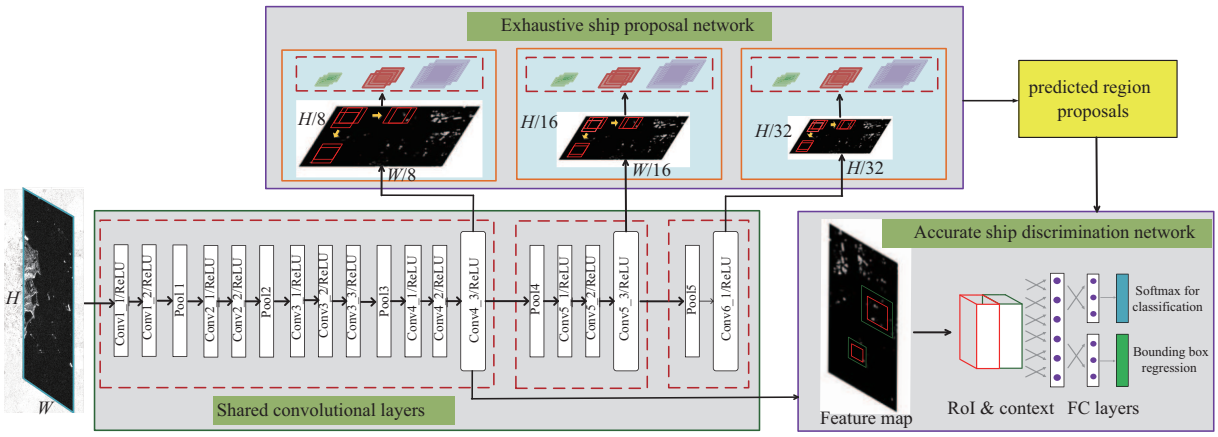

Figure 1 Overview of proposed method, mainly including an ESPN and an ASDN, both of which share convolutional layers for feature learning. In ASDN, "RoI" represents the regions of interest and "FC" layer indicates the fully connected layer.

layers use filters to generalize the brief representation of the convolutional layers to reduce the number of parameters. They yield the maximum or average values in each local region of input image or feature map, and provide a form of translation invariance. Rectified linear units (ReLU) [13] is used as an activation function. The mathematical transformation between each input pixel value x and its output y can be formulated as

$$y = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Therefore, the derivative for ReLU activation function is deduced as

$$y' = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

ReLU has certain advantages to become a popular choice for deep neural networks [25]. First, it is simple to use, i.e., the output is 0 when the input pixel value is less than 0, and raw otherwise. Meanwhile, the derivative is 1 when receiving a positive input, and 0 otherwise. This characteristic could effectively prevent the occurrence of vanishing gradients. Moreover, ReLU saturates at exactly 0, which is potentially helpful when using hidden activations as input features for a classifier. Finally, ReLU has been verified to speed up training cost for the simplest gradient computation, i.e., 0 or 1.

After the linear and nonlinear transformations layer by layer, the resolutions of the feature maps are usually changed and features become more abstract and semantic. For example, if the size of the input image block is $W \times H$ (W and H denote the width and the height of the input image, respectively), then the output feature map's resolution becomes $\frac{W}{2} \times \frac{H}{2}$, after a 2×2 max pooling layer without padding. Specifically, for a ship target with length of less than 32 pixels, the features may be lost after five such pooling layers in VGG-16 [26]. Additionally, features for densely clustered ship targets may be obscured severely. Considering both factors, comprehensively utilizing feature maps from different layers is valuable.

In our experiment, the VGG-16 configuration [26] is adopted as the shared CNN trunk, as shown by the "shared convolutional layers" in Figure 1, through which multilevel feature maps with different resolutions could be extracted. In detail, features from three representative branches, namely, "Conv4_3" layer, "Conv5_3" layer, and "Conv6_1" layer, are extracted. The size of feature maps from these three layers are mapped to a resolution of $\frac{W}{8} \times \frac{H}{8}$, $\frac{W}{16} \times \frac{H}{16}$ and $\frac{W}{32} \times \frac{H}{32}$, respectively. The lower layers' feature maps contain more finegrained information. By contrast, features from a deeper layer carry a larger volume of abstract and semantic information, which is crucial in object detection.

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

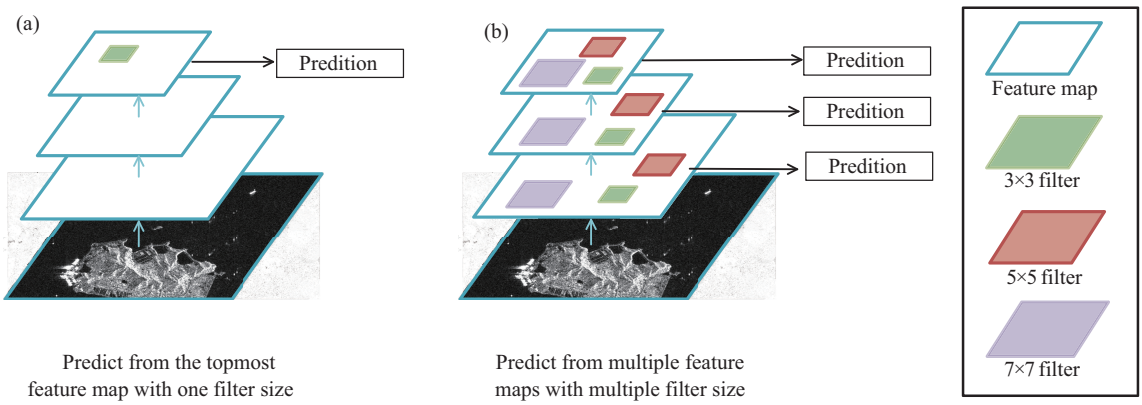

Figure 2 Ship proposal generation strategy of (a) RPN and (b) ESPN.

2.2 ESPN

ESPN takes a SAR image block as input and exhaustively outputs a set of ship-like regions. Instead of generating proposals from the topmost convolutional layer with single sized filters in RPN (Figure 2(a)), ESPN operates sliding window filtering from several intermediate layers with different-sized filters, as shown in Figure 2(b), to enable the final proposals to perceive as many ship targets as possible, even for the small and densely clustered targets.

In detail, after the sharing feature learning stage, ESPN predicts all ship-like regions through the three proposal branches mentioned in Subsection 2.1 and each of them implements the sliding operation with multiple specified filters to extract local feature representation X_i at each location, where i is the feature index. In this experiment, we set the sizes of multiple sliding windows for each proposal branch as 3×3, 5×5, and 7×7 in accordance with the ship size in our data set.

Without loss of generality, the anchor boxes assumed as $B_i = (b_i^x, b_i^y, b_i^w, b_i^h)$ are predicted according to the filter size. Here, b_i^x & b_i^y denote the top-left coordinates of the anchor box regions and b_i^w & b_i^h represent the width and height of the anchor box regions, respectively. Each anchor box has three different height-to-width ratios (i.e., 2:1, 1:1, and 1:2) to represent ship targets of different shapes, as shown in Table 1. Notably, a deeper layer and a larger sliding window are both in accordance with a lager receptive field, which is suitable to detect large and isolated objects. By contrast, a lower layer with a small sliding window is often effective in detecting small and densely clustered objects. Based on all these factors, the height of the proposals for "Conv4_3" layer is set as 10, 16, and 22 pixels. Meanwhile, for the "Conv5_3" layer, the ship proposals' height is set as 28, 34, and 40 pixels. In addition, for the "Conv6_1" layer, the value is set as 46, 52, and 58 pixels.

To construct training samples for each detection layer explicitly, the predicted region boxes located outside the image boundary are discarded and a class label $Y_i \in \{0, 1\}$ is assigned to the remaining ones. In the training phase, we assign a positive label $Y_i = 1$ to the predicted region box B_i, if the intersection-over-union (IoU) overlap ratio with a ground-truth box B_i^* reaches the highest. However, if a predicted region box's IoU ratio is lower than 0.3 for all ground-truth boxes, we assign a negative label $Y_i = 0$ to it, and others are discarded. The IoU overlap ratio is defined as

$$\text{IoU} = \frac{A(B_i \cap B_i^*)}{A(B_i \cup B_i^*)}, \tag{3}$$

where $A(B_i \cap B_i^*)$ represents the intersection area of the predicted region box and the ground-truth box, and $A(B_i \cup B_i^*)$ denotes their union.

With the aid of definitions and parameter settings mentioned, each detection layer's training samples are defined as $S^m = \{(X_i, Y_i, B_i)\}_{i=1}^N$, where $m \in \{1, 2, \ldots, M\}$ represents the index of the detection layer. Based on the work of Ren et al. [14], the loss for each sub-branch is a combination of classification

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

Table 1 Parameter configurations for three proposal branches in ESPN

Layer name	Conv4_3			Conv5_3			Conv6_1		
Filter size (pixel)	3×3	5×5	7×7	3×3	5×5	7×7	3×3	5×5	7×7
Anchor height (pixel)	10	16	22	28	34	40	46	52	58
Height-to-width ratio	1:2,1:1,2:1	1:2,1:1,2:1	1:2,1:1,2:1	1:2,1:1,2:1	1:2,1:1,2:1	1:2,1:1,2:1	1:2,1:1,2:1	1:2,1:1,2:1	1:2,1:1,2:1

and bounding box regression, which is defined as follows:

$$l^m(X,Y,B|W)=L_{\rm CLS}(P(Y|X),y^*)+\lambda[Y=1]L_{\rm BBR}(\hat{B},B^*), \tag{4}$$

where W stands for the parameters of the network, the balancing parameter λ is set as 0.05, y^* is the ground-truth label, the classification loss $L_{\rm CLS}(P(Y|X),y^*)=-(1-y^*)\log P(Y=0|X)-y^*\log P(Y=1|X)$, $P(Y|X)$ is the probability confidence over two classes, for example, ship-like and non-ship-like regions, computed by

$$P(Y|X)=\begin{pmatrix}P(Y=0|X)\\P(Y=1|X)\end{pmatrix}=\frac{1}{\sum_{k=0}^{K-1}\exp(\theta_k^{\rm T}X)}\begin{pmatrix}\exp(\theta_0^{\rm T}X)\\\exp(\theta_1^{\rm T}X)\end{pmatrix}, \tag{5}$$

where $K=2$ indicates the total number of classes. $\theta=[\theta_0,\theta_1]^{\rm T}$ is the parameter set for (5). Besides, $[Y=1]$ is equal to 1 when the regressed bounding box contains a real SAR ship; otherwise, $[Y=1]$ is equal to 0. This implies that the background is meaningless for training bounding box regression. $\hat{B}$ stands for the regressed bounding box, and $L_{\rm BBR}$ denotes the bounding box regression loss given by

$$L_{\rm BBR}(\hat{B},B^*)=f_{L_1}(\hat{B}^\dagger-B^{*\dagger}), \tag{6}$$

where

$$f_{L_1}(x)=\begin{cases}0.5x^2, & \text{if } |x|<1,\\ |x|-0.5, & \text{otherwise,}\end{cases} \tag{7}$$

$(\cdot)^\dagger$ represents the coordinate offsets of bounding box $(\cdot)$ with reference to the anchor box [14]. Studies [18, 27] have demonstrated that the smoothed L_1 loss is superior to L_2 loss for two reasons. First, L_1 loss is less sensitive to outliers than L_2 loss. Second, when the regression targets are bounded, training with L_2 loss can require careful turning of learning rates to prevent exploding gradients.

In accordance with the preceding definitions, the overall loss function of ESPN is a summation for each detection layer, given by

$$L_{\rm ESPN}(W)=\sum_{m=1}^{M}\sum_{i\in S^m}\alpha_m l^m(X_i,Y_i,B_i|W), \tag{8}$$

where the number of detection layers, denoted by M, is equal to 9, which stands for three proposal branches with three different detection layers and α_m denotes the weight for each detection layer's loss.

The optimal parameters $W^*=\arg\min_W L_{\rm ESPN}(W)$ are optimized by stochastic gradient descent (SGD) [27]. To prevent overfitting, we adopt the pre-trained image-net model [13] to finetune this network.

As is well-known, the lower layer of the CNN architecture affects gradients more than the deeper layer does. Thus, we set a slightly lower weight, i.e., $\alpha_m=0.9$, for the "Conv4_3" proposal branch and a slightly higher weight, i.e., $\alpha_m=1$, for the "Conv5_3" proposal branch and the "Conv6_1" proposal branch. When the training of the ESPN is finished, it takes image blocks as input and then outputs the location of ship targets through several proposal branches. Thereafter, these results are concatenated to form the final proposal detections.

2.3 ASDN

ASDN aims to eliminate false alarms generated by ESPN as accurately as possible. It takes an image block with proposals as input and then outputs the classification result and refined locations.

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

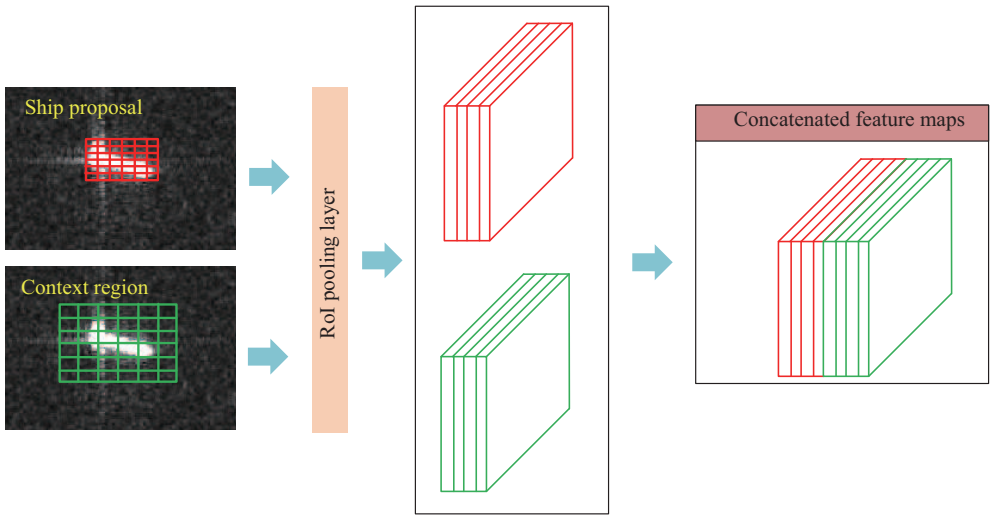

Figure 3 Ship proposals and their contextual regions are both bypassed by the RoI pooling layer. The upper row and the lower row show an example of a ship proposal and its context region, respectively. Both of them are pooled by the RoI pooling operation. Then, the feature maps are concatenated into one fused feature map to improve their representation capability.

Given the importance of context information for SAR images [23,24], in this paper, features from each proposal (red cube in ASDN of Figure 3) and its context information (green cube) are both analyzed. The optimal context region is determined by using the grid search strategy, i.e., each context region is set to be 1.5 times larger than the proposal region. Since these predicted ship-like regions and the context regions have different sizes, we adopt an RoI pooling layer for each proposal and each context region to generate features of a fixed dimension. Then, these fixed-size feature maps are concatenated into a single feature cube called fused features. Figure 3 illustrates the details of this process. Thereafter, the fused features are fed into subsequent fully connected layers and branched into two parts for further classification and bounding box regression.

The loss l_{ASDN} for ASDN is similar to (4), which combines a cross-entropy loss for classification and a smoothed L_1 loss for bounding box regression. In ASDN, the training samples S^{M+1} are collected in the same way as S^m. Therefore, the multi-task loss of (8) is extended to

$$L(W, W_d) = L_{\mathrm{ESPN}}(W) + \alpha_{M+1} \sum_{i \in S^{M+1}} l_{\mathrm{ASDN}}(X_i, Y_i, B_i | W_d), \tag{9}$$

where α_{M+1} denotes the weight of ASDN's loss, and W_d stands for the added parameter set of fully connected layers.

The pre-trained ESPN model is adopted to initialize the shared layers of ASDN, in that ESPN and ASDN share some convolutional layers for feature learning. The parameters are learned jointly by minimizing the final loss function (9), i.e., $(W^*, W_d^*) = \arg\min L(W, W_d)$, with SGD [28] strategy throughout the unified network. The additional weights of fully connected layers are randomly initialized by zero-mean Gaussian distribution with 0.01 standard deviation. The initial learning rate is 0.0005, which decreases by 0.1 every 5k iterations. The trade-off coefficient λ is equal to 1. The entire training procedure proceeds recursively until the overall loss function is converged. Actually, 35k iterations are required. Finally, non-maximum suppression (NMS) [29] is adopted to reduce the redundancies.

2.4 Computational cost

The detailed structure, number of parameters, and multiply-add computational cost (MAC) are illustrated in Table 2 based on the assumption that the size of the input image block is 1024×768. Given the computational cost, the number of parameters and the MAC are computed by referring to the detailed configurations of each layer. Table 3 summarizes the MAC and the number of parameters in the three

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

Table 2 Detailed structure, number of parameters, and MAC for each layer when using Coupled-CNN_E_A method with 1024×768 input

Part	Name	Type	Stride	Output	#Params.	MAC
Shared CNN layers	Conv1_1	3×3 convolution	1	$1024 \times 768 \times 64$	1.9k	1359.0M
	Conv1_2	3×3 convolution	1	$1024 \times 768 \times 64$	41.0k	28991.0M
	Pool1	2×2 max pooling	2	$512 \times 384 \times 64$		
	Conv2_1	3×3 convolution	1	$512 \times 384 \times 128$	81.9k	14495.5M
	Conv2_2	3×3 convolution	1	$512 \times 384 \times 128$	163.8k	28991.0M
	Pool2	2×2 max pooling	2	$256 \times 192 \times 128$		
	Conv3_1	3×3 convolution	1	$256 \times 192 \times 256$	327.7k	14495.5M
	Conv3_2	3×3 convolution	1	$256 \times 192 \times 256$	655.4k	28991.0M
	Conv3_3	3×3 convolution	1	$256 \times 192 \times 256$	655.4k	28991.0M
	Pool3	2×2 max pooling	2	$128 \times 96 \times 256$		
	Conv4_1	3×3 convolution	1	$128 \times 96 \times 512$	1310.7k	14495.5M
	Conv4_2	3×3 convolution	1	$128 \times 96 \times 512$	2621.4k	28991.0M
	Conv4_3	3×3 convolution	1	$128 \times 96 \times 512$	2621.4k	28991.0M
	Pool4	2×2 max pooling	2	$64 \times 48 \times 512$		
	Conv5_1	3×3 convolution	1	$64 \times 48 \times 512$	2621.4k	7247.8M
	Conv5_2	3×3 convolution	1	$64 \times 48 \times 512$	2621.4k	7247.8M
	Conv5_3	3×3 convolution	1	$64 \times 48 \times 512$	2621.4k	7247.8M
	Pool5	2×2 max pooling	2	$32 \times 24 \times 512$		
	Conv6_1	3×3 convolution	1	$32 \times 24 \times 512$	2621.4k	18119M
ESPN	SPN4_3	3×3 convolution	1	$128 \times 96 \times 6$	30.7k	339.7M
	SPN4_5	5×5 convolution	1	$128 \times 96 \times 6$	79.9k	943.7M
	SPN4_7	7×7 convolution	1	$128 \times 96 \times 6$	153.6k	1849.7M
	SPN5_3	3×3 convolution	1	$64 \times 48 \times 6$	30.7k	84.9M
	SPN5_5	5×5 convolution	1	$64 \times 48 \times 6$	79.9k	235.9M
	SPN5_7	7×7 convolution	1	$64 \times 48 \times 6$	153.6k	462.4M
	SPN6_3	3×3 convolution	1	$32 \times 24 \times 6$	30.7k	21.2M
	SPN6_5	5×5 convolution	1	$32 \times 24 \times 6$	79.9k	59.0M
	SPN6_7	7×7 convolution	1	$32 \times 24 \times 6$	153.6k	115.6M
ASDN	RoIPooling1	7×7 RoI pooling		$7 \times 7 \times 512$		
	RoIPooling2	7×7 RoI pooling		$7 \times 7 \times 512$		
	RoI_concat	3×3 convolution		$5 \times 5 \times 512$	5242.9k	118.0M
	FC	FC		4096	52428.8k	52.4M
	FC_cls	FC		2	8.2k	32.8K
	FC_bbr	FC		8	32.8k	32.8K
Total					**75.66M**	**256.77B**

Table 3 Comparison result of the number of parameters and the MAC for each part when using Coupled-CNN_E_A method with 1024×768 input

	#Params.				MAC			
	Shared CNN layers	ESPN	ASDN	Total	Shared CNN layers	ESPN	ASDN	Total
Coupled-CNN_E_A	18966.2k	792.6k	57712.7k	**75.66M**	258653.9M	4112.1M	170.4M	**256.77B**

parts: shared convolutional layers, ESPN and ASDN. The numbers in bold represent the total number of parameters and the MAC. The underlined numbers represent the number of parameters for "shared CNN layers" part and the MAC for this part. Specifically, for one image with size of 1024×768 as input, our method takes 256.77 billion MAC and 75.66 million parameters for one iteration. Observing the comparison results in Table 3, one can easily find that the shared CNN part takes much more burden on computational cost in terms of MAC. Therefore, we can conclude that the required time for training and testing mainly depends on the shared CNN part. Then, we can infer that the efficiency of the proposed method is similar to FRCN because the MAC brought by the special design in our method is far less than the shared CNN part.

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

Table 4 Information of Sentinel-1 imagery used in this study

Satellite	Imaging mode	Band	Polarization	Product type	Resolution (rg×az) (m)	Pixel spacing (rg×az) (m)	Average size per image (rg×az) (pixel)
Sentinel-1	IW	C	VH	GRD	20×22	10×10	25000×18000

Table 5 Information of GF-3 imagery used in this study

Satellite	Imaging mode	Band	Polarization	Pixel spacing (rg×az) (m)	Average size per image (rg×az) (pixel)
GF-3	NSC	C	VH	20×5	8800×21000

3 Experiments

3.1 Configuration

3.1.1 *Hardware environment*

Experiments are implemented based on the deep learning framework Caffe [30] and executed on a workstation with two Intel 32 Core i7 CPUs with 64 GB RAM and an NVIDIA GTX-1080 GPU with 8 GB memory. The operating system is Ubuntu 14.04.

3.1.2 *Data set description*

In this paper, the first experimental data set is collected from 60 wide-swath Sentinel-1 images, the main information of which is listed in Table 4. We conduct the experiments by adopting the interferometric wide-swath (IW) mode ground-range detected (GRD) Sentinel-1 imagery with C band. The resolution is 20 m in range direction and 22 m in azimuth direction. The pixel spacing is 10 m both in range and azimuth directions. The wide-swath Sentinel-1 image size used in our experiments is approximately 25000×18000 in pixel. These images are manually annotated by expert inspections on the Sentinel-1 application platform (SNAP) [31] partially with the help of AIS. In this case, among the 60 images, 52 are randomly selected for training and the remaining 8 are used for testing.

The second data set is the most recently collected GF-3 image. GF-3 is the first commercial C-band full polarimetric SAR satellite in China. It was launched in August 2016. It has multiple strip and scan imaging modes [32]. In this paper, we annotated 20 narrow-scan (NSC) mode images manually with the help of the corresponding AIS information and expert inspections. Among all the annotated images, 16 are used for training and the remaining 4 are used for testing. The average size of the images in this data set is approximately 8800×21000 in pixel. The pixel spacing is 20 m and 5 m along the azimuth and ground-range directions, respectively. The main meta data of GF-3 is shown in Table 5. Many small and densely clustered ship targets are in this data set.

Considering the limited GPU memory and processing speed, each SAR image is tiled into several adjacent image blocks whose size is approximately 1024×768 pixels for both training and testing. When training and testing the Sentinel-1 data set, we set an overlap of 50 pixels for each adjacent divided image blocks, which is larger than the average ship length (25 pixels for the Sentinel-1 data set). For the GF-3 data set, the overlap of adjacent image blocks is also set as 50 pixels because the average ship length in this data set is 23 pixels, which is slightly less than the average ship length in the Sentinel-1 data set. Figure 4 shows a sample image that has been divided into 3×3 blocks. The white rectangles represent the divided blocks without overlap, and the red rectangles represent divided blocks where 50 pixels are longer and wider than the white rectangles at the same upper-left corners (indicated by blue solid dots).

Thereafter, each image block is processed individually. According to the order of the division, the adjacent image blocks' detection results are stitched together by adding the coordinates of the corresponding upper-left corner. For overlapping of the adjacent image blocks, we adopt the NMS strategy to eliminate the redundant bounding boxes.

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

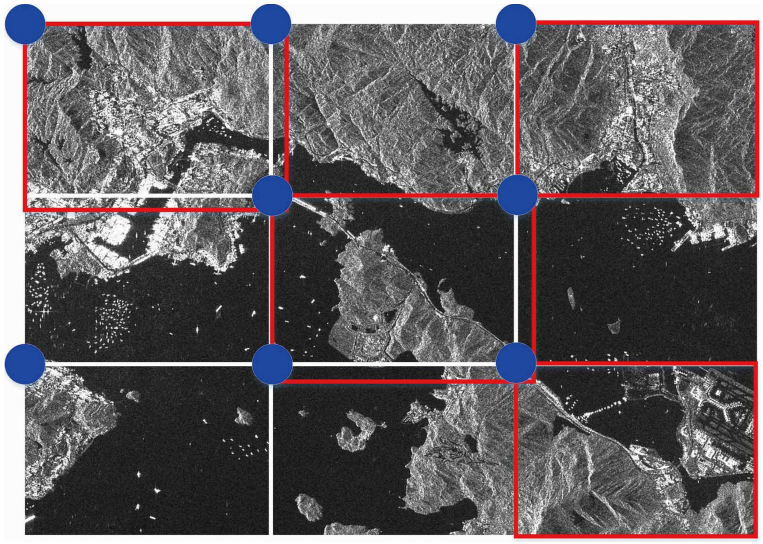

Figure 4 Illustration of image cropping strategy. The white grids are blocks without overlap. The red rectangles represent divided blocks with 50 pixel overlap. In both cases, blocks without ship targets are discarded and the remaining ones are used for training and testing.

3.1.3 *Evaluation metrics*

We adopted four widely used criteria to quantitatively evaluate the performance of detection, namely, precision, recall, average precision (AP), and $F1$ score. The precision measures the fraction of detections that are true positives, given by [33]

$$\text{precision} = \frac{\#\text{TruePositive}}{\#\text{TruePositive} + \#\text{FalsePositive}}. \tag{10}$$

Here, $\#(\cdot)$ represents the number of $(\cdot)$. The recall measures the fraction of positives over the number of ground-truths.

$$\text{recall} = \frac{\#\text{TruePositive}}{\#\text{TruePositive} + \#\text{FalseNegative}}. \tag{11}$$

The AP value and $F1$ score are obtained by combining the precision and recall metrics. This approach results in two single measurements to comprehensively evaluate the quality of a detection method. The AP metric is measured by the integral of the precision over the interval from recall $= 0$ to recall $= 1$, i.e., the area under the precision-recall curve (PRC) [34]. The $F1$ score is given by

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{12}$$

3.1.4 *Baseline methods*

To verify the performance of the special design in ESPN and ASDN, separately, our method is evaluated by employing three configurations, namely, "Coupled-CNN_E", "Coupled-CNN_A", and "Coupled-CNN_E_A", where "E" and "A" indicate the improvements in ESPN and ASDN, respectively. CFAR-based multi-stage (CFAR-MS) detector developed by the authors in [35] is implemented by SNAP. FRCN is selected as another baseline to demonstrate the superiority of the proposed techniques.

3.2 Results and discussions

3.2.1 *Sentinel-1 data set*

The proposal quality is evaluated in Figure 5(a), which is plotted to describe the recall of different methods under different IoU for all the testing images. This figure shows that the recall for each method drops when IoU increases. Compared with CNN-based methods, CFAR-MS achieves the lowest recall under a

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

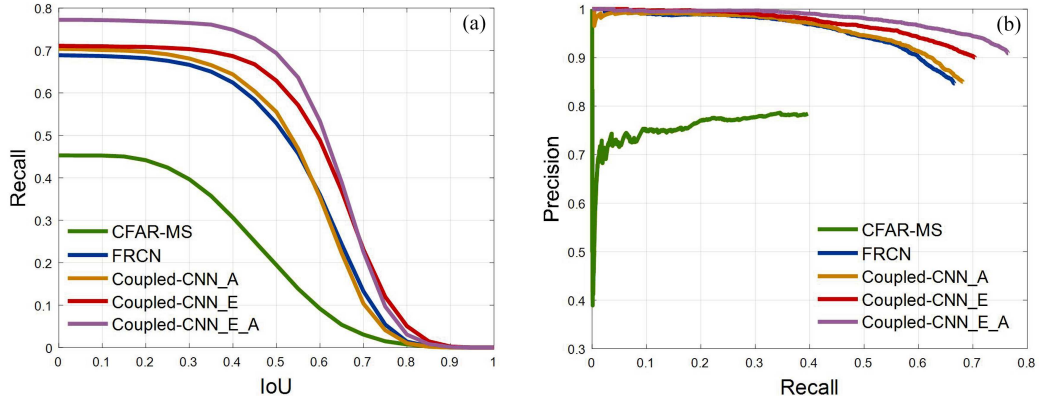

Figure 5 Performance curves over eight Sentinel-1 images. (a) Recall vs. IoU curve for each method; (b) precision vs. recall curve for each method.

Table 6 Performance comparison of different methods for the Sentinel-1 data set[a)]

Methods	Ground truth	True positive	False positive	Recall	Precision	Average precision	$F1$ score	Average time (s) per image
CFAR-MS	6814	2710	751	0.3977	0.7830	0.3123	0.5275	2550
FRCN	6814	4544	845	0.6669	0.8432	0.5812	0.7447	105
Coupled-CNN_E	6814	4843	560	<u>0.7107</u>	<u>0.8964</u>	<u>0.6519</u>	<u>0.7928</u>	113
Coupled-CNN_A	6814	4656	823	0.6833	0.8498	0.6069	0.7575	108
Coupled-CNN_E_A	6814	5260	570	**0.7719**	**0.9022**	**0.7151**	**0.8320**	115

a) The bold numbers denote the optimal values in each column. The underlined numbers denote the suboptimal values in each column.

constant IoU, and it declines most sharply when IoU increases. This condition illustrates its relatively low recall and poor localization performance. Determining a suitable IoU in the beginning is important because a higher IoU means a more rigorous prerequisite and a lower IoU means a loose prerequisite. In Figure 5(a), the turning point of CFAR-MS emerges when the IoU is equal to 0.3. By contrast, in the interval IoU $\in [0, 0.3]$, the recall remains stable for all the CNN-based methods. Therefore, in the testing phase, the predicted region is considered to be a true positive if the IoU is larger than 0.3. Otherwise, without any loss of generality, the predicted region is assumed to be a false positive. By analyzing the comparison results of two groups, namely, Coupled-CNN_E_A & Coupled-CNN_A and Coupled-CNN_E & FRCN, the recall is improved when IoU is equal to 0.3.

Based on this finding, the PRC over the 8 testing images is plotted in Figure 5(b). The area below the PRC is the AP value. It shows clear improvements on the AP results from this figure. From this figure, one can also observe that Coupled-CNN_E_A achieves the best precision and recall because the PRC for this method is the highest and longest. Similarly, the precision is improved while preserving the recall by comparing the results in Figure 5(b) via the comparative groups Coupled-CNN_E_A & Coupled-CNN_E and Coupled-CNN_A & FRCN.

Table 6 lists the numerical detection results and average testing time per image for each method under the condition IoU = 0.3. In terms of precision, recall, AP, and $F1$ score, the performance of Coupled-CNN_E_A is superior to the other four methods mentioned. Based on the observations, the following analysis is obtained. (1) The CNN-based methods achieve better performance than CFAR-MS, thereby demonstrating the advantage of the ship detector with deep learning techniques. (2) The new method surpasses FRCN by a noticeable margin due to the specific design in ESPN and ASDN. (3) The computational time needed in the new method is almost identical to that in FRCN. Nevertheless, the CNN-based methods operate considerably 22 times faster than CFAR-MS.

Furthermore, to analyze why the proposed Coupled-CNN_E_A method could achieve such a desirable performance, the corresponding visualization is necessary. Figure 6 presents the detection results in offshore areas of Sentinel-1 images. The green box indicates the correctly detected targets, the red one indicates false alarms, and the blue one represents the ground-truth. In this figure, the upper row (Fig-

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

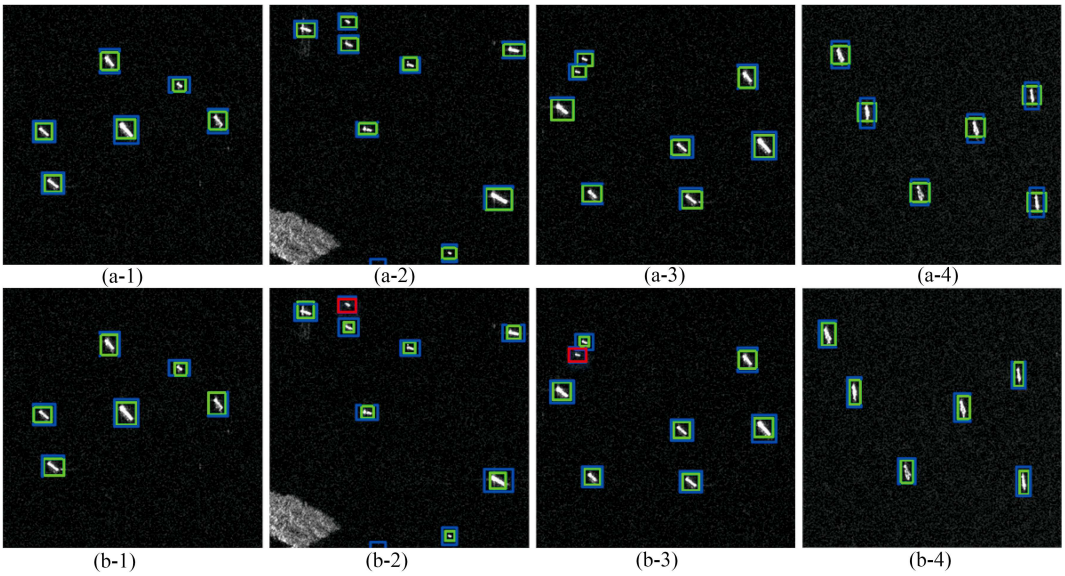

Figure 6 Ship detection results in offshore area of Sentinel-1 images. (a-1)–(a-4) exhibit the visualization results by using the proposed Coupled-CNN_E_A method. (b-1)–(b-4) show the detection results of the CFAR-MS method. The green boxes indicate the correctly detected targets, the red ones indicate false alarms, and the blue ones represent the ground-truth.

ure 6(a-1)–(a-4)) shows the visualization results by using the proposed Coupled-CNN_E_A method, while the lower row (Figure 6(b-1)–(b-4)) exhibits the detection results of the CFAR-MS method. According to the visualized examples, the performance of CFAR-MS is nearly the same as that of Coupled-CNN_E_A in areas where the ships are not very small or densely packed. Specifically, few false alarms of extremely small objects emerged in the visualization result of CFAR-MS.

Figure 7 exhibits a comparative visual effect of the proposed detector and CFAR-MS detector. The selected image contains many small and densely clustered ship targets, and ships in this image face many different conditions, including inshore targets, offshore targets, and targets in inland rivers and other locations. One can easily observe that the missing detections are significantly decreased in Figure 7(a), and the false alarms are slightly less in Figure 7(a) than in Figure 7(b). The visualization performance illustrates a clear improvement of the proposed detector. Therefore, we can conclude the experimental results through the following points:

- In areas filled with small and densely clustered ships, the proposed method can lower the missing detections to a great extent, which is crucial for SAR ship detection;
- For offshore targets, the proposed method can achieve slightly less false alarms and missing detections than CFAR-MS;
- Nearly no false alarms are on land for the proposed detector even without any prior information.

3.2.2 *GF-3 data set*

Figure 8(a) plots the recall versus IoU curve on the GF-3 data set, which illustrates the proposal quality with the proposed Coupled-CNN_E_A method and other four baseline methods. This figure could indicate, for certain IoU between detected boxes and ground-truth regions, how many true positive samples can be fetched. Then, the PRC over the four testing images is shown in Figure 8(b). This figure shows that the proposed Coupled-CNN_E_A method surpasses other baseline methods in terms of precision and recall. Furthermore, the AP value is improved because the area under the PRC is enlarged.

The quantitative detection results are shown in Table 7. The numerical values of recall and precision for the proposed method and the baseline methods are listed to illustrate obvious improvements of the proposed method over other baselines. With the recall and precision available, the comprehensive evaluation metrics of the AP value and $F1$ score are computed, both confirming the clear improvements of the proposed method on the collected and annotated GF-3 data set. The efficiency of the proposed

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

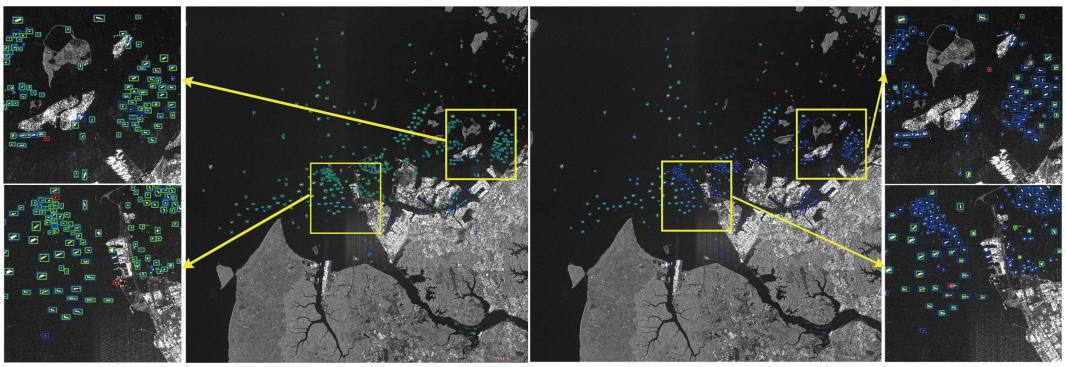

Figure 7 Ship detection results with (a) Coupled-CNN_E_A and (b) CFAR-MS for an image block cropped from the wide-swath Sentinel-1 SAR imagery over the Strait of Malacca, Singapore. In both subfigures, two areas (highlighted by the yellow boxes) are enlarged to exhibit a clear visual effect. The green box indicates the correctly detected targets, the red indicates false alarms, and the blue represents the ground-truth.

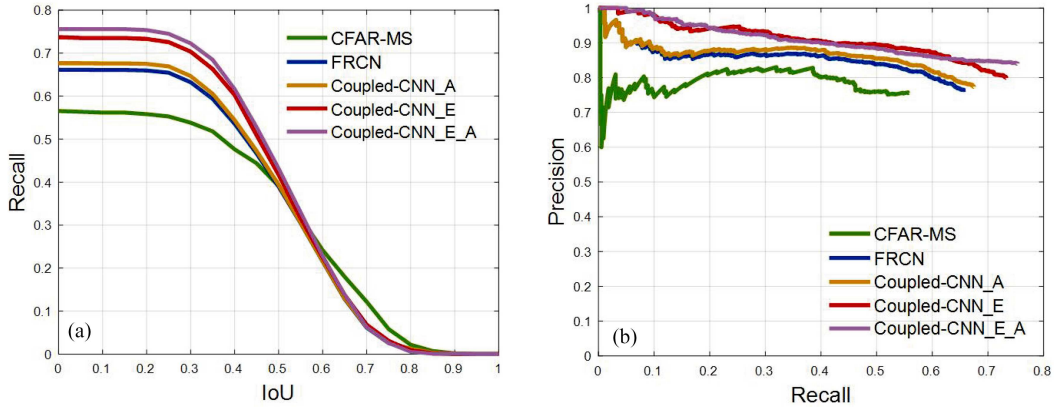

Figure 8 Performance curves over four GF-3 images. (a) Recall vs. IoU curve for each method; (b) precision vs. recall curve for each method.

Table 7 Performance comparison of different methods on GF-3 data set[a)]

Methods	Ground truth	True positive	False positive	Recall	Precision	Average precision	$F1$ score	Average time (s) per image
CFAR-MS	1757	981	316	0.5582	0.7562	0.4832	0.6423	1630
FRCN	1757	1179	373	0.6710	0.7597	0.5772	0.7126	85
Coupled-CNN_E	1757	1210	364	0.7433	0.7906	0.6784	0.7662	86
Coupled-CNN_A	1757	1306	346	0.6887	0.7687	0.5997	0.7265	89
Coupled-CNN_E_A	1757	1324	252	**0.7536**	**0.8401**	**0.6865**	**0.7945**	90

a) The bold numbers denote the optimal values in each column. The underlined numbers denote the suboptimal values in each column.

method is reflected by the average testing time per image in the last column of Table 7, i.e., 18 times faster than CFAR-MS.

Figure 9 shows the detection results on an image with many small and densely clustered ship examples of the proposed Coupled-CNN_E_A method and the CFAR-MS method. The size of the image is 2000×2000 in pixel. The detection results indicate that although most of the ships in this image are small and densely clustered, a desirable performance can be achieved. However, CFAR-MS missed many of them and false alarms remain a challenging problem for CFAR-MS.

By contrast, Figure 10 presents the detection results in offshore areas. The upper row (Figure 10(a-1)–(a-4)) visualizes the detection results of the proposed Coupled-CNN_E_A method. The detection results of the CFAR-MS method are shown in the lower row (Figure 10(b-1)–(b-4)). The detection performance of CFAR-MS is nearly the same as that of the proposed method in offshore areas, except for a few

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

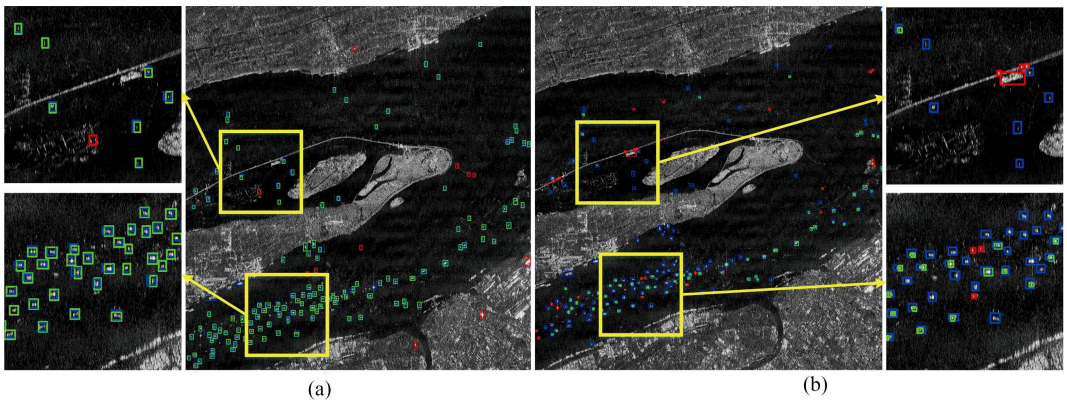

Figure 9 Ship detection results with (a) Coupled-CNN_E_A and (b) CFAR-MS for an image block cropped from the GF-3 SAR imagery. In both subfigures, two areas (highlighted by the yellow boxes) are enlarged for a clear visual effect. The green boxes indicate the correctly detected targets, the red ones indicate false alarms, and the blue ones represents the ground-truth.

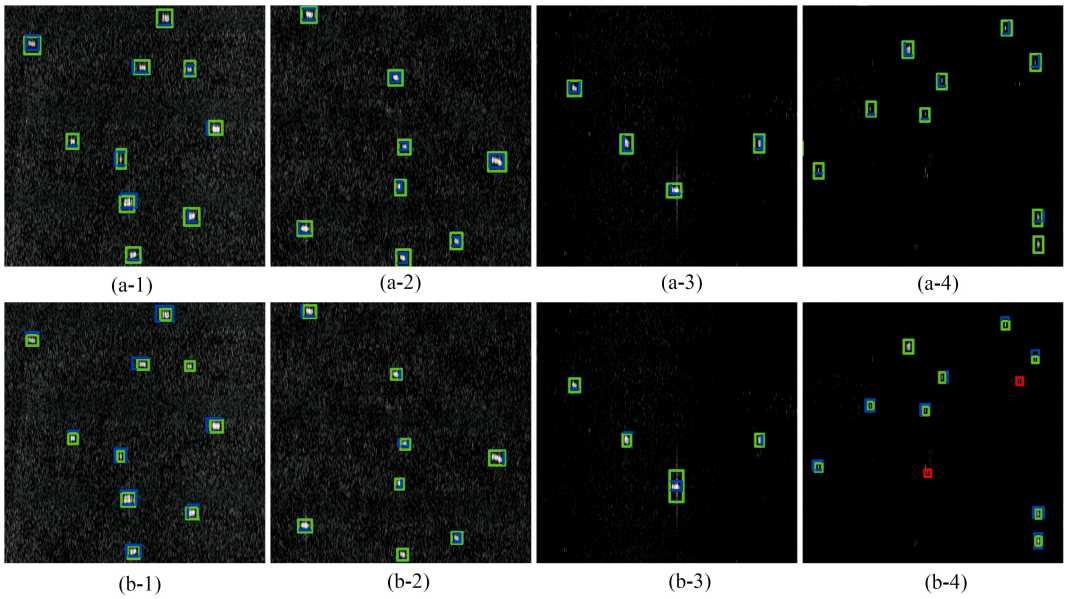

Figure 10 Ship detection results in offshore areas of GF-3 image. (a-1)–(a-4) exhibit the visualization results by using the proposed Coupled-CNN_E_A method. (b-1)–(b-4) shows the detection result of the CFAR-MS method. The green boxes indicate the correctly detected targets, the red ones indicate false alarms, and the blue ones represent the ground-truth.

extremely small ship targets (visualized by red boxes in Figure 10 as false alarms).

4 Conclusion

A coupled CNN-based detector that could detect small and densely clustered SAR ship targets is developed in this study. The special design in ESPN, i.e., sliding windows operate on multiple layers with multiple filter sizes and generate proposals from each sub-branch, enables one to propose ship candidates exhaustively. In ASDN, the joint use of information on the ship-like regions generated from ESPN and their context information further improves the detection performance. Experiments and analyses on the challenging Sentinel-1 data set and the GF-3 data set indicate that, among the CFAR-MS, FRCN, Coupled-CNN_E, Coupled-CNN_A and the new Coupled-CNN_E_A method, the AP value and $F1$ score of the new method are higher by 0.4151 and 0.3126 for the Sentinel-1 data set compared with the CFAR-MS method. The values are verified to be 0.2033 and 0.1522 higher than CFAR-MS for the GF-3 data set.

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

Both the quantitative and qualitative experimental results demonstrate the superiority and practicality of the new method. Our work is expected to focus on sensor-adaptive SAR ship detectors that may enable the natural use of transfer learning. In this case, models trained on one data set can achieve a desirable performance on another data set, which will reduce the limitations of the data-driven detector.

Acknowledgements This work was partially supported by National Natural Science Foundation of China (Grant No. 61331015) and China Postdoctoral Science Foundation (Grant No. 2015M581618). The authors are grateful to thank Prof. T. K. Truong for his helpful comments and suggestions that significantly improved this manuscript.

References

1 Wang S G, Wang M, Yang S Y, et al. New hierarchical saliency filtering for fast ship detection in high-resolution SAR images. IEEE Trans Geosci Remote Sens, 2017, 55: 351–362
2 Gao G, Shi G T. CFAR ship detection in nonhomogeneous sea clutter using polarimetric SAR data based on the notch filter. IEEE Trans Geosci Remote Sens, 2017, 55: 4811–4824
3 Zeng T, Zhang T, Tian W M, et al. A novel subsidence monitoring technique based on space-surface bistatic differential interferometry using GNSS as transmitters. Sci China Inf Sci, 2015, 58: 062304
4 Ma L, Chen L, Zhang X J, et al. A waterborne salient ship detection method on SAR imagery. Sci China Inf Sci, 2015, 58: 089301
5 Crisp D. The state-of-the-art in ship detection in synthetic aperture radar imagery. Org Lett, 2004, 35: 2165–2168
6 Wackerman C C, Friedman K S, Pichel W G, et al. Automatic detection of ships in RADARSAT-1 SAR imagery. Canadian J Remote Sens, 2001, 27: 568–577
7 Ferrara M N, Torre A. Automatic moving targets detection using a rule-based system: comparison between different study cases. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium Proceedings, Seattle, 1998. 1593–1595
8 Wang C L, Bi F K, Zhang W P, et al. An intensity-space domain CFAR method for ship detection in HR SAR images. IEEE Geosci Remote Sens Lett, 2017, 14: 529–533
9 Bi H, Zhang B, Zhu X X, et al. L_1-regularization-based SAR imaging and CFAR detection via complex approximated message passing. IEEE Trans Geosci Remote Sens, 2017, 55: 3426–3440
10 Iervolino P, Guida R, Whittaker P. A novel ship-detection technique for Sentinel-1 SAR data. In: Proceedings of the 5th Asia-Pacific Conference on Synthetic Aperture Radar, Singapore, 2015. 797–801
11 Feng J, Ma L, Bi F K, et al. A coarse-to-fine image registration method based on visual attention model. Sci China Inf Sci, 2014, 57: 122302
12 Wu X M, Du M N, Chen W H, et al. Salient object detection via region contrast and graph regularization. Sci China Inf Sci, 2016, 59: 032104
13 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, 2012. 1097–1105
14 Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intel, 2017, 39: 1137–1149
15 Dai J F, Li Y, He K M, et al. R-FCN: object detection via region-based fully convolutional networks. In: Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, 2016. 379–387
16 Bell S, Zitnick C L, Bala K, et al. Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 2874–2883
17 Li X, Zhao L M, Wei L N, et al. DeepSaliency: multi-task deep neural network model for salient object detection. IEEE Trans Image Process, 2016, 25: 3919–3930
18 Girshick R. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, 2015. 1440–1448
19 Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In: Proceedings of European Conference on Computer Vision, Amsterdam, 2016. 21–37
20 Cai Z W, Fan Q F, Feris R S, et al. A unified multi-scale deep convolutional neural network for fast object detection. In: Proceedings of European Conference on Computer Vision, Amsterdam, 2016. 354–370
21 Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 936–944
22 Xiang Y, Choi W, Lin Y Q, et al. Subcategory-aware convolutional neural networks for object proposals and detection. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision, Santa Rosa, 2017. 924–933
23 Zhai L, Li Y, Su Y. Inshore ship detection via saliency and context information in high-resolution SAR images. IEEE Geosci Remote Sens Lett, 2016, 13: 1870–1874
24 Zhu J W, Qiu X L, Pan Z X, et al. An improved shape contexts based ship classification in SAR images. Remote Sens, 2017, 9: 145
25 Schmidhuber J. Deep learning in neural networks: an overview. Neur Netw, 2015, 61: 85–117

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6

26 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. ArXiv:1409.1556
27 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 580–587
28 Bottou L. Large-scale machine learning with stochastic gradient descent. In: Proceedings of the 19th International Conference on Computational Statistics, Paris, 2010. 177–186
29 Neubeck A, van Gool L. Efficient non-maximum suppression. In: Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, 2006
30 Jia Y Q, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, 2014. 675–678
31 Zuhlke M, Fomferra N, Brockmann C, et al. SNAP (sentinel application platform) and the ESA Sentinel-3 Toolbox. In: Proceedings of Sentinel-3 for Science Workshop, Venice, 2015
32 Pan Z X, Liu L, Qiu X L, et al. Fast vessel detection in Gaofen-3 SAR images with ultrafine strip-map mode. Sensors, 2017, 17: 1578
33 Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, 2007
34 Flach P, Kull M. Precision-recall-gain curves: PR analysis done right. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, 2015. 838–846
35 Qin X X, Zhou S L, Zou H X, et al. A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images. IEEE Geosci Remote Sens Lett, 2013, 10: 806–810

https://engine.scichina.com/doi/10.1007/s11432-017-9405-6