

S-CNN-BASED SHIP DETECTION FROM HIGH-RESOLUTION REMOTE SENSING IMAGES

Ruiqian Zhang^a, Jian Yao^{a,*}, Kao Zhang^a, Chen Feng^a, Jiadong Zhang^b

^a School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, P.R. China

^b China Aerospace Science and Technology Corporation, Beijing City, P.R. China

Email: jian.yao@whu.edu.cn Web: <http://cvrs.whu.edu.cn/>

Commission VII, WG VII/4

KEY WORDS: Ship Detection, Convolutional Neural Networks (CNN), S-CNN, Ship Model Construction

ABSTRACT:

Reliable ship detection plays an important role in both military and civil fields. However, it makes the task difficult with high-resolution remote sensing images with complex background and various types of ships with different poses, shapes and scales. Related works mostly used gray and shape features to detect ships, which obtain results with poor robustness and efficiency. To detect ships more automatically and robustly, we propose a novel ship detection method based on the convolutional neural networks (CNNs), called S-CNN, fed with specifically designed proposals extracted from the ship model combined with an improved saliency detection method. Firstly we creatively propose two ship models, the “V” ship head model and the “||” ship body one, to localize the ship proposals from the line segments extracted from a test image. Next, for offshore ships with relatively small sizes, which cannot be efficiently picked out by the ship models due to the lack of reliable line segments, we propose an improved saliency detection method to find these proposals. Therefore, these two kinds of ship proposals are fed to the trained CNN for robust and efficient detection. Experimental results on a large amount of representative remote sensing images with different kinds of ships with varied poses, shapes and scales demonstrate the efficiency and robustness of our proposed S-CNN-Based ship detector.

1. INTRODUCTION

The detection of inshore and offshore ships has an important significance both in military and civilian fields. It can not only supervise fisheries, but also manage marine traffics to ensure the safety of coast and sea. Traditional ship detection requires manual observation with all kinds of ships, which consumes manpower and material resources greatly. With the development of remote sensing, satellite and artificial intelligence technologies, fast and accurately detecting ships based on remote sensing images has been becoming an urgent issue.

Two kinds of images are widely used for ship detection: synthetic aperture radar (SAR) images (Tello et al., 2005; Xing et al., 2015) and optical remote sensing images (Liu et al., 2014; Xu et al., 2014). The SAR images have a wealth of information on the sea. With the big difference of ships and sea in SAR images, ships can be separated with sea easily (Wang et al., 2014). However, both the low resolution of SAR images and the complicated sea clutter make it much harder to detect small-sized and cluttered ships. Another kind of images is optical ones. With the advancement of satellite technology, high resolution remote sensing images can be provided quickly and easily (Shi et al., 2014). Optical satellite images with high resolutions have more details for observed ships in textures, shapes and edges, which are very important for detecting small-sized and cluttered ships.

Gray intensity and shape features have been widely used to detect ships. The ship detection approaches based on gray intensity statistical characteristics extract ships by segmenting images with gray intensity features. Due to the great gray intensity distinction of background and objects, these approaches can produce good detection results. Chu et al. (2007) used information entropy to quantify this distinction and Zhang et al.

(2010) put morphological contrast into calculation to describe gray feature difference. Huo et al. (2015) presented a saliency analysis in ship detection, with the same principles as the human vision system. The results are stable in calm water conditions and can efficiently deal with offshore ship detection. However, when cloud is covered or large sea wave exists, this approach has more false alarm and leakage phenomenon. Also, it cannot efficiently detect inshore ships often connected or overlapped with lands.

The shape feature based approaches detect ships with rich edge information of ships. Due to the need of ship design consideration, ships are artificially constructed with some specific shapes, such as narrow bow area and parallel hull edges. Liu et al. (2014) applied edge information to robustly detect offshore ships from high resolution remote sensing images. But when extracting an object shape in a complex coastal environment, the detection accuracy based on specific characteristics decreases quickly. To solve this problem, Kun and Xuequan (2004) proposed an approach with prior knowledge of harbor in the land. It can separate the harbor area in the land and the ocean one, which result in that the ship targets located in the ocean area can be detected quickly and accurately. However, for densely distributed ships, how to remove shadows between ships and the land area with good segmentation is very difficult. To detect ships more automatically, Xu et al. (2014) introduced robust invariant generalized Hough transform (RIGHT) into shape extraction by a robust shape model automatically learned via iterative training. Although it can get more widely applicable features from training, RIGHT cannot represent most of ship characteristics with only shape feature.

In summary, all these methods take significant efforts on ship detection in optical images. However, some issues still exist, like how to integrate more ship features into detection, and how to detect all kinds of inshore and offshore ships together more automatically and quickly. This challenge also happens in other object detection or recognition issues. To solve these problems,

*Corresponding author

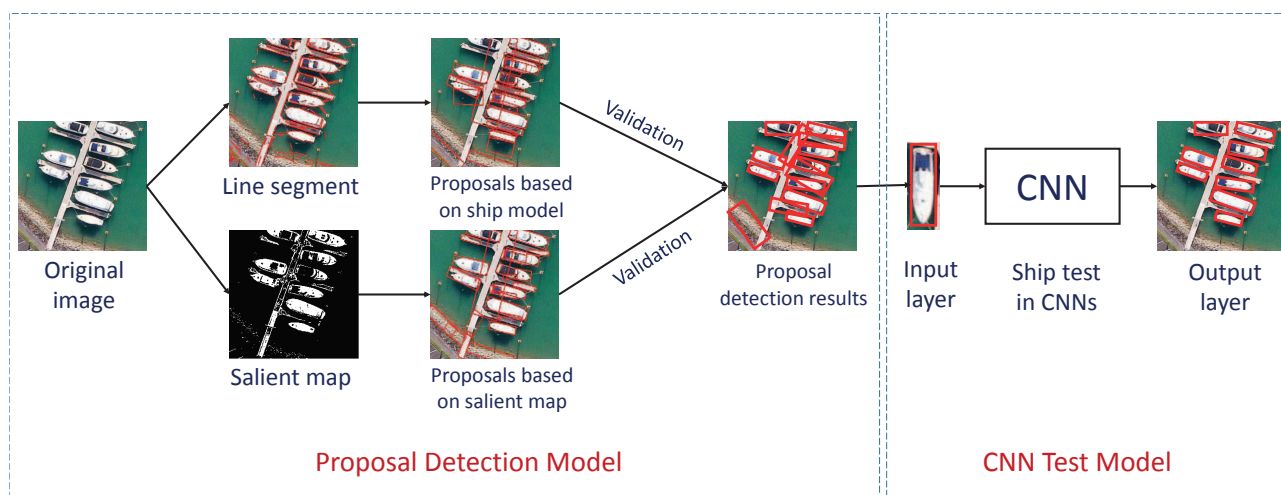


Figure 1: The flowchart of our proposed ship detection system.

a lot of machine learning technologies had been put into target detection, like neural network (Roth, 1990) and support vector machine (SVM) technology (Burges, 1998). With the rise of deep learning, scientific researchers pay more attention on object detection by convolutional neural networks (CNNs) (Hu et al., 2015). It can not only deal with large scale images, but also train features automatically with high efficiency. However, in ship detection domain with high resolution remote sensing images, there is still a lack of evaluation experiments based on CNNs.

The concept of CNN was inspired by (Fukushima, 1980), who presented “neocognitron” as a hierarchical and sift-invariant model, which is an early attempt in pattern recognition. Neocognitron puts one vision system into many subschemes and uses multi-level cascade feature plane for recognition. Whereafter, Lcun et al. (1998) proposed to train CNN by the backpropagation algorithm and firstly obtained feature recognition results. In the past two decades, CNN had been gradually ignored with the rise of other efficient machine learning techniques, such as boosting and support vector machine (SVM). While with the breakthrough in (Krizhevsky et al., 2012), people focused on CNN once again. CNN can achieve a big success with a large-scale dataset. During Large Scale Visual Recognition Challenge 2013 (ILSVRC2013), the top five rankers with minimum error ratios were all achieved based on CNN. It can not only deal with large scale images, like remote sensing image, but also train features automatically with high efficiency.

Based on the advantages of CNNs, we propose to incorporate the high-capacity CNNs to automatically extract ship features for detection in this paper. Connected with CNN requiring a large number of ship samples in training part, more kinds of ship features can be put in the feature model. Detection requires localizing objects within a test image, which has a big difference with image classification. A lots of approaches did localization by using a sliding-window detector, which has been used widely, typically on constrained object categories, such as faces (Baluja et al., 1998) and pedestrians (Sermanet et al., 2013). In our detection application, the ships varied in types (such as barges and warships), shapes, and poses, which make the use of the sliding-window detector very challenged and even impossible due to the need of a large amount of length-width ratios for sliding windows. Inspired with Regions with CNN feature approach (R-CNN) (Girshick et al., 2014) that combines region proposals with CNNs, we propose an integrated ship detection system that bridges the gap between ship proposal extraction and CNNs, named S-CNN. Our proposed system consists of

two modules, as illustrated in Figure 1. The first module aims to extract ship proposals from a high-resolution remote sensing image with two effective methods, the ship model based detection and saliency based one. Two ship models, the “V” ship head model and the “|” ship body one, are proposed to localize the ship proposals from the line segments extracted from a test image, which can produce stable results for both inshore and offshore ships with enough large sizes and clear boundaries. For ships with small sizes and/or without clear boundaries due to the ship wave, the saliency based method is more efficient. The second module focuses on the training and detection of CNN. The CNN model is trained from a large set of training ship samples in advance. The ship proposals extracted in the first module are fed into the trained CNN for efficient detection. Experimental results on a large amount of representative remote sensing images with different kinds of ships with varied poses, shapes and scales demonstrate the efficiency and robustness of our proposed S-CNN-Based ship detector.

The remainder of this paper is organized as follows. The proposed two ship proposal detection methods are detailedly described in Section 2. The training and detection of CNN are presented in Section 3. Experimental results on a large amount of representative high-resolution remote sensing images are presented in Section 4 followed by the conclusions drawn in Section 5.

2. SHIP PROPOSAL DETECTION

2.1 Preprocessing

To make the ship detection more efficiently, we apply a series of preprocessing operations on a test image, which are comprised of denoising, enhancement and segmentation. The image denoising operation reduces the noises caused by optical image sensors and/or atmospheric radiation and the image enhancement increases the image contrast. These two operations can make ships in the image more prominent for detection. The common image denoising methods are median filtering, spatial low-pass filtering, local smoothing and so on. While image enhancement uses some operator, such as Sobel, Prewitt, Roberts and Laplace operators. Rough image segmentation is used to quickly separate the sea area and the land one based on their color difference. In addition, via image segmentation, some individual ships with clear sea background can be efficiently segmented into one region, which makes the offshore detection more easy. In our system, we used

median filtering and Laplace operator for image denoising and enhancement. The test image was segmented using the Edge Detection and Image Segmentation (EDISON) system by the Mean Shift method (Comaniciu and Meer, 2002). The Mean Shift achieves the image segmentation results by clustering gradually, which can separate ships and sea accurately, especially in an offshore ships image.

2.2 Simple Feature Extraction

Simple feature extraction is significant for efficiently detecting valid ship proposals. By extracting simple features, line segments and saliency, the expression of ship shape and texture features are convenient to be described. Line segment features are the prior step for detecting ship proposals based on two ship models, the ship “V” head model and the ship “||” body one. While the saliency feature is needed in the saliency detection.

Two highly efficient and advanced line segment detectors, LSD (Grompone et al., 2010) and EDLines (Akinlar and Topal, 2011), can be efficiently used for extracting line segment features in our application. LSD improves the algorithm proposed by Burns et al. (1986) based on the Helmholtz principle, which combines phase grouping. EDLines extracts the line segments from the edge map based on the least square algorithm and the minimum error principle, which has a high precision in a short time.

Saliency feature extraction aims to achieve a similar function of human eye observation system. The famous ITTI salient detection model (Itti and Koch, 2000) integrates gray feature, direction and color information to construct basic physiological visual similar features. However, the results with the orientation channel usually turns out to be less useful for salient object segmentation (Frintrop et al., 2015) since it assigns high saliency values to object edges, and makes object segmentation difficult. Thus we extract the same saliency features: color and intensity ones with (Frintrop et al., 2015) to generate the salient map. We extract color features with the intensity and newly defined RG and BY color channels, which are calculated as $I = (R + G + B)/3$, $RG = R - G$ and $BY = B - (R + G)/2$, respectively. Thus, the salient map equation is defined as follows:

$$S(\mathbf{p}) = g(I(\mathbf{p}) + RG(\mathbf{p}) + BY(\mathbf{p})), \quad (1)$$

where $S(\mathbf{p})$ represents the salient map value at the point \mathbf{p} in the image and $g(\cdot)$ denotes a combination function of three channels. In our system, we calculated three channels with equal weights, so the salient map value for a pixel is obtained by simple addition of three channel values, i.e., $S(\mathbf{p}) = (I(\mathbf{p}) + RG(\mathbf{p}) + BY(\mathbf{p}))/3$.

2.3 Ship Proposal Detection

Based on the above extracted two kinds of features, line segment and saliency ones, we combine two approaches to detect ship proposals. The first one is the ship model based detection originally and specifically designed for our application. In most of ships, there exist two significant characteristics, V-shaped ship head and parallel ship body, which is described as the “V” ship head model and the “||” ship body one, as illustrated in Figures 2(a) and (b), respectively. As the “V” ship head model shown in Figure 2(a), we search the line segments around the ship head candidate points, and calculate the angle of the nearby line segments. If the acute angle formed by two line segments meets some condition, i.e., less than some predefined threshold, these two line segments will be considered as a ship head candidate. Since ship heads may be confused with some land structures or

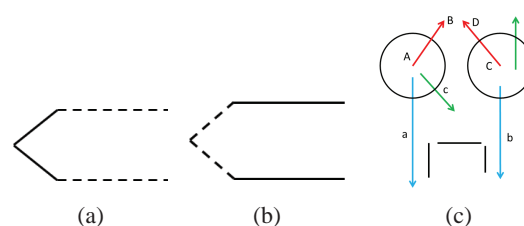


Figure 2: Illustration of ship models: (a) the “V” ship head model; (b) the “||” ship body model; (c) the ship model candidates.

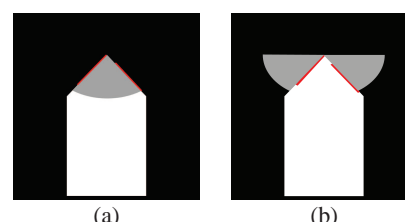


Figure 3: An illustration of two mask regions for ship validation: (a) the sector mask region M_S ; (b) the semi-circle mask region M_C .

other nearby ships, we propose to further validate each ship head candidate with at least one ship body line segment existed close to the head candidate within some suitable range and angle with respective to the two line segments of the ship head.

Due to the existence of many other non-ship constructions or even ship internal structures with multiple lines similar as the ship head structure, we need to further validate the ship head candidates. Figure 4(e) shows a plausible ship head candidate formed by a ship body line segment and a land one. Figure 4(i) shows another plausible ship head candidate formed by two line segments extracted from ship internal structures instead of from the ship boundary. To eliminate such these plausible candidates, the segmentation mask image for non-sea areas can be sufficiently utilized for validation. Firstly we define two mask regions, the sector mask region M_S and the semi-circle one M_C , as illustrated in Figure 3. The sector mask region M_C is a fan-shaped one with some radius formed by two line segments of a ship head candidate, as shown in Figure 3(a). While, the semi-circle mask region M_C is formed by the semi-circle with the same radius as that of M_C followed by the removal of M_C from the semi-circle region, as shown in Figure 3(b). In Figure 4(e), a ship head candidate is formed by a real ship head line segment and a land one. Most of pixels in this angle region fall in the sea area so the area of the sector mask region M_S is close to zero, which can be efficiently utilized to eliminate such these plausible candidates. While, in Figure 4(i), a plausible ship head candidate is formed by two line segments from ship internal structures. We observe that the area ratio of M_S and M_C is much smaller than that in the real ship head candidate as shown in Figure 4(a). Naturally, we further validate those ship head candidates based on the following two conditions:

$$\begin{cases} Area(M_S) > T_1, \\ \frac{Area(M_S)}{Area(M_C)} > T_2, \end{cases} \quad (2)$$

where T_1 and T_2 are two thresholds adaptively set according to the resolution of the test image.

As the “||” ship body model shown in Figure 2(b), we search two approximately parallel line segments as a ship body candidate. Similarly, we further validate each ship body candidate with at least one ship head line segment existed close to the body

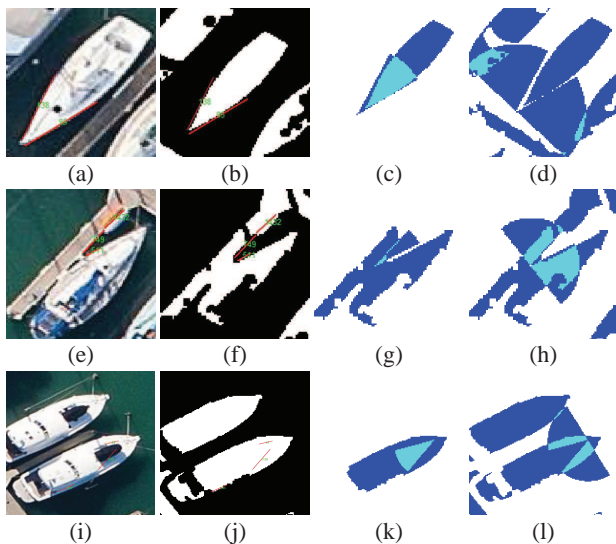


Figure 4: Examples of ship head candidate validation. (a), (e) and (i) show the partial regions of a test remote sensing image containing ship candidates. (b), (f) and (j) show the segmentation binary results of (a), (e) and (i), respectively. In (c), (g) and (k), the cyan regions are the overlaps of their section mask regions M_S and the connected regions containing ship candidates, respectively. In (d), (h) and (l), the cyan regions are the overlaps of their semi-circle mask regions M_C and their segmentation images, respectively.

candidate within some suitable range and angle with respect to the two line segments of the ship body. Figure 2(c) presents an illustration of the ship candidates. Two red vectors named AB and CD are met with the ship head model, and a and b in blue are two line segments satisfy the ship body model. However, the black line segments are not met with the ship body model due to that they are too short and the candidate comprised of two green vectors c and d is eliminated by the angle constraint.

Based on the saliency map of different characteristics extracted from a test image, we propose to detect ship proposals from salient image regions from the saliency map. By simply thresholding the saliency map with some suitable threshold, we get a salient binary image, from which we collect separated blobs as the ship proposal candidates. To suppress the noise, those blobs without an enough saliency value, which is the sum of all pixel saliency values in that blob, will be eliminated.

2.4 Multi-Scale Searching

Different kinds of ships have different shapes and sizes, and present details to various extent in images captured at a high resolution. With too much details in a high resolution remote sensing image as shown in the Figure 5(a), a lot of small and fragmented line segments were detected, it is hard to discern the ships out from the detected line segments. In order to efficiently detect ships with different sizes from a high resolution remote sensing image, we need to collect all the ship candidates in multiple scales as the ship proposals. A high resolution test image is sequentially down-sampled with different down-sampling scales, which results in an image pyramid. When we only consider to detect the warships with relatively fixed large sizes, we just search the candidates from some suitable low-resolution down-sampled image. If we expect to detect all kinds of ships with different levels of sizes, we need to search all candidates from the original image and all the down-sampled ones. Figure 5(b) shows an example of the low-resolution image

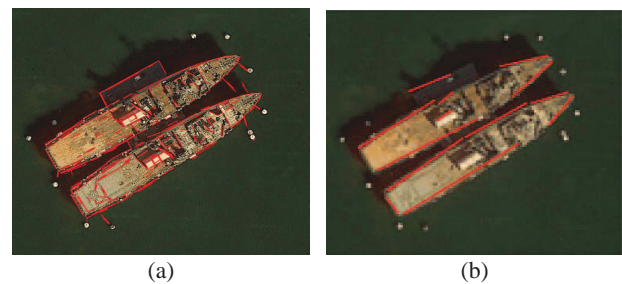


Figure 5: An illustration of line segments extracted from a warship image in different resolutions: (a) in the original resolution; (b) in the low resolution with a down-sampling scale factor of 1/4.

down-sampled from the original image shown in Figure 5(a) with a down-sampling scale factor of 1/4. From Figure 5, we observe that there exist a lot of small and fragmented line segments in the original image, which are not suitable for figuring out the warships based on the ship models, however, just several main line segments are extracted from its corresponding low-resolution image, which are much easier to form the ship candidates from them.

2.5 MBR Adjustment

After we detect out the ship proposals based on the ship models and the saliency map, the boundaries of proposals are extracted by their minimum bounding rectangles (MBRs) on the segmented binary mask image. The MBR is obtained totally depending on the segmentation result in preprocessing. Due to inaccurate segmentation, some ship pixels are not included in the MBRs of ships. To generate more accurate ship proposals, we need to adjust their MBRs, which will be fed into the later CNN. For the ship proposals determined by the ship models, we can adjust the MBRs by line segments of ship heads or bodies.

In the “V” ship head model, the symmetrical axis of the ship body can be figured out by equally splitting the acute angle formed by two straight line segments of the ship head. If the body orientation of a MBR and the symmetrical axis of the ship is almost consistent, the MBR has the right way. However, if the body orientation of the MBR is far away from the symmetrical axis, we need to adjust the MBR direction to make them consistent. In the “||” ship body model, we figure out the symmetrical axis from the almost parallel line segments of the “||” ship body in the same way by splitting the acute angle formed by those parallel line segments. We adjust the MBR of a ship proposal as the same for the “V” ship candidates.

Specifically, for the “V” ship head model, we calculate the inclination angles of two ship bow line segments, from which the direction of bisector line of “V” can be obtained. Let α be the inclination angle of the bisector line, which is the correct main direction of the MBR, and β be the original MBR inclination angle. The MBR adjustment is implemented by rotating the inclination angle of the MBR main direction from β to α . We do the same operation for the “||” ship body model. As for an initial vertex \mathbf{x} of the MBR, the corresponding coordinates of rotated MBR’s vertex \mathbf{x}' can be calculated as:

$$\mathbf{x}' = \mathbf{R}(\mathbf{x} - \mathbf{x}_c) + \mathbf{x}_c, \quad (3)$$

where \mathbf{x}_c denotes the coordinate centroid of the original MBR and the \mathbf{R} stands for the 2D rotation matrix defined by the rotation angle $\theta = \alpha - \beta$. Figure 6 shows two examples for the MBR adjustment, from which we observe that the MBRs after adjustment more accurately cover the whole ship bodies.

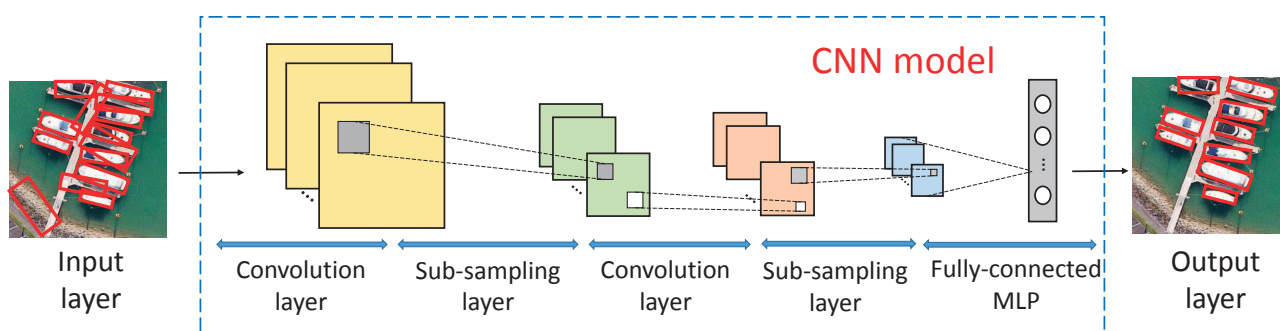


Figure 7: The flowchart of the CNN-based ship detection system.

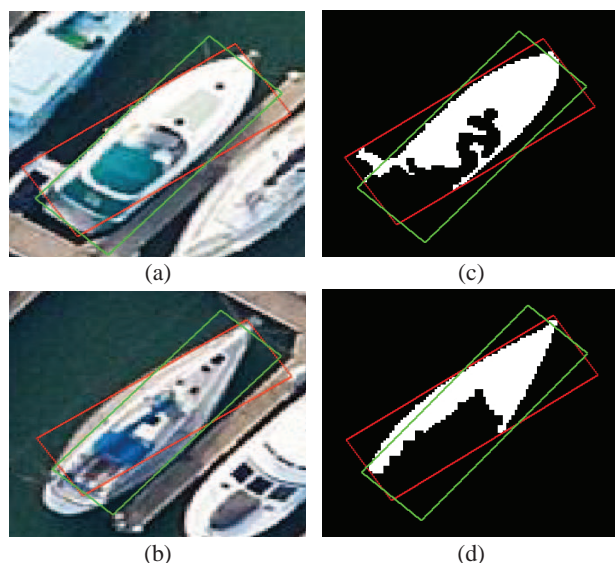


Figure 6: Two examples of MBR adjustment: (a)-(b): original remote sensing images; (c)-(d): the corresponding connected mask region of (a) and (b), respectively. The red rectangles stand for the original MBRs obtained from the segmentation mask image while the adjusted MBRs are marked in green rectangles.

3. SHIP DETECTION WITH S-CNN

3.1 Introduction to CNN

The popular approach of training a multi-layer neutral network is the back-propagation algorithm, in which, there are an input layer, multiple hidden ones and an output one. With each 2D input vector fed to the input layer, each output result of the input unit is equal to the corresponding component of the input vector. In general, each output layer can be given by (Duda et al., 2000):

$$X_j^l = f\left(\sum_{i=1}^d X_i^{l-1} \omega_{ij}^l + b_j^l\right), \quad (4)$$

where i is an index value of the input layer unit, j is an index value of the hidden layer unit, ω_{ij}^l denotes the weight between the input layer unit i and the hidden layer unit j , b_j^l denotes additive bias given by each layer map, and $f(\cdot)$ expresses the output activation function.

As a kind of neutral network based on multi-layer networks, CNN also has the input, hidden and output layers. In order to efficiently deal with large-scale data, CNN usually has a deep neutral network with plenty of hidden layers and parameters. In traditional neutral networks, training with so many parameters

will be a big challenging, so CNN introduces three ideas to reduce the number of parameters: local receive fields, shared weights, and multiple convolution kernels. Local receive fields mean to change the neuron sensory area from global to local. As human recognition to outside world, the spatial relation of an image is closely related to local pixels, and those far away always have weak correlation. The conversion from global to local can cut down the number of parameters greatly. The other way introduced by CNN is to share weights, which is inspired by the theory that humans can use the feature from one part to another in an image when the statistical properties of two parts are similar. Therefore, making a convolution between learned features based on small parts of the image and the original large-sized image, a different feature activation value for any position in this image can be acquired. Feature activation values and learned features from small parts can replace features learned from all images, which can quickly decrease the number of parameters. However, when parameters are cut down greatly, a new problem is arose, i.e., only one convolution kernel cannot take sufficient features. The third idea comes up to solve this problem: multiple convolution kernels. The CNN derives multiple kernels for image convolution and generates multiple convoluted images which can be seen as different channels of the image. Based on these three ideas and the neutral network layer calculation approach, the convolution layer can be calculated as follows (Bouvier, 2006):

$$X_j^l = f\left(\sum_{i \in M_j} X_i^{l-1} \otimes k_{ij}^l + b_j^l\right) \quad (5)$$

where M_j represents a selection of input maps, k_{ij} is convolution kernels between the input layer unit i and the hidden layer unit j , and the symbol ' \otimes ' means operation of convolution. Compared Eq. (5) with Eq. (4), we can see that for a particular output map, the input maps will be convolved with different kernels, which is an improvement in CNN. The l -th convolution layer can be obtained from the $(l-1)$ -th layer with Eq. (5).

With three above introduced ideas, the feature training in CNN can derive on the right track, but it may also generate too many features, which possibly result in that the over-fitting problem comes out with feature classification. Thus the sub-sampling layer is proposed. Just like a natural idea to aggregate the characteristics of different positions when people try to describe a large image, the sub-sampling is another kind of pooling to collect statistically with a part of image's feature. It is also a good way to reduce the computation cost and to gradually build up spatial invariance. The sub-sampling layer is calculated as follows:

$$X_j^l = f\left(B_j^l D(X_j^{l-1}) + b_j^l\right) \quad (6)$$

where $D(\cdot)$ represents a sub-sampling function and B_j^l is a mul-

multiplicative bias while b_j^i is an additive bias. Both multiplicative bias and additive one are all corresponding to each output layer.

With the above mentioned ideas and core equations calculating the convolution layer and the sub-sampling one, the convolutional neural network becomes more integrated. Figure 7 shows the flowchart of the CNN-based ship detection system, from which, the commutative construction of the convolution layer and the sub-sampling layer is clear.

3.2 Detection with S-CNN

For our specific detection task, we first train the CNN model for ship detection from a large dataset and then test the trained CNN model on the ship proposals extracted by the ship models and the saliency map. In this paper, we call this specific CNN model as S-CNN.

To achieve a good S-CNN model widely suitable for efficiently detecting different kinds of ships with different shapes and different sizes, we need to collect enough samples, especially positive ones for training. We firstly collected a large set of high-resolution images containing different kinds of ships and then marked all the positive images manually. In general, two kinds of object samples can be put into CNNs. The first one is comprised of small scale images drawn from original ones, which just contain one object in each image. Another consists of large scale images in which there exists one or multiple objects marked manually in an image. In our training system, we manually marked the second of ship samples for training. A large amount of negative ship samples can be automatically collected from those images containing marked ships by ensuring that the collected negative samples have no enough overlap with marked positive ones.

After training the S-CNN model is finished, the ship proposals extracted by the ship models and the saliency map will be fed into S-CNN for testing. In this procedure, we compute the CNN features of each proposal to derive the similarity of candidate proposals with ship features, whose details can be seen in (Girshick et al., 2014). Proposals extracted by our approaches make the S-CNN test more pertinent because these proposals tell us which areas have ship targets with high possibilities. It also solves the uncertain length-width ratio problem in the sliding-window method with higher speed and efficiency. While training features and detecting ships just with CNN having features extracted by machine (deep learning) with generous samples generate more preeminent results in wide applicability.

Completed by the CNN test, each ship proposal has an activation value. Due to that ship proposals may be overlapped, we apply a non-maximum suppression operation to eliminate the proposals with low activation values.

4. EXPERIMENTAL RESULTS

To completely test our proposed S-CNN-based method to detect ships from high-resolution optical remote sensing images, we individually evaluated our proposed ship proposal extraction method and the whole ship detection performance with the trained S-CNN model.

4.1 Evaluation on Ship Proposal Extraction

In this experiment evaluating the performance of our proposed ship proposal extraction method, the test data set we used is comprised of high-resolution optical remote sensing images

collected from Google Earth with a preferred resolution of $0.12m$ and with a preferred resolution of $0.25m$ containing some warship objects. These remote sensing images were all collected by Landsat-7, QuickBird, IKONOS and SPOT5, which are from city harbors or naval military bases where a lot of ships or/and warships gathered. In order to test the effectiveness of our proposed ship proposal extraction method, lots of images containing different types of ships with different sizes and shapes were picked out for testing. This collected test data set is divided into three categories. The first category of images contains inshore ships connected with the land area, as shown in Figure 8(a). The second one includes far apart offshore ships as shown in Figure 8(g). The third one contains large-scale warships with a large length-width ratio as shown in Figure 8(m). All these test images were clipped with a same resolution of 2000×2000 pixels.

In this experiment, we extracted the line segments by ED-Lines (Akinlar and Topal, 2011) with their recommended parameters. In the ship model based proposal extraction and validation parts, the parameters were set empirically. The two thresholds T_1 and T_2 in Eq. (2) were set as 0.4 and 0.6, respectively. In the salient map extraction part, we weighted the I , RG and BY channels equally. The calculated saliency maps were thresholded by a fixed value of 70. In total, our ship proposals were extracted from four-layer image pyramid. In general, warship proposals can be more efficiently extracted with a down-sampling scale factor of 1/4.

Figure 8(b), 8(h) and 8(n) show extracted line segments from three representative example images by EDLines, from which we observed that our used line segment detector is very efficiency, which extracted most of line segments in object edges and textures, like ship body edges “|” structures and head line segments “V” shapes, which are necessary for applying the ship model to successfully extract the ship proposals. Compared with inshore ships, offshore ones and warships, EDLines has more line segment details on land in offshore ships and warships, and it illustrates that the extraction results of inshore ships as shown Figure 8(d) and warships as shown in Figure 8(p) with the ship model method have a better effect than those of offshore ships as shown in Figure 8(j). Figure 8(c), 8(i) and 8(o) illustrate the salient maps of these images, from which the ships are obviously observed with high saliency values. Due to that some land areas have the high saliency values too, the offshore ship detection without cluttered lands as shown in Figure 8(i) have better proposal extraction results, compared with those shown in Figure 8(c) and 8(o). Therefore, the combination of the ship model and saliency methods creates a preeminent performance to extract ship proposals.

4.2 Evaluation on Ship Detection

Our proposed S-CNN-based detection algorithm was implemented using the open source CNN, which is depended on the Caffe CNN library Jia et al. (2014). Totally, we trained the proposed S-CNN model with 1010 high-resolution remote sensing positive images containing 5270 ship positive samples with varied types, sizes and shapes, and 1576 negative images without any ships as the negative samples. Averagely, there are 5.2 ships in each positive image, which is in accordance with the general standard in deep learning. The used non-maximum suppression value in post-processing after detection was set as 30%, and the activation value of the ship estimation was set as 0.7. Some S-CNN detection results after CNNs testing are shown with blue and green rectangles in Figure 9 where the ground truth that we labeled manually is shown with red rectangles. Sometimes, the

positively detected ships perfectly fit the ground truth. In these cases, just red rectangles are visible. Green rectangle means the wrong detected results. In Figure 8, the first row of images are detection results with inshore ships. Almost all detection results are accurate, which demonstrates that our algorithm can detect inshore ships very well. And it produced the same good results in the second row of offshore ships. In the specially selected third row of images, there exist some wrongly detected ships. The wrong detection mostly appears in the land area, where some objects, like rooftop, have similar structures with ship textures. This shortcoming can be improved by selecting enough negative samples, especially similar to the real ships in the local structures, for training to some extent. In some images especially with different-sized ships, omission detection happens occasionally, which indicates that our proposal detectors need to be further improved.

Both precision and recall are often used to quantitatively evaluate detection performance. In the test data set containing 3545 inshore ships and 1520 offshore ships, our detection results are listed in Table 1 where N_g represents the number of ground truth, N_t is the number of ships validated by S-CNN, and N_r represents the number of correctly detected ships. With these numbers, the recall and precision can be calculated as:

$$\begin{cases} \text{Recall} = N_r/N_t, \\ \text{Precision} = N_r/N_g. \end{cases} \quad (7)$$

From the results shown in Table 1, the recall and precision of inshore and offshore ships are both above 90%. Specifically, a high precision of 99.1% was achieved for offshore ships.

Table 1: Precision and recall of our proposed S-CNN-based ship detector on different data sets.

	N_g	N_t	N_r	Recall (%)	Precision (%)
Inshore ships	3545	3731	3399	91.1	95.9
Offshore ships	1520	1539	1506	97.9	99.1

5. CONCLUSION

In this paper, we proposed a novel S-CNN-based ship detection framework for high-resolution optical remote sensing images. To sufficiently train the proposed S-CNN model, we collected 5270 positive samples with different types, sizes, and shapes from more than one thousand high-resolution remote sensing images and a large amount of negative samples. The whole detection framework is comprised of two stages. In the stage, we creatively extracted ship proposals based on two ship models, the “V” ship head model and the “|” ship body one, formed by line segments and the saliency map. The minimum bounding rectangles (MBRs) of these ship proposals, which were calculated from the segmentation binary map, were further refined based on extracted line segments. In the second stage, we fed the extracted ship proposals with refined boundaries into the trained S-CNN model for validation. Experimental results demonstrate that our proposed S-CNN-based method can provide enough proposals with a very high recall and improves the ship detection accuracy compared with the traditional R-CNN method.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 91438203, the Open Research Fund of The Academy of Satellite Application under grant NO. 2014_CXJJ-YG_13, and the National Natural Science Foundation of China (Project No. 41571436).

References

- Akinlar, C. and Topal, C., 2011. (edlines): A real-time line segment detector with a false detection control. *Pattern Recognition Letters* 32(13), pp. 1633–1642.
- Baluja, S., Rowley, H. A. and Kanade, T., 1998. Neural network-based face detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 20(1), pp. 23–38.
- Bouvier, J., 2006. Notes on convolutional neural networks. *Neural Nets*.
- Burges, C. J., 1998. A tutorial on support vector machines for pattern recognition. *Data Mining And Knowledge Discovery* 2, pp. 121–167.
- Burns, J. B., Hanson, A. R. and Riseman, E. M., 1986. Extracting straight lines burns. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 8(4), pp. 425C455.
- Chu, Z., Wang, Q., Chen, H. and Xu, S., 2007. Ship auto detection method based on minimum error threshold segmentation. *Computer Engineering* 33(11), pp. 239–241.
- Comaniciu, D. and Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, pp. 603–619.
- Duda, R. O., Hart, P. E. and Stork, D. G., 2000. Pattern Classification (2nd Edition). *Wiley-Interscience*.
- Frintrop, S., Werner, T. and Garcia, G., 2015. Traditional saliency reloaded: A good old model in new shape. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Fukushima, K., 1980. Neocognitron : A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36(4), pp. 193–202.
- Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Grompone, v. G. R., Jakubowicz, J., Morel, J. M. and Randall, G., 2010. LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Software Engineering* 32(4), pp. 722–732.
- Hu, F., Xia, G. S., Hu, J. and Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*.
- Huo, L., Jiao, L., Wang, S. and Yang, S., 2015. Object-level saliency detection with color attributes. *Pattern Recognition*.
- Itti, L. and Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research* 40, pp. 1489–1506.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. *Eprint Arxiv* pp. 675–687.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E., 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25, pp. 2012.
- Kun, C. and Xuequan, C., 2004. A new ship and dock target segmentation method based on geometry feature. *Computer Engineering & Applications* 41(31), pp. 197–200.
- Leun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), pp. 2278–2324.
- Liu, G., Zhang, Y., Zheng, X., Sun, X., Fu, K. and Wang, H., 2014. A new method on inshore ship detection in high-resolution satellite images using shape and context information. *IEEE Geoscience & Remote Sensing Letters* 11(3), pp. 617–621.

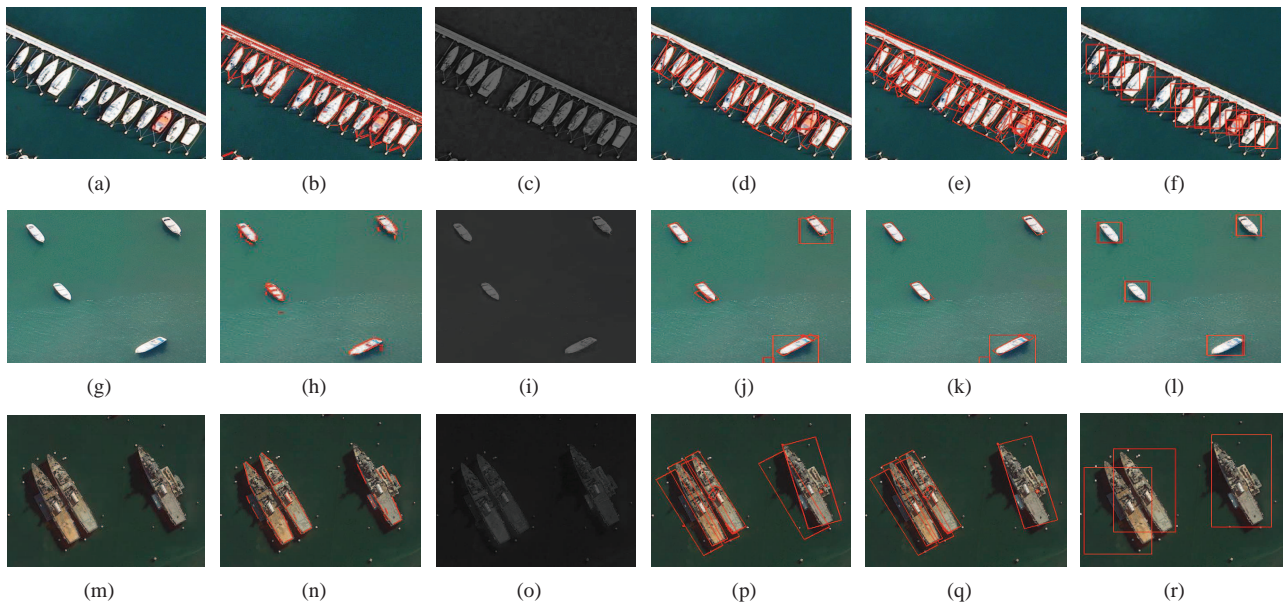


Figure 8: The results of ship detection system process. (a), (g) and (m) are images with inshore ships, offshore ships and warships respectively. (b), (h) and (n) show the line segment results and (c), (i) and (o) are images with results of salient map. In (d), (j) and (p), the ship proposal detection results based on ship model method are showed, and in (e), (k) and (q), ship proposal detection are labeled with salient map. (a), (g) and (m) are ship detection final results.

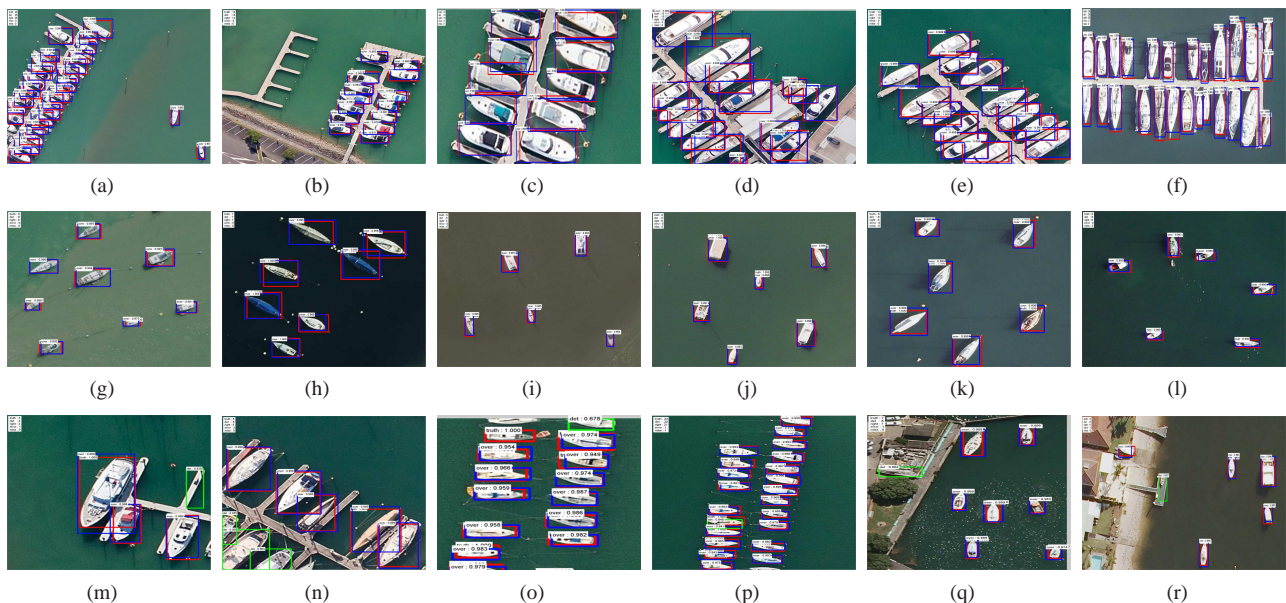


Figure 9: The results of our proposed ship detection system. (a) - (f) are inshore ship detection results, and (g) - (l) show some offshore ship detection results. (m) - (r) are special selection results which have error or lack detection.

Roth, M. W., 1990. Survey of neural network technology for automatic target recognition. *IEEE Transactions on Neural Networks* 1(1), pp. 28–43.

Sermanet, P., Kavukcuoglu, K., Chintala, S. and LeCun, Y., 2013. Pedestrian detection with unsupervised multi-stage feature learning. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shi, Z., Yu, X., Jiang, Z. and Li, B., 2014. Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. *IEEE Transactions on Geoscience & Remote Sensing* 52(8), pp. 4511–4523.

Tello, M., Lopez-Martinez, C. and Mallorqui, J. J., 2005. A novel algorithm for ship detection in SAR imagery based on the wavelet transform. *IEEE Geoscience & Remote Sensing Letters* 2(2), pp. 201–205.

Wang, C., Jiang, S., Zhang, H., Wu, F. and Zhang, B., 2014. Ship detection for high-resolution SAR images based on feature analysis. *IEEE Geoscience & Remote Sensing Letters* 11(1), pp. 119–123.

Xing, X., Ji, K., Kang, L. and Zhan, M., 2015. Review of ship surveillance technologies based on high-resolution wide-swath synthetic aperture radar imaging. *Journal of Radars* 4(1), pp. 107–121.

Xu, J., Sun, X., Zhang, D. and Fu, K., 2014. Automatic detection of inshore ships in high-resolution remote sensing images using robust invariant generalized Hough transform. *IEEE Geoscience & Remote Sensing Letters* 11(12), pp. 2070–2074.

Zhang, Y., Xu, Z., Chen, W. and Peng, X., 2010. Faint targets detection based on retinex theory and gray morphological filtering method. *Infrared Technology* 32(1), pp. 46–51.