Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region Convolutional Neural Network

XUE YANG [1,2], (Student Member, IEEE), HAO SUN[1], XIAN SUN[1], MENGLONG YAN[1], ZHI GUO[1], AND KUN FU [1,2,3]

[1]Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China
[2]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[3]Institute of Electronics, Chinese Academy of Sciences, Suzhou 215123, China

Corresponding author: Kun Fu (fukun@mail.ie.ac.cn)

ABSTRACT Ship detection is of great importance and full of challenges in the field of remote sensing. The complexity of application scenarios, the redundancy of detection region, and the difficulty of dense ship detection are all the main obstacles that limit the successful operation of traditional methods in ship detection. In this paper, we propose a brand new detection model based on multitask rotational region convolutional neural network to solve the problems above. This model is mainly consisting of five consecutive parts: dense feature pyramid network, adaptive region of interest (ROI) align, rotational bounding box regression, prow direction prediction and rotational nonmaximum suppression (R-NMS). First of all, the low-level location information and high-level semantic information are fully utilized through multiscale feature networks. Then, we design adaptive ROI align to obtain high quality proposals which remain complete spatial and semantic information. Unlike most previous approaches, the prediction obtained by our method is the minimum bounding rectangle of the object with less redundant regions. Therefore, the rotational region detection framework is more suitable to detect the dense object than traditional detection model. Additionally, we can find the berthing and sailing direction of ship through prediction. A detailed evaluation based on SRSS for rotation detection shows that our detection method has a competitive performance.

INDEX TERMS Convolutional neural network, remote sensing, ship detection.

I. INTRODUCTION

With the development of remote sensing technology, high-resolution remote sensing images can be easily obtained. Automatic ship detection has been playing a significant role in the field of remote sensing for a long time and has made a great progress in promoting national defense construction, port management, cargo transportation, and maritime rescue [1]. Simultaneously, the information of ship's berthing and sailing direction are also of huge significance. However, the characteristics of the large aspect ratio make ship detection become more difficult than other object detections, such as vehicles [2]–[6], buildings [7]–[12] and aircrafts [13]–[18]. What's more, the complexity of application scenarios, the redundancy of detection region, and the difficulty of dense ship detection have posed a great challenge for ship detection and direction prediction.

In recent years, deep learning has achieved great success in computer vision [19]–[24]. Much attention has been paid to object detection based on deep learning methods and has made great achievement. Region proposals with convolutional neural network (RCNN) [25] provide an excellent pipeline for object detection. Although RCNN has some obvious defects in computation speed and storage space, its detection results are far better than traditional detection methods. Fast-RCNN [26] significantly improve the efficiency of detection and effectively reduce the storage by using shared computing. Faster-RCNN [27] adopts a trainable region proposal network (RPN) instead of Selective Search method to achieve end-to-end training while improving detection efficiency and accuracy. It is consist of two stage: region proposal and region classification.

IEEE *Access*

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

The methods above are known as horizontal region detection which are suitable for natural scene detection but not for satellite remote sensing ship detection. In satellite remote sensing images, the ships have a large aspect ratio and are often densely arranged in complex scenes. Once the ship is inclined, the redundant regions of the horizontal bounding box and the regions of overlap between the ships will be relatively large. The disadvantage of this situation is obvious and disastrous. Specifically, complex scenes often contain many noise objects, which greatly affect the performance of the ship detection. In addition, large redundant regions introduce a lot of noise, causing the feature information to be interfered or even submerged. As shown in Fig. 1(a), a large object overlap region causes the object to be discarded after the operation of nonmaximum suppression (NMS). To address these problems above, we propose a new, end-to-end, rotational-region-based object detection framework for ship detection in high-resolution satellite images which can handle different complex scenarios, detect intensive objects, and reduce redundant detection regions, as illustrated in Fig. 1(b). Moreover, our framework can predict the berthing and sailing direction of ship, which cannot be achieved by the horizontal region detection method.
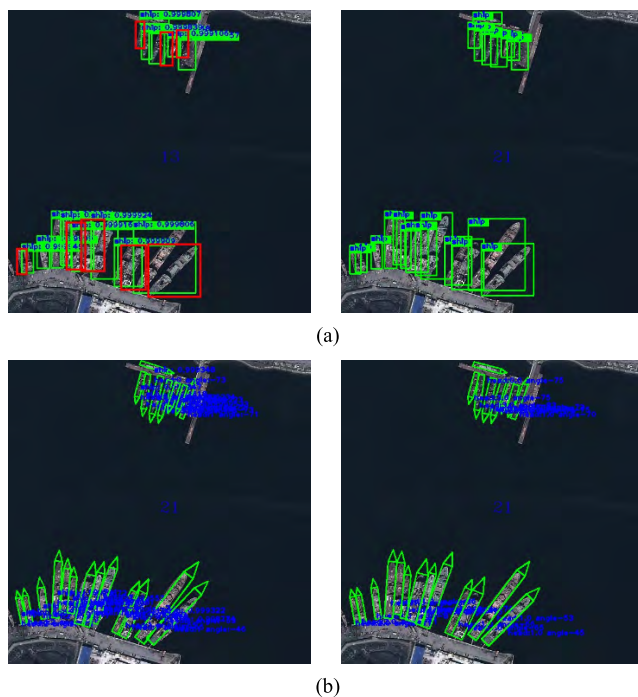

FIGURE 1. Rotational region detection algorithm perfectly solves the problem of traditional detection algorithm in dense object detection. The green, red bounding boxes represent predictions boxes and missing prediction boxes respectively. (a) Horizontal region detection. (b) Rotational region detection.

Our framework mainly is consist of five consecutive parts: Dense Feature Pyramid Network (DFPN), adaptive region of interest (ROI) Align, rotational bounding box regression, prow direction prediction and rotational nonmaximum suppression (R-NMS). Compared with detection methods based

on convolutional neural network (CNN), our framework is more suitable for ship detection and has achieved more promoting performance.

The main contributions of this paper are as follows:

1) DFPN
We use a new multiscale feature fusion network called DFPN, which can effectively integrate the low-level location information and high-level semantic information to provide more advanced features for object detection.

2) ADAPTIVE ROI ALIGN
Adaptive ROI Align is proposed in this paper to mitigate the effects of redundant noise regions in the proposals and keep the completeness of semantic and spatial information.

3) PROW DIRECTION PREDICTION
The berthing and sailing direction of ship can be found through prediction. This method is simple but effective, with a high prediction accuracy.

4) R-NMS
In order to obtain more accuracy prediction results, we propose R-NMS which has more stringent constraints.

The rest of this paper is organized as follows. Section II we briefly review related work on object detection. Section III introduces the details of the proposed method. Section IV presents experiments conducted on a remote sensing dataset to validate the effectiveness of the proposed framework. Finally, section V discusses and concludes the results of the proposed method.

II. RELATED WORK
Ship detection has been investigated by a wide variety of methods in recent years. In this section, we briefly review the existing machine-learning-based ship detection algorithm and deep-learning-based ship detection algorithm.

In the past few years, some machine-learning-based methods have been proposed for ship detection [28]–[32]. Yu *et al.* [33] and Zhu *et al.* [34] propose features of texture and shape by sea-land segmentation, then an algorithm such as the contrast box algorithm or semi-supervised hierarchical classification is used to get the candidate object region. Bi *et al.* [35] use a bottom-up visual attention mechanism to select prominent candidate regions throughout the detection scene. Yang *et al.* [36] propose a novel detection framework via sea surface analysis to solve the task of automatic ship detection in high-resolution optical satellite images with various sea surfaces. This proposed method first use two new features to analyze whether the sea surface is homogeneous or not. Then, they propose a linear function combining pixel and region characteristics to select ship candidates. Finally, false alarms are filtered by adopting compactness and length-width ratio. Shi *et al.* [37] present a method to detect ships in a ''coarse-to-fine'' manner. Specially, they convert an optical image into a hyperspectral form by adopting

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN
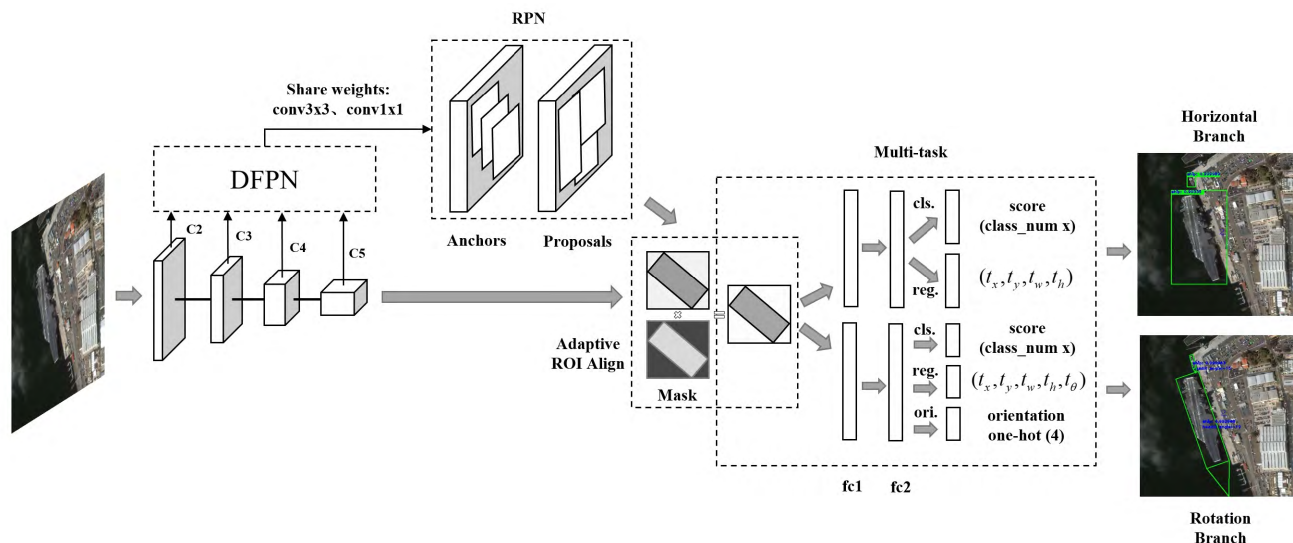
IEEE *Access*

FIGURE 2. Overall framework of rotational region ship detection. This framework mainly consists of five consecutive parts: DFPN, adaptive ROI Align, rotational bounding box regression, prow direction prediction and R-NMS.

anomaly detector and local shape features, then extract ships through hyperspectral algorithm. Corbane *et al.* [38] present a complete processing chain for ship detection based on statistical methods, mathematical morphology and other signal-processing techniques such as the wavelet analysis and Radon transform. This automatic ship detection model is used to complement existing regulations, especially the fishing control system.

Although these machine-learning-based ship detection algorithm above have shown promising performance, they have poor practicability in complex scenarios. With the application of deep CNN in object detection [39]–[43], deep-learning-based ship detection algorithm are also widely used in remote sensing ship detection. Kang *et al.* [44] take the objects proposals generated by Faster R-CNN for the guard windows of CFAR algorithm, then pick up the small objects, thus reevaluating the bounding boxes which have relative low classification scores in detection network. Zhang *et al.* [45] propose a new ship detection model based on CNN which is called SCNN, fed with specifically designed proposals extracted from the ship model combined with an improved saliency detection method. Kang *et al.* [46] build a contextual region-based CNN with multilayer fusion for SAR ship detection, which is an elaborately designed deep hierarchical network, and composed of a RPN with high network resolution and an object detection network with contextual features. Tang *et al.* [47] adopt compressed domain for fast ship candidate extraction, while DNN is exploited for high-level feature representation and classification, and ELM is used for efficient feature pooling and decision making.

These methods above are mostly based on horizontal region detection. Detection results tend to have very large redundant regions, and is not conducive to NMS operation. We propose a novel object detection model based on

multitask rotation region CNN which effectively integrates the low-level location information and high-level semantic information. Meanwhile, this method mitigate the effects of redundant noise regions in the proposals and get rotational bounding box with prow direction. Compared with other deep-learning-based ship detection framework, our method can achieve state-of-the-art detection performance, even in dense scenes.

III. PROPOSED METHOD

The overall framework of our rotational region ship detection method is illustrated in Fig. 2. DFPN, adaptive ROI Align, rotational bounding box regression, prow direction prediction and R-NMS are the five important components of our method. Firstly, DFPN [1] is an effective multiscale feature fusion network which enhances feature propagation, encourages feature reuse, and ensures the effectiveness of detecting multiscale objects. Then, we get proposals from the RPN to provide high-quality region proposals for the next stage. In order to keep the completeness of semantic and spatial information, we design an adaptive ROI Align to mitigate the effects of redundant noise regions in the proposals. Furthermore, compared with traditional framework, second stage of our model has horizontal and rotational branches that respectively predict the horizontal bounding box and the rotational bounding box. Meanwhile, the rotation branch can also predict the berthing and sailing direction of ship. Finally, we use R-NMS which has more stringent constraints so as to obtain the final prediction.

A. DENSE FEATURE PYRAMID NETWORK

The low-level location information and high-level semantic information are very important to object detection. The feature pyramid [32] is an effective multiscale method to
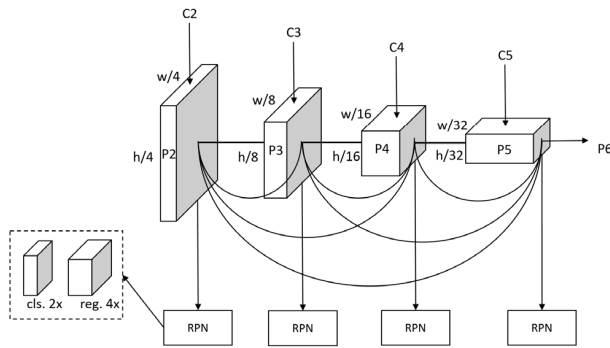
IEEE *Access*

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

FIGURE 3. A multiscale feature pyramid connection. Each feature map is densely connected, and merged by concatenation.

fuse multilevel information, so we adopt a multiscale feature pyramid connection which is called DFPN [1]. Fig. 3 depicts the structure of this densely connected multiscale pyramid feature fusion network.

In this paper, we use ResNet [48] as backbone and choose the last layer of each residual block as the feature maps $\{C_2, C_3, C_4, C_5\}$ in the bottom-up feedforward network. According to the residual network structure, the strides of each feature map correspond to $\{4, 8, 16, 32\}$ pixels. In the top-down network, we get higher resolution features $\{P_2, P_3, P_4, P_5\}$ through lateral connections and dense connections. We set the number of channels for all feature maps to 256 so as to reduce the number of parameters. The specific definition is as follows:

$$P_5 = Conv_{1\times1}(C_5)$$

$$P_i = Conv_{3\times3}\left[\sum_{j=i+1}^{5} Upsample(P_j)\oplus Conv_{1\times1}(C_i)\right] \quad (1)$$

where P_i is the fused feature map corresponding to C_i. $Conv_{k\times k}(.)$ represents the convolution operation, and k is the size of the convolution kernel. $Upsample(.)$ represents nearest neighbor up-sampling in this paper. $\oplus$ represents the operation of concatenation.

We assign five scales $\{32, 64, 128, 256, 512\}$ pixels to $\{P_2, P_3, P_4, P_5, P_6\}$ respectively (P_6 is simply a stride two subsampling of P_5). Taking into account of the characteristics of ships, the ratios of anchor are $\{1:7, 1:5, 1:3, 1:2, 1, 2, 3, 5, 7\}$. Each feature point for each feature map will generate 9 anchors (1 × 9), 36 outputs (4 × 9) for each regression layer, and 18 outputs (2 × 9) for each classification layer. A large number of experiments show that DFPN has a great feature of fusion, and significantly improve the detection performance

B. ADAPTIVE ROI ALIGN

The large aspect ratio is a major feature of the ship. However, once the ship is inclined, the redundant regions of the proposal are relatively large. A lot of noise will reduce the quality of feature extraction, or even submerged features.

Fig. 4(a-c) shows three methods for obtaining fixed-length feature vectors: ROI Align [1], RROI Align [49], and adaptive ROI Align. It is obvious that ROI Align is accompanied by a lot of noise, causing the target features to be overwhelmed. Although RROI removes all the noise through affine transformation, it loses the spatial information of the object. We designed Adaptive ROI Align, a method that automatically filters noise regions by introducing a mask. This mask is trainable and is obtained by convolving the proposals. Adaptive ROI Align makes the spatial information be retained, while leaving a small amount of noise to improve the stability of the network.

Fig. 4(d) visualizes the three methods for obtaining fixed-length feature vectors. As we can see, coast is the main noise in the image, ROI Align can't completely remove it. RROI Align lost spatial information and produced the feature deformation at the same time, which are not conducive to angle regression and prow direction prediction in the second stage. The adaptive ROI Align better solves the problem of the method above, and have access to high-quality feature maps.

C. PROW DIRECTION PREDICTION

As shown in Fig. 5(a), we use the four parameters $(x_{min}, y_{min}, x_{max}, y_{max})$ to represent the horizontal rectangle in the first stage, and use the five parameters (x, y, w, h, θ) to represent the arbitrary rotation rectangle in the second stage.

Considering the range of angles is $[-90, 0)$, we can see that rotation angle θ cannot represent the prow direction of the object, but the prow direction is certainly in the direction of the four sides of the minimum bounding rectangle. In the light of these facts, we label the four sides of the rotation bounding box counterclockwise, as shown in Fig. 5(b). Meanwhile, we predict the berthing and sailing direction of the ship in the rotation branch.

Fig. 6 shows the detection results of four different scenarios, we have achieved amazing performance with such simple predictions. Part of the wrong prediction is to judge the stern of the prow. In our opinion, this network first learned that the prow must be in the long edge direction and then judged which side of the long edge is the prow.

D. MULTITASK LOSS FUNCTION

During training of the RPN, each anchor is assigned a binary class label and five parametric coordinates. The feature maps were input to the RPN network through a 3 × 3 convolutional layer, followed by two sibling 1 × 1 convolution layers for regression and classification. We need to find positive and negative samples from all anchors, which we call a mini-batch. The positive sample anchors need to satisfy the following conditions: the Intersection-over-Union (IoU) overlap between an anchor and the ground-truth is greater than 0.6. The negative samples are defined as: IoU overlap less than 0.25. The total number of positive and negative samples is 256, the ratio is 0.5. Similar to the RPN stage, the second stage classifies each proposal and assign five and four parametric coordinates to regress the final bounding box
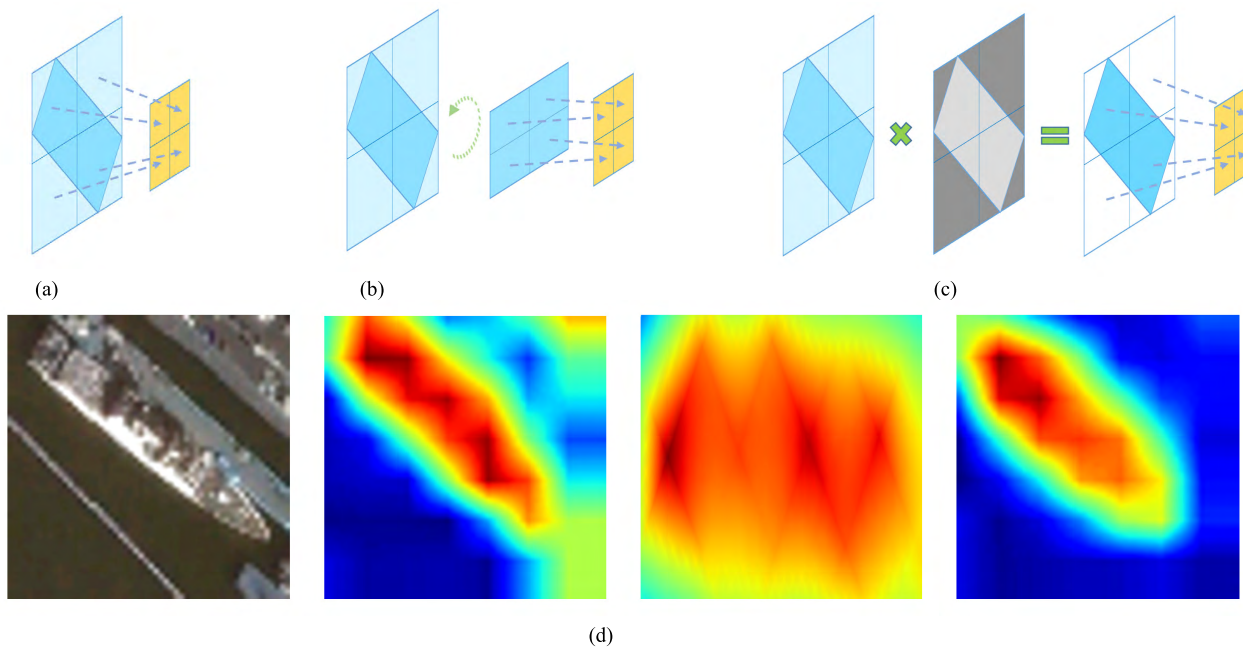
X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

IEEE*Access*

(a)   (b)   (c)

(d)

FIGURE 4. Structure and visualization of three Align methods. (a) ROI Align. (b) RROI Align. (c) Adaptive ROI Align. (d) Visualization of the three methods for obtaining fixed-length feature vectors.
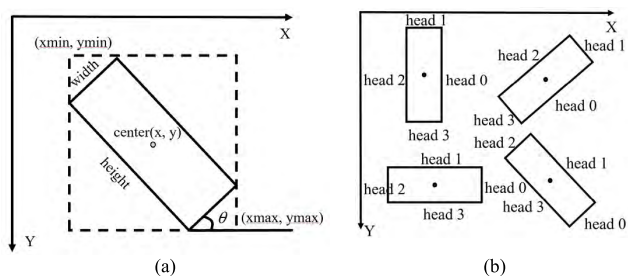

(a)   (b)

FIGURE 5. Representation of horizontal bounding box, rotational bounding box and prow direction. (a) Representation of horizontal bounding box and rotational bounding box. (b) Prow direction.

in rotational branch and horizontal branch respectively. The ratio of positive and negative samples in mini-batch is 0.5, the threshold is 0.5 and the total number of samples is 128.

After adding the angle information, rotational bounding box can locate the object more accurately. The regression of rotational bounding box is defined as follows:

$$t_x = (x - x_a)/w_a, \quad t_y = (y - y_a)/h_a,$$
$$t_w = \log(w/w_a), \quad t_h = \log(h/h_a),$$
$$t_\theta = \theta - \theta_a + k\pi/2 \tag{2}$$
$$t_x^* = (x^* - x_a)/w_a, \quad t_y^* = (y^* - y_a)/h_a,$$
$$t_w^* = \log(w^*/w_a), \quad t_h^* = \log(h^*/h_a),$$
$$t_\theta^* = \theta^* - \theta_a + k\pi/2 \tag{3}$$

where x, y, w, and h denote the box's center coordinates and its width and height. Variables x, x_a, and x^* are for the predicted box, anchor box, and ground-truth box, respectively

(a)   (b)

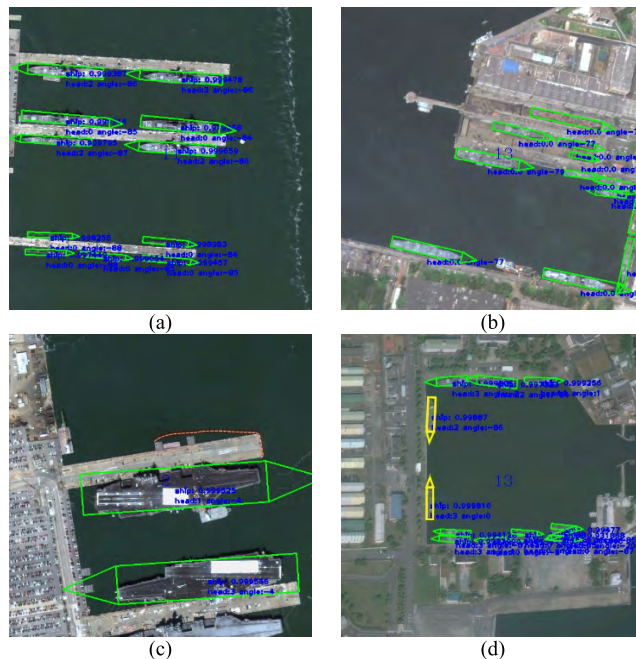(c)   (d)

FIGURE 6. Detection results with prow direction. The green, yellow bounding boxes represent predictions boxes and error prow direction prediction boxes respectively. (a) Scene one: The ships are arranged side by side but in different directions. (b) Scene two: The ships berth in the harbor and dock. (c) Scene three: Large objects. (d) Scene four: Small objects.

(likewise for y, w, h). The parameter $k \in Z$ keeps θ in the range $[-90, 0)$. In order to keep the bounding box in the same position, w and h need to be swapped when k is an odd number.

IEEE *Access*

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

We use multitask loss to minimize the objective function, which is defined as follows:

$$L(p_i, l_i, u_i^*, u_i, v_i^*, v_i, h_k^*, h_k)$$

$$= \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, l_i)$$

$$+ \lambda_1 \frac{1}{N_{reg-h}} \sum_j p_j L_{reg-h}(u_j^*, u_j)$$

$$+ \lambda_2 \frac{1}{N_{reg-r}} \sum_k p_k L_{reg-r}(v_k^*, v_k)$$

$$+ \lambda_3 \frac{1}{N_{reg-r}} \sum_k p_k L_{reg-r}(h_k^*, h_k) \qquad (4)$$

where l_i represents the label of the object, p_i is the probability distribution of various classes calculated by the soft-max function, u_i, v_i represent the predicted parameterized coordinate vectors, u_i^*, v_i^* represent the offset of ground-truth. h_k^*, h_k represent the prow direction of ground-truth and prediction respectively. The hyper-parameter λ_1, λ_2, λ_3 in (4) control the balance between the four task losses; all experiments use $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = 10$ in this paper. In addition, the functions L_{cls} and L_{reg} are defined as:

$$L_{cls}(p, l) = - \log pl \qquad (5)$$

$$L_{reg}(t_i^*, t_i) = smooth_{L_1}(t_i^* - t_i) \qquad (6)$$

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

E. ROTATIONAL NONMAXIMUM SUPPRESSION

NMS is to obtain high quality bounding boxes with small IoU overlap. When ships are densely arranged, the traditional NMS often faces such a dilemma that the bounding box has a large IoU overlap. Therefore, IoU computation on axis-aligned bounding box may lead to an inaccurate IoU of skew interactive bounding box and further ruin the bounding box prediction. An implementation for Skew IoU computation [49] with thought to triangulation is proposed to deal with this problem.

The sensitive relationship between IoU overlap and rotation angle often affect the detection results. For example, for a ship with aspect ratio of 1: 7, the IoU is only 0.38 when the angles differ by 15 degrees. Therefore, we design R-NMS which has two constraints: (a) preserve the prediction results with IoU less than 0.7; (b) if the IoU is in the range of [0.3, 0.7], discard the prediction results that the angle difference is greater than 15°.

IV. EXPERIMENTS

In this section, we will introduce our dataset first. Then we present several groups of comparative experiments to explore the detection performance of the proposed framework. All experiments are conducted on a computer with an NVIDIA GeForce GTX 1080 GPU, and 8 GB of memory.

A. DATASET AND SETTINGS

In the experiments, we evaluate the proposed framework on one data sets designed for remote sensing image rotation region detection, one of which is a satellite remote sensing ship image dataset (SRSS) that we have collected and labeled.

SRSS is collected publicly from Google Earth with 50 large scene images sized $10,000 \times 10,000$ pixels, covering 25 square kilometers. The resolution of satellite remote sensing images is 0.5 meters. In addition, these satellite remote sensing images have the tri-band information (include red, green, and blue) after geometric correction. Geotif is the format of satellite image with latitude and longitude information. The images contain scenes of civilian ports, naval base, offshore areas, and far seas. The annotation content is a set of contour points starting from the prow. We divide the images into $1,000 \times 1,000$ subimages with an overlap of 0.4, then filter out images that do not contain ships, resulting in about 8000 final images. Meanwhile, the ratio of training set to test is 1:4.

All experiments were implemented on the deep learning framework, tensorlfow [50]. We use the pretraining model ResNet-101 to initialize the network. For SRSS dataset, we train a total of 40 k iterations, with a learning rate of 0.001 for the first 20 k iterations, 0.0001 for the next 10 k iterations, and 0.00001 for the remaining 10 k iterations. The optimizer chosen is MomentumOptimizer. Furthermore, we flip the image randomly in the training process, while subtracting the mean value [103.939, 116.779, 123.68] which comes from ImageNet [22]. Subtracting mean can centralize all dimensions of the input data, which is conducive to model training.

B. EVALUATION INDICATORS

To quantitatively evaluate the performance of different framework in object detection, we use the precision–recall curve (PRC), and F-measure (F1), which are three well-known and widely applied standard measures approaches for comparisons [51].

PRC is obtained from four well-established evaluation components in information retrieval, true positive (TP), false positive (FP), false negative (FN), and true negative (TN) [46]. TP and FP indicate the number of correct predictions and error predictions. FN is the sum of regions not proposed. Based on these four components, we provide the definition of precision and recall rate as:

$$Precision = \frac{TP}{(TP + FP)} \qquad (8)$$

$$Recall = \frac{TP}{(TP + FN)} \qquad (9)$$

F1 is a statistic that is commonly used in the field of object detection. The higher the F1 value, the better the performance. The definition is as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (10)$$

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

IEEE *Access*

TABLE 1. Performance of different feature maps combination strategies.

Combination Strategies	RECALL (%)	Precision (%)	F1 (%)
P3	72.7	80.1	76.2
P2+P3	75.0	80.0	77.4
P3+P4	81.6	82.1	81.8
P4+P5	75.5	80.8	78.1
P3+P4+P5	84.7	84.2	84.4
P2+P3+P4+P5	**85.2**	**84.5**	**84.9**

TABLE 2. Performance of three methods for extracting fixed-length feature vectors.

Methods	RECALL (%)	Precision (%)	F1 (%)
ROI Align (R-DFPN)	83.7	82.4	83.0
RROI Align (RRPN)	81.6	83.2	82.4
Adaptive ROI Align	**85.2**	**84.5**	**84.9**

TABLE 3. Detection performance of five methods.

Methods	RECALL (%)	Precision (%)	F1 (%)	Time (s)
Faster-RCNN	75.9	88.3	81.7	0.1
FPN	77.0	**89.3**	82.7	0.15
FPN-Soft-NMS	80.5	87.6	83.9	0.15
LSTM-Based	76.4	82.4	79.3	**0.05**
Ours	**85.2**	84.5	**84.9**	0.16

C. EVALUATION OF DFPN

As we all know, low-level feature semantic information is relatively scarce, but the object location is accurate. On the contrary, high-level feature semantic information is rich, but the object location is relatively rough. Therefore, the choice of feature maps is particularly important. In this section, we chose six different feature maps combination strategies to explore the impact on the detection performance. The specific combination strategies are shown in Table 1. As we can see, the model achieved the worst detection performance when using only the P3 feature map. Furthermore, the combination of P3+P4 is significantly better than the combination of P2+P3 and the combination of P4+P5. This is due to the fact that most of the ships in our data match the anchors in P3 and P4. What's more, the P2 layer is mainly used for small object detection, P5 layer for large object detection. When using all feature maps, the detection performance is optimized: 85.2% for Recall, 84.5% for Precision, and 84.9% for F-measure.

In summary, multiscale detection networks are significantly better than single-scale detection networks, especially in the detection of small objects. Only make full use of effective fusion of various layers of feature information, can we achieve better results.

D. EVALUATION OF ADAPTIVE ROI ALIGN

Due to the large number of redundant regions in ship detection, the ultimate detection performance is often compromised. Fig. 6 shows three methods for extracting fixed-length
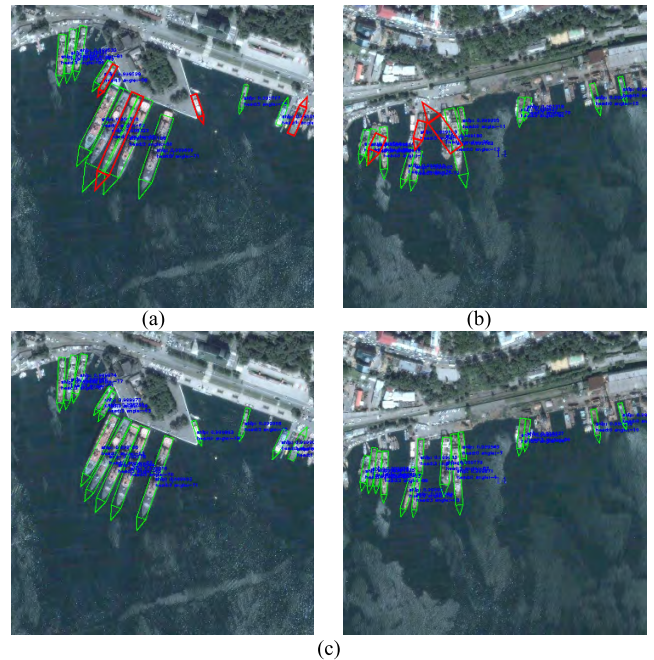

(a)

(b)

(c)

FIGURE 7. Detection results of three Align methods. The red boxes and green boxes represent mission prediction boxes and correct detection boxes respectively. (a) ROI Align. (b) RROI Align. (c) Adaptive ROI Align.
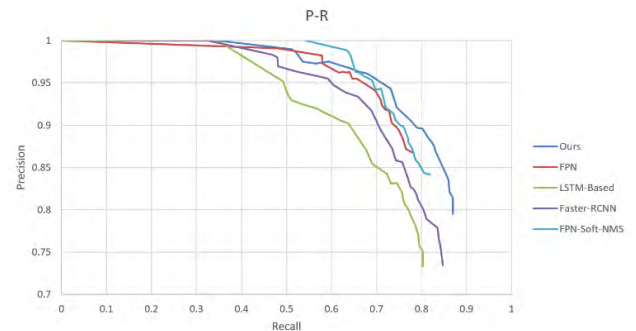

FIGURE 8. The P-R curves of different methods. The proposed method has the state-of-the-art performance.
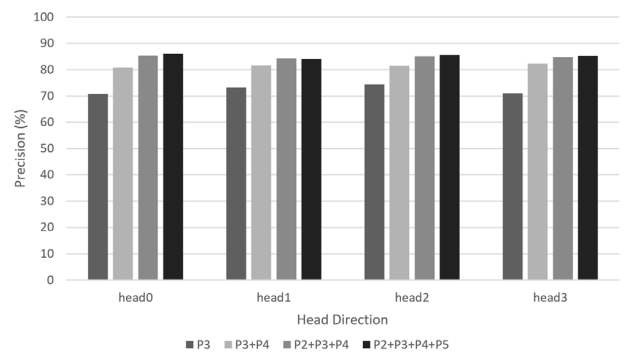

FIGURE 9. Accuracy of prow direction prediction.

feature vectors, while comparing their differences and their advantages and disadvantages. In this section, we will conduct specific experiments on these three methods, and parametrically compare the performance among them.
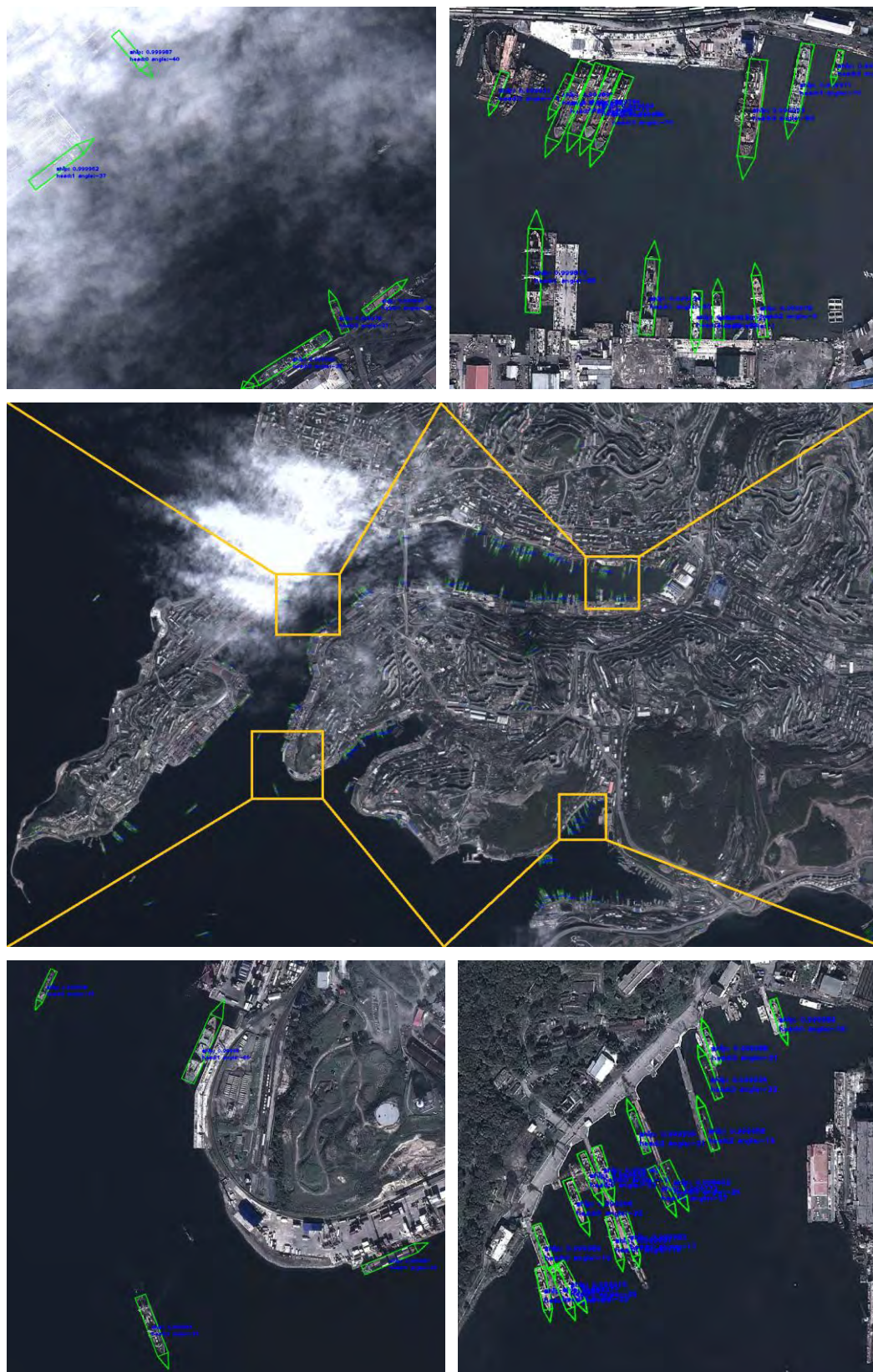
IEEE *Access*

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

FIGURE 10. The detection results of the proposed method near coast.

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

IEEE *Access*

Table 2 shows the detection performance of ROI Align (R-DFPN) [1], RROI Align (RRPN) [49] and Adaptive ROI Align. Obviously, adaptive ROI Align achieved the best results, especially the recall. Because the noise in the redundant regions often interfere with or even submerges the features, the ROI Align detection results often have missing detection and false alarms, as shown in Fig. 7(a). Although RROI Align completely eliminates the interference of redundant regions, it loses spatial information. The detection performance of RROI Align still unsatisfactory, which is reflected in the inaccurate prediction results and large angle deviation, as illustrated in Fig. 7(b). Because of the sensitive relationship between IoU overlap and rotation angle, the recall of RROI Align has been improved slightly. Fig. 7(c) shows adaptive ROI Align has the advantage of reducing the influence of noise and preserving the spatial information. Therefore, the detection result of adaptive ROI Align is accurate and the recall is high.

E. COMPARISONS WITH OTHER SHIP DETECTION METHODS

In order to prove our proposed method is more competitive than traditional detection methods which are state-of-the-art in computer vision, we compare our proposed method with Faster-RCNN, FPN, FPN-Soft-NMS and LSTM-Based [52].

Table 3 show the quantitative comparison results of five methods, measured by F1. In the comparison of traditional detection methods (Faster-RCNN and FPN), FPN based on multiscale network has better performance. Meanwhile, FPN obtains the highest Precision value among (89.3%) the five methods. Soft-NMS [53] makes the bounding box, whose IoU exceeds the threshold have a certain probability to be reserved, so it is helpful to the dense scene detection. The results show that the FPN-Soft-NMS achieves an increase of about 1.2% without adding additional training and computational burden. LSTM-Based is a novel structure of the detection network, which introduces the LSTM structural unit. At the same time, it uses the Hungarian algorithm to serialize the output of the final detection result without any post-processing. This approach avoids the use of nonmaximum suppression operation and is suitable to use in dense scenes. However, this method still has certain limitations to highly overlapping objects, such as large aspect ratio ships. In the final detection results, detection performance of LSTM-Based is not prominent. Compared with the four detection framework above, the detection model proposed in this paper has achieved the best detection performance, and has the highest recall. Although our method offers superior performance in both multiscale and high-density object, we can see from Table 3 that the Precision of our method is not the highest, being behind that of the traditional method. This phenomenon shows that our method has a higher false alarm.

The time required for computing is summarized in Table 3. LSTM-Based has the fastest detection speed. Although our

method is the slowest of all detection algorithms, it is still very efficient.

Different Recall and Precision can be obtained by changing the confidence score threshold of detection. Fig. 8 plots the performance curves of different methods. As shown in the figure, the proposed method has the highest recall in a given precision. Similarly, with a specific recall, the proposed method has the highest precision. In short, the method we proposed has the best performance.

F. EVALUATION OF PROW DIRECTION PREDICTION

In this section, we will estimate the accuracy of the prow direction prediction of our rotational region detection method. Fig. 9 is the statistical results. The accuracy of prow predictions depends on the recall and we discover that the prediction accuracy in each direction is close to the recall, therefore it is a strong evidence that this simple prediction method is efficient and feasible.

An interesting phenomenon can be found in the experiment, that the final prediction often appears only in the prow and stern. We speculate that our network first learned that the prow must be in the long side and then judged which side of the long edge is the prow.

V. CONCLUSION

In this paper, we build an end to end ship detection framework based on rotation regions which can handle different complex scenarios, detect intensive objects, and reduce redundant detection regions. Many novel structures were designed for this model. For example, we design a new multiscale feature fusion network, called DFPN, which can effectively integrate the low-level location information and high-level semantic information to provide more advanced features for object detection. Meanwhile, we explore the detection performance of different feature maps combination strategies. Then, adaptive ROI Align is proposed in this paper to mitigate the effects of redundant noise regions in the proposals and keep the completeness of semantic and spatial information. In addition, the berthing and sailing direction of ship has been found through prediction. At last, we adopt R-NMS which has more stringent constraints to obtain more accurate prediction. Experiments based on SRSS dataset for rotation region detection show that our detection method has a competitive performance, as shown in Fig. 10.

REFERENCES

[1] X. Yang *et al.*, "Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens.*, vol. 10, no. 1, p. 132, 2018.

IEEE Access

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

[2] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[3] W. Liu, F. Yamazaki, and T. T. Vu, "Automated vehicle extraction and speed determination from QuickBird satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 75–82, Mar. 2011.

[4] D. E. Bar and S. Raboy, "Moving car detection and spectral restoration in a single satellite WorldView-2 imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2077–2087, Oct. 2013.

[5] L. Eikvil, L. Aurdal, and H. Koren, "Classification-based vehicle detection in high-resolution satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 1, pp. 65–72, 2009.

[6] A. Kembhavi, D. Harwood, and L. S. Davis, "Vehicle detection using partial least squares," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1250–1265, Jun. 2011.

[7] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1701–1717, Mar. 2013.

[8] S. Ghaffarian and S. Ghaffarian, "Automatic building detection based on Purposive FastICA (PFICA) algorithm using monocular high resolution Google Earth images," *ISPRS J. Photogramm. Remote Sens.*, vol. 97, pp. 152–159, Nov. 2014.

[9] M. I. Elbakary and K. M. Iftekharuddin, "Shadow detection of man-made buildings in high-resolution panchromatic satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5374–5386, Sep. 2014.

[10] X. Tong *et al.*, "Building-damage detection using pre- and post-seismic high-resolution satellite stereo imagery: A case study of the May 2008 Wenchuan earthquake," *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 13–27, Mar. 2012.

[11] X. Tong *et al.*, "Use of shadows for detection of earthquake-induced collapsed buildings in high-resolution satellite imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 53–67, May 2013.

[12] X. Huang, L. Zhang, and T. Zhu, "Building change detection from multi-temporal high-resolution remotely sensed images based on a morphological building index," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 105–115, Jan. 2014.

[13] S. Rastegar, A. Babaeian, M. Bandarabadi, and Y. Toopchi, "Airplane detection and tracking using wavelet features and SVM classifier," in *Proc. Southeastern Symp. Syst. Theory*, Tullahoma, TN, USA, Mar. 2009, pp. 64–67.

[14] H. Cai and Y. Su, "Airplane detection in remote sensing image with a circle-frequency filter," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Wuhan, China, Nov. 2005, pp. 529–534.

[15] Z. An and Z. Shi, "An airplane detection method for panchromatic image," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Beijing, China, Jul. 2013, pp. 189–192.

[16] W. Li, S. Xiang, H. Wang, and C. Pan, "Robust airplane detection in satellite images," in *Proc. 18th IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 2821–2824.

[17] S. Bo and Y. Jing, "Region-based airplane detection in remotely sensed imagery," in *Proc. 3rd Int. Congr. Image Signal Process.*, Yantai, China, Oct. 2010, pp. 1923–1926.

[18] W. Siyu, G. Xin, S. Hao, Z. Xinwei, and S. Xian, "An aircraft detection method based on convolutional neural networks in high-resolution SAR images," *J. Radars*, vol. 6, no. 2, pp. 195–203, 2017.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 2818–2826.

[20] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[21] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105.

[23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 448–456.

[24] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 5353–5360.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 580–587.

[26] R. Girshick, "Fast R-CNN," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1440–1448.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[28] D. J. Crisp, "A ship detection system for RADARSAT-2 dual-pol multi-look imagery implemented in the ADSS," in *Proc. IEEE Int. Conf. Radar*, Adelaide, SA, Australia, Sep. 2013, pp. 318–323.

[29] D. J. Crisp, "The state-of-the-art in ship detection in synthetic aperture radar imagery," *Organic Lett.*, vol. 35, pp. 2165–2168, May 2004.

[30] C. Wang, F. Bi, W. Zhang, and L. Chen, "An intensity-space domain CFAR method for ship detection in HR SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 529–533, Apr. 2017.

[31] M. F. Fingas and C. E. Brown, "Review of ship detection from airborne platforms," *Can. J. Remote Sens.*, vol. 27, no. 4, pp. 379–385, 2001.

[32] X. Leng, K. Ji, S. Zhou, X. Xing, and H. Zou, "An adaptive ship detection scheme for spaceborne SAR imagery," *Sensors*, vol. 16, no. 9, p. 1345, 2016.

[33] Y. D. Yu, X. B. Yang, S. J. Xiao, and J. L. Lin, "Automated ship detection from optical remote sensing images," *Key Eng. Mater.*, vol. 500, pp. 785–791, Jan. 2012.

[34] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010.

[35] F. Bi, B. Zhu, L. Gao, and M. Bian, "A visual search inspired computational model for ship detection in optical satellite images," *IEEE Geosci. Remote Sens Lett.*, vol. 9, no. 4, pp. 749–754, Jul. 2012.

[36] G. Yang, B. Li, S. Ji, F. Gao, and Q. Xu, "Ship detection from optical satellite images based on sea surface analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 641–645, Mar. 2014.

[37] Z. Shi, X. Yu, Z. Jiang, and B. Li, "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4511–4523, Aug. 2014.

[38] C. Corbane, L. Najman, E. Pecoul, L. Demagistri, and M. Petit, "A complete processing chain for ship detection using optical satellite imagery," *Int. J. Remote Sens.*, vol. 31, n. 22, pp. 5837–5854, 2010.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[40] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 779–788.

[41] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

[42] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. (Dec. 2016). "Feature pyramid networks for object detection." [Online]. Available: https://arxiv.org/abs/1612.03144

[43] J. Dai, Y. Li, K. He, and J. Sun. (Jun. 2016). "R-FCN: Object detection via region-based fully convolutional networks." [Online]. Available: https://arxiv.org/abs/1605.06409

[44] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proc. IEEE Int. Workshop Remote Sens. Intell. Process.*, May 2017, pp. 1–4.

[45] R. Zhang, J. Yao, K. Zhang, C. Feng, and J. Zhang, "S-CNN-based ship detection from high-resolution remote sensing images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 41-B7, pp. 423–430, Jul. 2016.

[46] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, p. 860, 2017.

[47] J. Tang, C. Deng, G.-B. Huang, and B. Zhao, "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1174–1185, Mar. 2014.

X. Yang *et al.*: Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region CNN

IEEE *Access*

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 770–778.

[49] J. Ma *et al.* (Mar. 2017). "Arbitrary-oriented scene text detection via rotation proposals." [Online]. Available: https://arxiv.org/abs/1703.01086

[50] M. Abadi *et al.* (Mar. 2016). "TensorFlow: Large-scale machine learning on heterogeneous distributed systems." [Online]. Available: https://arxiv.org/abs/1603.04467

[51] J. Han *et al.*, "Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding," *ISPRS J. Photogramm. Remote Sens.*, vol. 89, pp. 37–48, Mar. 2014.

[52] R. Stewart, M. Andriluka, and A. Y. Ng, "End-to-end people detection in crowded scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2325–2333.

[53] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis. (Aug. 2017). "Soft-NMS—Improving object detection with one line of code." [Online]. Available: https://arxiv.org/abs/1704.04503
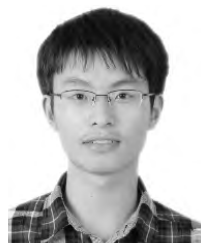
XUE YANG (S'18) received the B.Sc. degree from Central South University, Changsha, China, in 2016, where he is currently pursuing the M.S. degree with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China.

His research interests include computer vision, pattern recognition, and remote sensing image processing, especially on object detection and classification.

HAO SUN received the B.Sc. degree from the Beijing Institute of Technology, Beijing, China, in 2007, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences, Beijing, 2012.

He is currently an Associate Professor with the Institute of Electronics, Chinese Academy of Sciences, Beijing. His research interests include computer vision, pattern recognition, and remote sensing image processing.

XIAN SUN received the B.Sc. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2004, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2009.

He is currently an Associate Professor with the Institute of Electronics, Chinese Academy of Sciences, Beijing. His research interests include computer vision, and remote sensing image understanding.

MENGLONG YAN received the B.Sc. degree from Wuhan University, Wuhan, China, in 2007, and the M.Sc. and Ph.D. degrees from Beijing University, Beijing, China, in 2012.

He is currently an Associate Professor with the Institute of Electronics, Chinese Academy of Sciences, Beijing. His research interests include computer vision, pattern recognition, and remote sensing image processing.

ZHI GUO received the B.Sc. degree from Tsinghua University, Beijing, China, in 1998, and the M.Sc. and Ph.D. degrees from the Chinese Academy of Sciences University, Beijing, in 2003.

He is currently a Professor with the Institute of Electronics, Chinese Academy of Sciences, Beijing. His research interests include remote sensing information processing and application.

KUN FU received the B.Sc., M.Sc., and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 1995, 1999, and 2002, respectively.

He is currently a Professor with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision, remote sensing image understanding, geospatial data mining, and visualization.

• • •