IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 14, 2020, accepted March 26, 2020, date of publication April 3, 2020, date of current version April 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985637

Efficient Low-Cost Ship Detection for SAR Imagery Based on Simplified U-Net

YUXING MAO[1], (Student Member, IEEE), YUQIN YANG[2], ZIYUAN MA[3], (Member, IEEE), MINGZHE LI[4], HAO SU[5], AND JUN ZHANG[2]

[1]Department of Aerospace Science and Technology, Space Engineering University, Beijing 101416, China
[2]School of Architecture and Planning, Yunnan University, Kunming 650500, China
[3]Automation College, Nanjing University of Aeronautics and Astronautics, Nanjing 210000, China
[4]PLA Academy of Military Science, Beijing 100091, China
[5]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Jun Zhang (zhangjun650500@163.com)

ABSTRACT Due to the rapid development of chip technology and deep learning revolution, many ship detection frameworks for synthetic aperture radar (SAR) imagery based on convolutional neural networks (CNNs) have been proposed and achieved great success. However, there are problems hampering their development: 1) For the SAR ship detection task, it is uneconomic to apply heavy backbone network to extract features because it results in heavy computing load and prolongs the inference time cost; 2) The anchor-based methods usually have massive hyper-parameters, which typically need to be tuned carefully and easily lead to weak detection performance. To alleviate the problems, an efficient low-cost ship detection network for SAR imagery is proposed in this paper. Firstly, a simplified U-Net as the backbone to extract features is proposed. It only contains $\sim$ 0.47 million learnable weights, which is 2.37%, 0.76%, 0.34%, 1.01%, 0.55% and 1.07% of DarkNet-19, DarkNet-53, VGG-16, ResNet-50, ResNet-101 and ResNext-101, respectively. Secondly, an anchor-free SAR ship detection framework consisting of a bounding boxes regression sub-net and a score map regression sub-net based on simplified U-Net is proposed. To evaluate the effectiveness of our method, extensive experiments have been conducted and a more comprehensive set of evaluation metrics have been applied. Results demonstrate that the proposed network achieves 68.1% average precision and 67.6% average recall on the SAR ship detection dataset (SSDD), respectively. Compared with the state-of-the-art works, our proposed network achieves very competitive detection performance and extreme lightweight ($\sim$ 0.93 million learnable weights in total).

INDEX TERMS Bounding box, score map, simplified U-Net, anchor-free, low-cost, SAR ship detection.

I. INTRODUCTION

Marine traffic is increasingly crucial for global, regional and national economies and security because it significantly benefits seaborne trade and defends illegal activities, including smuggling, territorial sea invasion and maritime terrorism. To ensure effective and efficient marine traffic control, intelligent ocean surface ship surveillance, which is based on remote sensing imagery and computer-vision technique, has become a hot research field in recent years. High-resolution synthetic aperture radar (SAR) is regarded as one of the most suitable sensors for instances detection and maritime monitoring in the field of space technology, for it offers high-resolution images regardless of weather and light conditions.

During the past decades, numerous research achievements have been published in the field of SAR ship detection. The most widely used method is Constant False-Alarm Rate (CFAR) algorithm. This algorithm sets a threshold so that we can identify targets that are statistically significant above the background pixel while maintaining a constant false alarm rate [1]–[13]. In [5], Ai *et al.* proposed a 2-D joint log-normal distribution algorithm utilizing a strong gray intensity correlation to model the clutter of ship targets. Wang *et al.* presented a new hierarchical scheme for detecting ships in SAR images [6], which consisted of detection and discrimination modules, so that ship candidates were obtained by applying CFAR and ship discrimination was performed by using one-class classification. Hou *et al.* detected ships by measuring the visual conspicuity of each water region. And then, the ship targets were detected in the interested

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson.

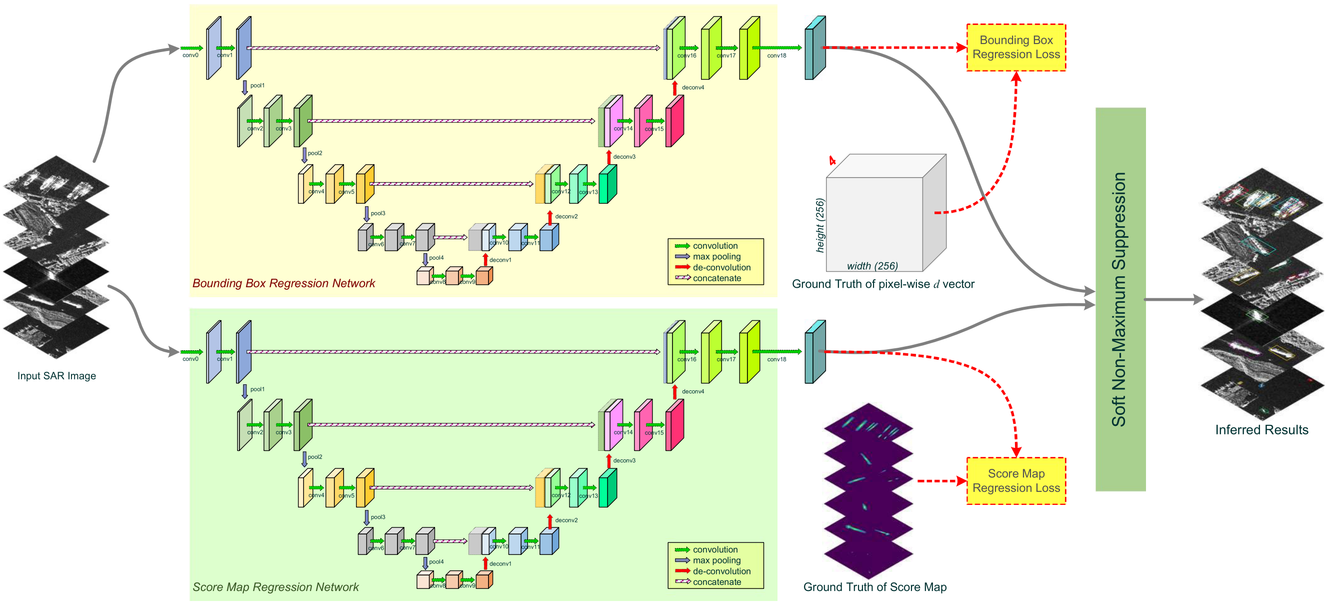


FIGURE 1. Architecture of proposed SAR ship detection network based on simplified U-Net backbone.

regions by the k-means clustering algorithm [7]. Wang *et al.* designed a fast block detector to extract sea clutter in a uniform local area [8], and then CFAR was employed. Ships were identified based on the kernel density estimation of ships, aspect ratio and pixel points. However, due to the high similarity between the harbor and the ship body on gray and texture features, the methods described above are unable to achieve the effective detection of inshore ships. Zhai *et al.* presented a novel approach via saliency and context information to deal with this issue [9]. Zhang *et al.* proposed a ship segmentation scheme with non-local processing to handle the speckle noise and the complicated backscattering phenomenology in SAR images [10]. In [12], both ship bounding boxes and contours are extracted based on the active contour algorithm. These ship detection approaches are based on manually extracted features, experienced statistic model and traditional image processing methods. A great deal of prior knowledge is required, and end-to-end optimization is difficult to achieve. In recent years, with the rapid development of hardware computing, deep learning has encountered another revival. Lots of convolutional neural network (CNN) architectures have already achieved great success for image features extraction and classification tasks in computer vision field, such as AlexNet [14], VGG-Net [15], GoogleNet [16], ResNet [17] ResNeXt [18], and EfficientNet [19]. Taking these CNNs with strong feature extraction capability as the backbone, region proposal-based object detectors, such as Faster R-CNN [20] and its variants [21]–[24], and regression-based detectors, such as YOLO [25] and its variants [26]–[29], have proved their remarkable results on various object detection benchmarks. With the release of a series of ship detection datasets for SAR image [30]–[32], many researchers began to apply universal object detectors or design application-specific network to SAR ship detection [32]–[52]. Li *et al.* firstly proposed to apply Faster R-CNN in the SAR ship detection task and they believed that deep learning based detector would be the focus of future mainstream research [32]. Kang *et al.* took the objects proposals generated by Faster R-CNN for the guard windows of the CFAR algorithm so that in this way these small-sized ship targets could be identified both accurately and efficiently [33]. Ai *et al.* used an adaptive-threshold-based CFAR detector as target prescreening to remove the high-intensity outliers before CNN [34]. An improved Faster R-CNN based on maximum stability extremal region decision criterion for SAR ship detection in harbor was proposed in [35], aiming to achieve effective inshore ship detection. To improve the ship detection accuracy, much efforts have been devoted from different perspectives to improving the ship detection accuracy, such as sea-land segmentation [36], contextual information fusion [37], attention mechanism [43]–[46], oriented proposal [38], [39] and transfer learning [47]–[49].

In view of the above works, the deep-learning-based SAR ship detectors usually perform better than traditional ones since CNN can extract features adaptively and it has a strong ability of non-linear mapping to regress the location and bounding boxes of ship instances. However, most works focus on improving ship detection accuracy, whereas the detection speed, which is extraordinarily significant, especially in maritime rescue and emergent military decision-making, is neglected to some extent. Although there are several recent studies which regard speed as an essential aspect, their backbone networks, which are originally constructed for universal multi-categories object detection task, are still too heavy. Unlike the universal multi-categories object detection task, the ship detection task only contains ship and background categories, and it is therefore uneconomic to apply heavy backbone network to extract features because it results in heavy computing load and prolongs the inference time cost. It is necesary to design specific lightweight backbone network for SAR ship detection. In addition, the approaches mentioned above are all anchor-based. The anchor-based object detectors usually have massive hyper-parameters, which typically need to be tuned carefully and easily lead to weak detection performance.

In order to alleviate the problems above, an efficient low-cost ship detection network for SAR imagery is proposed in this paper, starting with a simplified U-Net [53] as the backbone to extract features. Different from other feature pyramids network (FPN) [54] based object detectors which are computing and memory intensive, the proposed simplified U-Net has a concatenation mechanism behaving as pyramidal hierarchy concept but achieves extremely low-cost. Then, we construct a bounding boxes regression network with a score map regression network in parallel based on simplified U-Net backbone. The former is used to regress the bounding boxes size in polar representation and the latter is exploited to predict the probability of current pixel as centers of ship instances. After the two sub-nets, a soft non-maximum suppression (NMS) [55] post-processing module is cascaded to fuse all bounding boxes proposals. In theory, the proposed lightweight network can directly regress ships' bounding boxes without any predefined anchors.

To verify the practicability and robustness of the proposed method, we conducted extensive experiments and used a more comprehensive set of metrics than previous papers. Compared with other state-of-the-art methods, our proposed architecture showed competitive performance with much lower computing costs. The main contributions of this paper are as follows: 1) A simplified U-Net is proposed specifically for extracting the features of SAR ship instances. The proposed simplified U-Net (~0.47 million learnable weights) is significantly lighter than the backbone network used by the mainstream detectors. 2) An anchor-free SAR ship detection framework consists of a bounding boxes regression sub-net and a score map regression sub-net based on simplified U-Net is proposed. The proposed network architecture not only achieved encouraging performance and extreme lightweight (<1 million learnable weights in total), but also can be easily modified for multi-tasks.

The remainder of this paper is organized with the following sections: Section II illustrates the proposed ship detection architecture. Experiments and results are demonstrated and discussed in Section III, Section IV, respectively. A brief conclusion is given in Section V.

II. ARCHITECTURE OF PROPOSED NETWORK

A. ARCHITECTURE

Fig. 1 illustrates the architecture of the proposed ship instance detection network, which consists of two branches: 1) ship bounding boxes regression network and 2) score map regression network. Both branches use the proposed simplified U-Net to extract features. Subsequently, a soft non-maximum suppression (NMS) post-processing module is cascaded to fuse all bounding boxes proposals. The ship bounding box is represented by 4-tuple vector, and is expected to be regressed based on each pixel in input image. The score map regression network is designed to predict a 2-D probability distribution in which each score at each position $[i, j]$ indicates the likelihood of current position as the centers of any ship instances. Based on the predicted score map, one can choose these pixels with high scores and construct the bounding box using a $4-D$ vector using the current position as the origin point. Though related, ship bounding boxes regression and score map regression are two different types of regression tasks. If only one simplified U-Net is used to extract features, which means sharing the weights and calculation, many convolution layers need to be cascaded in parallel for bounding boxes regression and score map regression. This will not only increase the number of parameters, but also cause the multi-scale information of the simplified U-Net not to be significantly transmitted backward, so it is necessary to use two parallel sub-networks. Due to the fact that neighboring bounding boxes usually have correlated scores and result in false positives, a soft NMS module is therefore cascaded to fuse all these bounding boxes.

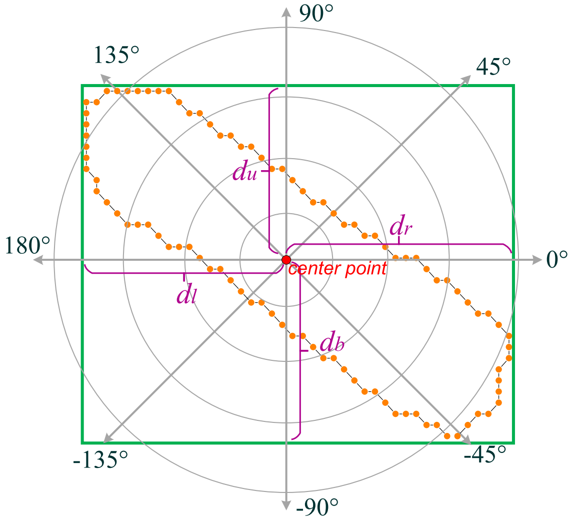


FIGURE 2. Ship instance's contour in the polar coordinate system, each point is represented with (θ, d), the bounding box could be represented with $(d_r, d_u, d_l, d_b)^T$ vector.

B. SHIP BOUNDING BOX REPRESENTATION

As shown in Fig. 2, each point locates at the ship instance boundary is represented with (θ, d) in polar coordinate system. This concept is quite simple and was initially proposed in [56] for instance segmentation. It inspired us to represent the bounding box in the same way. Once given a center point $[x_c, y_c]$, one can compute the intersection points $(x_i, y_i), i = 1, 2, 3, 4$ of ship bounding box and four rays

TABLE 1. Architecture of simplified U-Net backbone.

Operation	Input Blob Size	# of Kernels	Kernel/Pooling Size	Stride	Output Blob Size	# of learnable weights
conv0	$512 \times 512 \times 3$	16	3×3	1	$512 \times 512 \times 16$	448
conv1	$512 \times 512 \times 16$	16	3×3	1	$512 \times 512 \times 16$	2320
pool1	$512 \times 512 \times 16$	-	2×2	2	$256 \times 256 \times 16$	-
conv2	$256 \times 256 \times 16$	32	3×3	1	$256 \times 256 \times 32$	4640
conv3	$256 \times 256 \times 32$	32	3×3	1	$256 \times 256 \times 32$	9248
pool2	$256 \times 256 \times 32$	-	2×2	2	$128 \times 128 \times 32$	-
conv4	$128 \times 128 \times 32$	48	3×3	1	$128 \times 128 \times 48$	13872
conv5	$128 \times 128 \times 48$	48	3×3	1	$128 \times 128 \times 48$	20784
pool3	$128 \times 128 \times 48$	-	2×2	2	$64 \times 64 \times 48$	-
conv6	$64 \times 64 \times 48$	64	3×3	1	$64 \times 64 \times 64$	27712
conv7	$64 \times 64 \times 64$	64	3×3	1	$64 \times 64 \times 64$	36928
pool4	$64 \times 64 \times 64$	-	2×2	2	$32 \times 32 \times 64$	-
conv8	$32 \times 32 \times 64$	64	3×3	1	$32 \times 32 \times 64$	36928
conv9	$32 \times 32 \times 64$	64	3×3	1	$32 \times 32 \times 64$	36928
deconv1	$32 \times 32 \times 64$	32	3×3	3	$64 \times 64 \times 32$	18464
conv10	$64 \times 64 \times 96$	64	3×3	1	$64 \times 64 \times 64$	55360
conv11	$64 \times 64 \times 64$	64	3×3	1	$64 \times 64 \times 64$	36928
deconv2	$64 \times 64 \times 64$	32	3×3	2	$128 \times 128 \times 32$	18464
conv12	$128 \times 128 \times 80$	48	3×3	1	$128 \times 128 \times 48$	34608
conv13	$128 \times 128 \times 48$	48	3×3	1	$128 \times 128 \times 48$	20784
deconv3	$128 \times 128 \times 48$	32	3×3	2	$256 \times 256 \times 32$	13856
conv14	$256 \times 256 \times 64$	32	3×3	1	$256 \times 256 \times 32$	18464
conv15	$256 \times 256 \times 32$	32	3×3	1	$256 \times 256 \times 32$	9248
deconv4	$256 \times 256 \times 32$	32	3×3	2	$512 \times 512 \times 32$	9248
conv16	$512 \times 512 \times 48$	48	3×3	1	$512 \times 512 \times 48$	20784
conv17	$512 \times 512 \times 48$	48	3×3	1	$512 \times 512 \times 48$	20784
conv18[b]	$512 \times 512 \times 48$	4	1×1	1	$512 \times 512 \times 4$	196
conv18[c]	$512 \times 512 \times 48$	2	1×1	1	$512 \times 512 \times 2$	98

[a] Data blob size is $width \times height \times channels$.

[b] For bounding boxes regression network, number of kernels and output blob size are 4 and $512 \times 512 \times 4$, respectively.

[c] For score map regression network, number of kernels and output blob size are 2 and $512 \times 512 \times 2$, respectively.

starting from center point with $\theta = 0°, 90°, 180°, -90°$ easily. The corresponding distance d_r, d_u, d_l, d_b between the center point and each contour intersection point compose a $4 - D$ vector. Based on the center point $[x_c, y_c]$ and the distance vector $(d_r, d_u, d_l, d_b)^T$, the ship bounding box could be reconstructed handily. In this way, the ship instance detection task is formulated as instance center classification and four-direction distance regression.

C. SHIP BOUNDING BOX REGRESSION NETWORK

In this work, a simplified U-Net [53] based end-to-end training convolutional neural network is constructed to regress the distance vector $(d_r, d_u, d_l, d_b)^T$ at each potential pixel, as is illustrated in the upper branch in Fig. 1. The proposed simplified U-Net follows the encoder-decoder framework and it consists of a contracting path (left partition) and an expansive path (right partition). The contracting path follows the typical architecture of a convnet. It consists of the repeated application of two 3×3 convolutions, each followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation with stride 2 for down-sampling. After each down-sampling operation, we increase the number of convolutional kernels. Every step in the expansive path consists of an up-sampling of the feature map (3×3 up-convolution) followed by a concatenation operation from the same level from contracting path, and then the two 3×3 convolutions, each followed by a ReLU, are stacked. Since the high resolution features from the contracting path are combined with the up-sampled output, the successive convolution layer can then learn to assemble a more precise output based on it. Unlike other feature pyramids network (FPN) based object detectors, which are computing and memory intensive, simplified U-Net's concatenation mechanism combines coarse, high layer information with fine, low layer information together, which behaves as pyramidal hierarchy concept but achieves low-cost. Table 1 lists the ship detection network detail configuration based on 512×512 input SAR image size and reports the amount of learnable weights. Final convolutional layer "conv18" after U-Net backbone maps feature map to 4 channels, regressing d_r, d_u, d_l, d_b, respectively. Our proposed simplified U-Net only contains $\sim$ 0.47 million learnable weights, which is 2.37%, 0.76%, 0.34%, 1.01%, 0.55%, 1.07% of Darknet-19, Darknet-53, VGG-16, ResNet-50, ResNet-101, ResNext-101, respectively.

D. SHIP BOUNDING BOX REGRESSION LOSS

In object detection task, the metric intersection over union (IoU) measures how well the predicted bounding boxes and their ground truths align and overlap. IoU loss is an effective way to supervise the learning phase, which directly optimizes the metric of interest. In this paper, IoU loss is used to train the

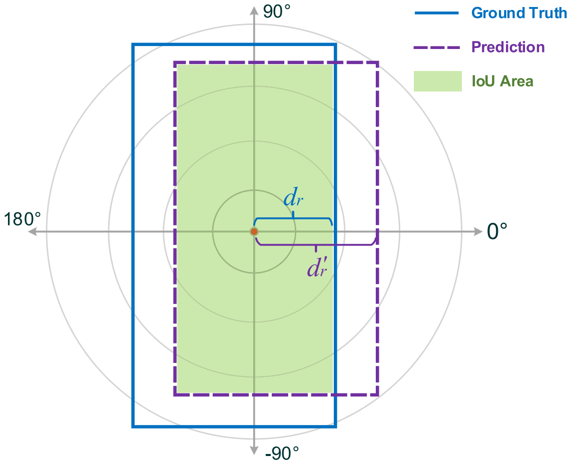


FIGURE 3. IoU(interaction area over union area) of ship instance.

bounding boxes regression network, which is defined below:

$$L_{bbox} = 1.0 - \frac{2 * |\mathcal{B}_p \cap \mathcal{B}_{gt}|}{|\mathcal{B}_p| + |\mathcal{B}_{gt}|} \tag{1}$$

where $\mathcal{B}_p$ and $\mathcal{B}_{gt}$ are predicted bounding box and its corresponding ground truth, respectively. $|\cdot|$ is a function return to the area of input's rectangle. As is depicted in Fig. 3, the IoU area could be figured out by:

$$|\mathcal{B}_p \cap \mathcal{B}_{gt}| = (min(d_r, d_r') - min(d_l, d_l')) \times (min(d_u, d_u') - min(d_b, d_b')) \tag{2}$$

where $(d_r', d_u', d_l', d_b')^T$ and $(d_r, d_u, d_l, d_b)^T$ are predicted distance vector at $\theta = 0°, 90°, 180°, -90°$ in polar coordinate system and its ground truth.

E. SCORE MAP REGRESSION NETWORK

Score map regression network in Fig. 1 is used for predicting score map $\mathcal{S}'$, in which each element locates at $[i, j]$ indicates its probability as centerness of any ship instances. The network also follows simplified U-Net structure (as shown in Table. 1) and the final convolutional layer outputs a 512 × 512 × 2 data blob with softmax probability normalization. During the inference phase, one can choose these pixels with the high predicted score as a reference point (center point in Fig. 2) to reconstruct the bounding box of ship instances. To get the ground truth of score map $\mathcal{S}$ for network training, $d[i, j]$ is defined as the distance between pixel $[i, j]$ to its closest pixel which belongs to background category. It is clear that larger $d[i, j]$ implies pixel $[i, j]$ is far away from the ship instance border, i.e., closer to ship instances' centerness. As illustrated in Fig. 4)(a), there are totally 5 ship instances in the given image, $d[i, j], \forall(i, j)$ is figured out and normalized to $0.0 \sim 1.0$ for each ship instance based on our ship semantic annotations. Finally, these score maps of 5 ship instances are accumulated together as $\mathcal{S}$, which is shown in Fig. 4(b).

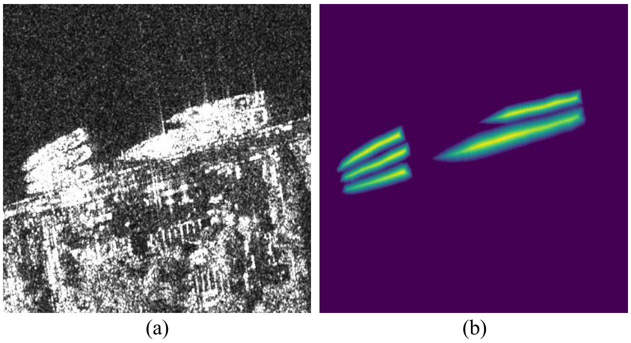


FIGURE 4. Input SAR image (a) and its ground truth of score map (b).

F. SCORE MAP REGRESSION LOSS

Due to the fact that the ship instances usually occupy a fraction of the input SAR image, it is a typical positive/negative samples imbalance situation for score map regression network training and all the neural network weights could be easily converged to zeros during training phase if we directly use $L2$ Euclidean distance loss. In order to alleviate the samples imbalance problem, we derive an effective loss formula based on two expectations: i). the non-zero (e.g. $> 10^{-4}$) area of predicted score map $\mathcal{S}'$ should be converged to non-zero area of its ground truth $\mathcal{S}$. ii). the difference between predicted $\mathcal{S}'$ and its ground truth $\mathcal{S}$ at non-zero area should converge to zero. Therefore, the score map regression loss L_{score} is formulated below:

$$L_{score} = L^{dist} + L^{closeness} \tag{3}$$

$$L^{dist} = \frac{2 \times g(\mathcal{S}') \cap g(\mathcal{S})}{g(\mathcal{S}') \cup g(\mathcal{S})} \tag{4}$$

$$L^{closeness} = \frac{1}{|g(\mathcal{S})|} \sum_{\forall(x,y) \in |g(\mathcal{S})|} (\mathcal{S}[x, y] - \mathcal{S}'[x, y])^2 \tag{5}$$

where L_{dist} implies the IoU between non-zeros areas of $\mathcal{S}'$ and $\mathcal{S}$. $L^{closeness}$ measures the proximity of score values at non-zero area. $g(\mathcal{Z})$ is a function to return all non-zero coordinates list of input matrix $\mathcal{Z}$. $|g(\mathcal{S})|$ is the non-zero values' amount of $\mathcal{S}$.

G. SHIP INSTANCES FUSION

Due to the fact that each pixel in resized input SAR image outputs the bounding box proposal vector $(d_r, d_u, d_l, d_b)^T$ with its corresponding score p which implies the probability of the current pixel locates as centerness of a ship instance. Neighboring bounding boxes usually have correlated scores and it results in false positives. Therefore, a soft non-maximum suppression (soft NMS) is exploited as a post-processing phase to get final detections, as is illustrated in Algorithm 1. For a detection bounding box $\mathcal{F}$ with the maximum score, soft NMS decays the detection scores of all other objects as a continuous function of their overlap with $\mathcal{F}$. It is clear that scores for detection boxes which have a higher overlap with $\mathcal{F}$ should be decayed more, as they have a higher likelihood of being false positives. The decay function used in our work

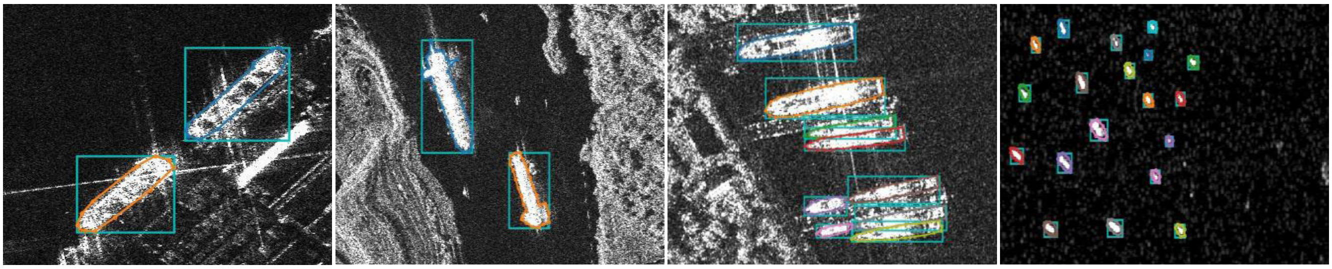
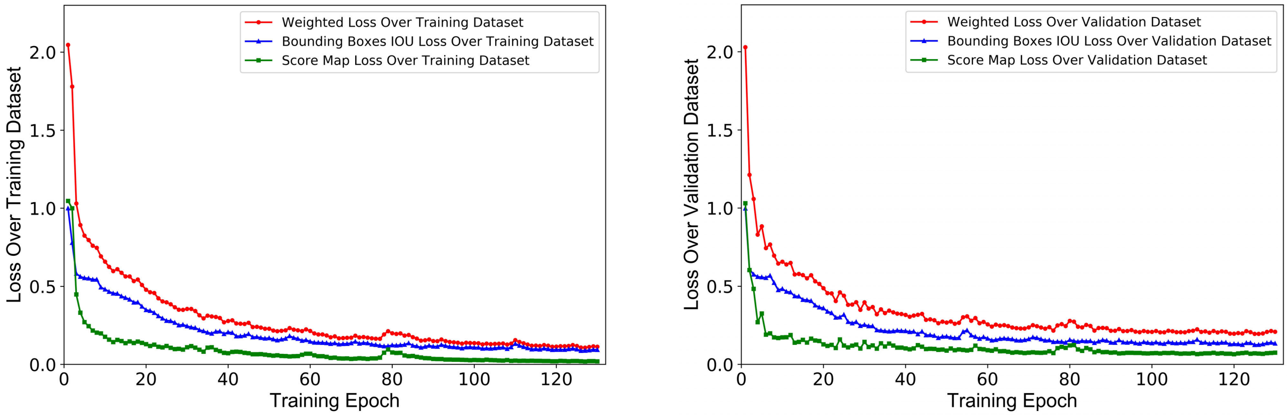


FIGURE 5. SAR image samples with bounding boxes annotation from SSDD, semantic polygon boundary is constructed for score map regression.



(a) Training Loss Curve vs. Training Epochs

(b) Validation Loss Curve vs. Training Epochs

FIGURE 6. Loss convergence curve of training dataset and validation dataset.

is the same with [55], it is a linear weight function defined below:

$$f(iou(\mathcal{F}, D_i)) = \begin{cases} 1, & \text{if } iou(\mathcal{F}, D_i) < P_t \\ 1.0 - iou(\mathcal{F}, D_i), & \text{otherwise} \end{cases} \tag{6}$$

Algorithm 1: Soft Non-Maximum Suppression

```
Input: Initial detected bounding boxes list $\mathcal{D} = \{D_1, D_2, D_3, \cdots, D_M\}$ and its corresponding confidence probability (score value) list $\mathcal{C} = \{c_1, c_2, c_3, \cdots, c_M\}$, NMS threshold $P_t$
Output: Bounding boxes list $\hat{\mathcal{D}}$ with corresponding confidence probability list $\hat{\mathcal{C}}$.
$\hat{\mathcal{D}} \leftarrow \{\}, \hat{\mathcal{C}} \leftarrow \{\}$
while $len(\mathcal{D}) \neq 0$ do
    $e \leftarrow \text{argmax } \mathcal{C}$
    $\mathcal{F} \leftarrow D_e,\ p \leftarrow c_e$
    $\hat{\mathcal{D}} \leftarrow \hat{\mathcal{D}} \cup \mathcal{F},\ \hat{\mathcal{C}} \leftarrow \hat{\mathcal{C}} \cup p$
    $\mathcal{D} \leftarrow \mathcal{D} - \mathcal{F},\ \mathcal{C} \leftarrow \mathcal{C} - p$
    for $D_i$ in $\mathcal{D}$ do
        $c_i \leftarrow c_i \times f(iou(\mathcal{F}, D_i))$
return $\hat{\mathcal{D}}, \hat{\mathcal{C}}$
```

III. EXPERIMENTS SETUP

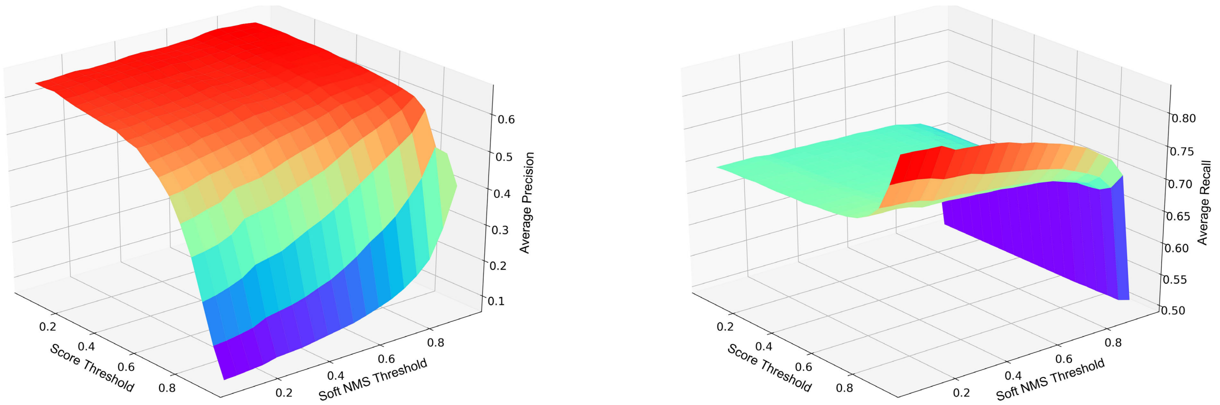
A. DATASET

In this paper, we used the SAR ship detection dataset (SSDD) [32] to train and validate our proposed network's performance. SSDD is a widely used benchmark for researchers to evaluate their approaches because it includes ships in various environments, such as a variety of ships with adjacent docks and land, isolated oceans, and side by side, as shown in Fig. 5. SSDD follows the standardized image object's ground truth labeling approach in PASCAL VOC [57] to construct annotations. There are totally 1160 SAR image slices (with indexing from 1 to 1160) and 2456 ship instances in SSDD from RadarSat-2, TerraSAR-X, and Sentinel-1 in Yantai, China and Visakhapatnam, India. They vary in terms of polarization (including HH, VV, VH and HV) and resolution (from 1m to 15m). The average number of ships per image is 2.1. We divided SSDD dataset into three parts: training subset (754 images), validation subset (174 images) and test subset (232 images), these images with indexes' suffix 1 and 9 were collected as test subset, others as training subset and validation subset randomly. It is worth mentioning that we constructed the semantic label artificially for score

TABLE 2. Comparison of different U-Net configurations based on SSDD (%).

Model	AP	$AP^{.50}$	$AP^{.75}$	AP^{S}	AP^{M}	AP^{L}	AR	AR^{S}	AR^{M}	AR^{L}	# of learnable weights
Standard U-Net [53]	68.2	94.1	**83.1**	**63.6**	75.9	70.4	67.4	**63.1**	74.7	**72.0**	30782016
U-Net[a]+ResNet-Block	**68.6**	**95.0**	82.0	62.9	**78.2**	71.7	66.3	62.9	**77.0**	69.9	13429248
U-Net[a]+DCK	59.0	81.5	72.0	54.9	65.7	63.5	58.7	60.5	71.9	68.3	9899968
Proposed U-Net[a]	68.1	94.0	83.0	63.3	75.7	**73.2**	**67.6**	63.0	74.9	71.2	**933894**

[a] U-Net is based on our modified version which is illustrated in Fig. 1.



(a) Average Precision Distribution.

(b) Average Recall Distribution.

FIGURE 7. Average precision and recall at different S_t and P_t.

map supervised learning, which is illustrated with polygon boundary in Fig. 5.

B. EVALUATION METRICS

To evaluate the performance of our proposed low-cost ship detection network, the following metrics were used: average precision (AP), average recall (AR), network parameters amount N and inference time cost $T_{inf.}$. For a given IoU threshold η, the predicted bounding box $\mathcal{B}_p$ is regarded as a true positive (TP) only if the IoU between $\mathcal{B}_p$ and its ground truth $\mathcal{B}_{gt}$ is greater than η; otherwise, it is a false negative (FN). The precision P and recall R at IoU threshold η are defined below, respectively:

$$P(\eta) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{7}$$

$$R(\eta) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

where FP implies false positive and TP + FP represents the total number of ship bounding boxes recognized by the detection network. TP + FN is the total amount of ship bounding boxes in annotations (ground truth). Precision refers to the percentage of detected ship instances that are relevant and recall refers to the percentage of total relevant ship instances correctly gathered by the ship detector. It is not possible to maximize both precision and recall at the same time, as one comes at the cost of another. AP and AR are figured out over multiple η:

$$AP = \frac{1}{|\mathcal{T}|} \sum_{\eta \in \mathcal{T}} P(\eta) \tag{9}$$

$$AR = \frac{1}{|\mathcal{T}|} \sum_{\eta \in \mathcal{T}} R(\eta) \tag{10}$$

where $\mathcal{T} = [0.5, 0.55, 0.60, \cdots, 0.95]$ is the IoU thresholds set and $|\mathcal{T}|$ is the length of $\mathcal{T}$. In our experiments, the (average) precision at $\eta = 0.5$ ($AP^{.5}$) and $\eta = 0.75$ ($AP^{.75}$) were reported as well. Meanwhile, AP^S, AR^S, AP^M, AR^M, AP^L, AR^L were adopted for analyzing the average precision and average recall for small size (area of $\mathcal{B}_{gt} < 32^2$), medium size ($32^2 <$ area of $\mathcal{B}_{gt} < 96^2$) and large size (area of $\mathcal{B}_{gt} > 96^2$) ship instances, respectively. To emphasis the model complexity and its real-time capability, network parameters amount N and inference speed $T_{inf.}$ were taken into consideration for further comparison with state-of-the-art works.

IV. RESULTS AND DISCUSSION

A. TRAINING

The architecture in Fig. 1 was implemented based on Tensorflow [60] framework and was trained end-to-end using Intel Xeon(Cascade Lake) Platinum 8269CY CPU @ 2.5GHz and a NVIDIA GeForce TITAN V with 12G memory. The training batch was 32, the initial learning rate was 0.001 and decayed 2 times every 20 epochs. The gradient optimizer was Adam. Fig. 6 plots the training loss curve distribution, and the validation loss after each epoch is also reported. The weighted loss was the sum of bounding boxes IoU loss and score map regression loss. One can observe that the loss converged after 100 epochs, then the overfilling phenomena occured on the training dataset. We chose the network weights after epoch 119, in which the bounding boxes IoU

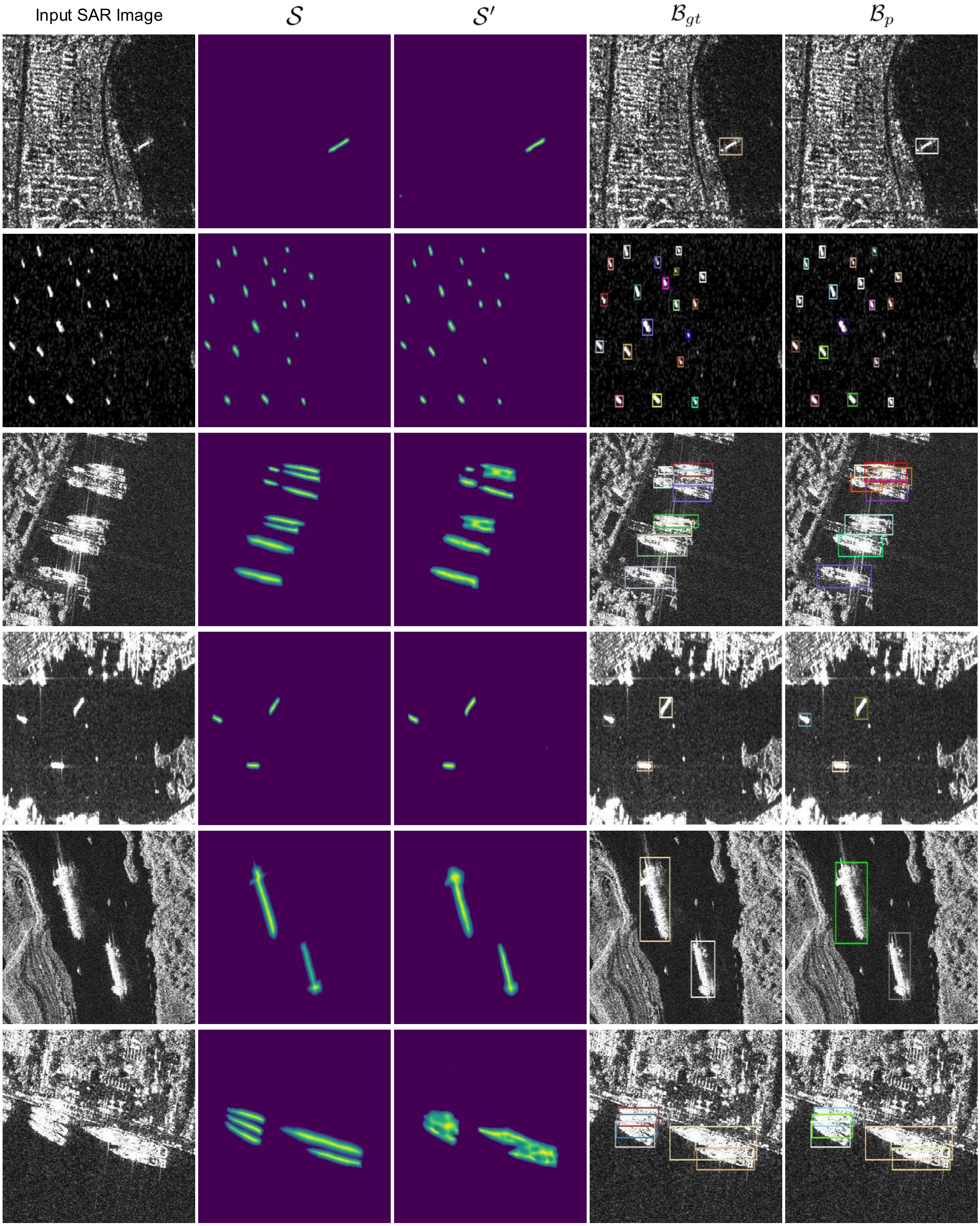


FIGURE 8. Visualization of score map prediction and bounding boxes prediction for SAR images. The bounding boxes are presented in different colors randomly.

TABLE 3. Comparison of ship detection performance with other state-of-the-art instance detection models based on SSDD (%).

Model	Backbone	AP	$AP^{.50}$	$AP^{.75}$	AP^S	AP^M	AP^L	AR	AR^S	AR^M	AR^L	$Params(M)$
YOLO-v2 [26]	DarkNet-19	50.8	88.4	46.2	54.4	53.7	57.4	54.9	50.3	62.9	55.9	25.8
YOLO-v3 [27]	DarkNet-53	58.8	94.9	63.8	57.1	64.6	43.6	61.7	60.3	65.4	49	61.5
SSD@300 × 300 [58]	VGG-16	61.2	92.8	71.5	57.2	69	63.5	65	61.1	71.7	66.9	23.8
SSD@500 × 500 [58]	VGG-16	63.9	94.4	74.9	59.8	70.6	72.6	66.8	63.3	72.8	70.6	24.4
ATSS [59]	ResNet-50+FPN	65.5	94.8	78.1	61.7	75.5	51.2	67.7	63.4	75.8	62.7	32.1
FCOS [29]	ResNet-101+FPN	47.2	82.6	46	49.0	47.0	30.2	54.9	54.9	55.9	47.4	51.1
RetinaNet [28]	ResNet-50+FPN	63.6	92.9	75.3	59.5	72.2	56	65.8	61.4	74.2	60.6	36.3
	ResNet-101+FPN	64.8	93.8	77.5	60.6	73.1	59.6	66.8	62.6	74.4	64.8	55.3
	ResNeXt-101+64×4d+FPN	65.7	94.1	80	61.5	73.9	64	67.5	63.3	74.7	68	94.2
Cascade R-CNN [21]	ResNet-50+FPN	69	95.3	83.2	64.1	76.9	73.2	68.8	64.1	76.4	**74.9**	69.2
	ResNet-101+FPN	69.8	96.4	84.5	64.1	**78.8**	73.2	69.5	64.4	**78**	71.7	88.1
	ResNeXt-101+64×4d+FPN	**70.3**	95.5	85.1	**65.4**	78.1	75.3	**69.7**	**65.3**	77.2	72.7	127.1
Faster R-CNN [20]	ResNet-50+FPN	67.4	95.9	80.9	63.8	75	60.9	67.8	64	75	63.3	41.4
	ResNet-101+FPN	67.3	95.4	81.4	63.1	75.2	65.1	68.3	64.1	76.3	64.3	60.3
	ResNeXt-101+64×4d+FPN	69	97.1	83.4	65.1	76.2	67.1	69.4	65.3	76.9	68	99.3
Mask R-CNN[b] [22]	ResNet-50+FPN	67.9	94.7	84.1	63.6	75.1	67.2	68.1	63.9	75.1	72.1	44
	ResNet-101+FPN	68.3	95.9	83.7	63.5	76.2	72.6	68.5	63.9	76.3	72.2	63
Cascade Mask R-CNN[b] [23]	ResNet-50+FPN	69.8	94.6	**86.7**	64.8	78.6	**77.7**	69.7	65	77.5	73.8	77
	ResNet-101+FPN	69.4	94.2	83.9	64.1	78	77.3	69.3	64.6	76.7	74.8	96
Mask Scoring R-CNN[b] [24]	ResNet-50+FPN	68.9	94.9	85	64.7	77.1	66.2	69.2	64.9	77	65.4	60.2
	ResNet-101+FPN	68	**97.2**	82.2	63.4	76.3	72.6	68.4	63.7	76.3	72.2	79.2
Proposed	U-Net[a]	68.1	94	83	63.3	75.7	73.2	67.6	63	74.9	71.2	**0.93**

[a] U-Net is based on our modified version which is illustrated in Fig. 1.
[b] with segmentation function enabled.

loss converged to 0.1285 and the score map loss converged to 0.0687, to inference the validation dataset and evaluate the model performance.

B. ABLATION STUDY

1) INFLUENCE OF DIFFERENT U-NET CONFIGURATION

In order to evaluate the influence of different U-Net configurations, we performed ablation experiments from 3 perspective: 1) Increasing convolutional kernels's amount. 2) Increasing the feature extraction network based on residual block of ResNet [17], which retains shallow context by a connection from the first to the last layers of a block layer. The residual mechanism has already been proven that can effectively take equal or higher features expression ability. 3) Expanding the receptive field by replacing the convolution kernel in Fig. 1 with dilated convolutional kernel (DCK), which inserts zeros between pixels in convolutional kernels and was expected to increase the resolution of intermediate feature map responses [50]. Table 2 lists the performance results based on SSDD dataset. All these 4 experiments were trained 200 epochs with the same training configurations and selected optimal model parameters using validation weighted loss. We selected these pixels with predicted score > 0.2 as potential centerness of bounding boxes. The soft NMS threshold $P_t = 0.75$. We can observe that the dilated convolutional kernel is ineffective for detecting ships in the complicated backscattering phenomenology of SAR images. With residual block, due to the fact that different layers can contribute through the residual bypass principle, there is a minor improvement occurred in the detection performance, however, it is at the cost of high computing resource requirement and increasing model complexity. Therefore, our proposed simplified U-Net, with reduced convolutional kernels, is an economic solution with acceptable detection performance.

2) INFLUENCE OF DIFFERENT SCORE/SOFT NMS THRESHOLDS

After training the network, we need to set a score threshold S_t and highlight these pixels with predicted score $> S_t$, then reconstruct the bounding boxes based on these selected pixels as origin point together with predicted $(d_r, d_u, d_l, d_b)^T$ vector. Neighbour pixels usually generate bounding boxes with high correlation, so soft NMS with overlap threshold P_t was used to fuse all potential bounding boxes of the same ship instance, as is introduced in Algorithm 1. It is obvious that decreasing S_t will introduce more bounding boxes in input SAR image and changing P_t will influence the final bounding boxes quantity and quality directly. To analyze the detailed impact of both S_t and P_t, Fig. 7 plots the AP and AR distribution under various S_t and P_t. Findings show that optimal S_t and P_t for average precision is 0.20, 0.75, respectively. The corresponding AR is 0.676. One can obverse that AP and AR can not be optimized at the same time, one is at the cost of another. Meanwhile, AP and AR metrics are insensitive with respect to $S_t < 0.50$ and P_t, which indicates the robustness degree of bounding boxes regression for each pixel. By setting $S_t = 0.20$, $P_t = 0.75$, Fig. 8 illustrates some typical SAR examples in SSDD with their ground truths, predicted score map and predicted bounding boxes. For small size ships, if the score map could not predict it with high response, then it is easy to occur false negative, which can be observed from the second row in Fig. 8.

C. PERFORMANCE COMPARISON

To verify the effectiveness of our proposed method, YOLO-v2 [26], YOLO-v3 [27], SSD [58], RetinaNet [28], ATSS [59], FCOS [29], Cascade R-CNN [21], Faster R-CNN [20], Mask R-CNN [22], Cascade Mask R-CNN [23], Mask Scoring R-CNN [24] were implemented to evaluate

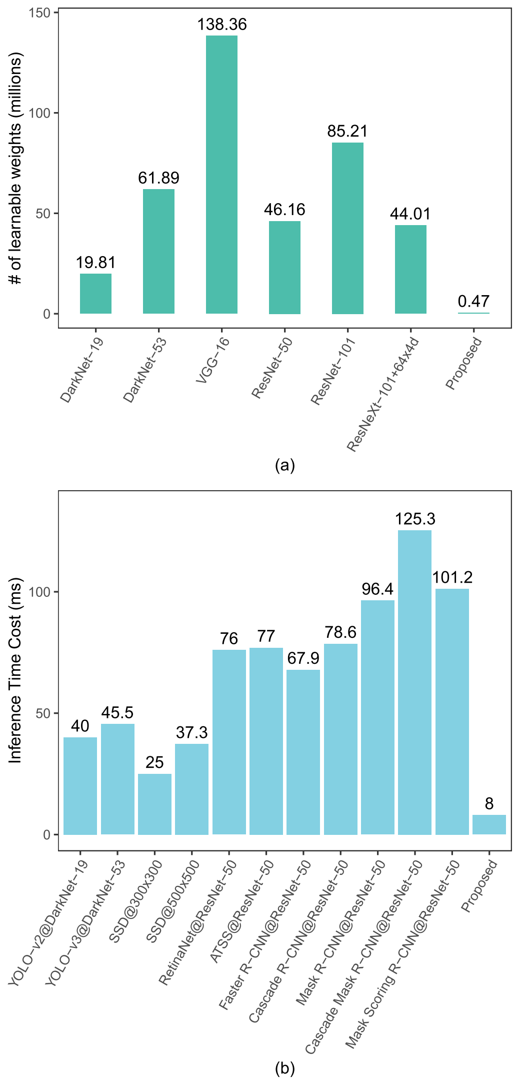


FIGURE 9. Backbone network learnable weights statistic (a) and inference time cost of different detection frameworks based on NVIDIA GeForce TITAN V GPU (b).

the detection performance based on Pytorch [61] framework. In addition, MMdetection toolbox [62] was used to implement RetinaNet, ATSS, FCOS, Cascade R-CNN, Faster R-CNN, Mask R-CNN, Cascade Mask R-CNN, Mask Scoring R-CNN. Table 3 reports the performance metrics of these state-of-the-art works. Different backbones with FPN including ResNet-50 [17], ResNet-101 [17], ResNeXt-101 [18] with 64 × 4d template were integrated in these competitors. Experiment results demonstrated that our proposed network achieved competitive performance with mainstream object detection frameworks. It is worth mentioning that our proposed network is much light-weighted as compared others. Fig. 9 illustrates the parameters amount of different backbones and the inference time cost of different frameworks based on NVIDIA GeForce TITAN V. Compared with other competitors, our proposed network, with ∼0.93 million learnable weights in total, only cost 8 milliseconds to infer a 512 × 512 SAR image. Therefore, even through it performed a little bit inferior than other state-of-the-art frameworks in detection performance, the proposed simplified U-Net based bounding boxes regression network with score map regression branch is absolutely a much more suitable solution for SAR ship detection in high-speed and low-cost scenarios.

V. CONCLUSION

In this paper, a novel simplified U-Net based bounding boxes regression network, together with another score map regression network in parallel is proposed to detect ships in SAR images. Score map indicates the probability of current position as the center of ships, and ship bounding boxes are represented in a 4 tuple format in polar coordinate system, which can be easily fitted and converged during the training phase. By filtering all pixels with predicted score above a given threshold, one can select all pixels with high confidence as "good" locations of ships. The proposed detection mechanism no longer depends on any region proposals and there is no necessity to configure any anchors. Therefore, our proposed network is of low-cost and carries slight quantity of learnable weights. It achieved encouraging performance in both detection and real-time capacity. The proposed methodology is not only limited to ship detection in SAR image, but can also be easily modified for multi-tasks.

REFERENCES

[1] M. Sciotti and P. Lombardo, "Ship detection in SAR images: A segmentation-based approach," in *Proc. IEEE Radar Conf.*, May 2001, pp. 81–86.

[2] K. Ouchi, S. Tamaki, H. Yaguchi, and M. Iehara, "Ship detection based on coherence images derived from cross correlation of multilook SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 3, pp. 184–187, Jul. 2004.

[3] C. Wang, M. Liao, and X. Li, "Ship detection in SAR image based on the alpha-stable distribution," *Sensors*, vol. 8, no. 8, pp. 4948–4960, 2008.

[4] F. Zhang and B. Wu, "A scheme for ship detection in inhomogeneous regions based on segmentation of SAR images," *Int. J. Remote Sens.*, vol. 29, no. 19, pp. 5733–5747, Oct. 2008.

[5] J. Ai, X. Qi, W. Yu, Y. Deng, F. Liu, and L. Shi, "A new CFAR ship detection algorithm based on 2-D joint log-normal distribution in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 806–810, Oct. 2010.

[6] Y. Wang and H. Liu, "A hierarchical ship detection scheme for high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 4173–4184, Oct. 2012.

[7] B. Hou, W. Yang, S. Wang, and X. Hou, "SAR image ship detection based on visual attention model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. - IGARSS*, Jul. 2013, pp. 2003–2006.

[8] C. Wang, S. Jiang, H. Zhang, F. Wu, and B. Zhang, "Ship detection for high-resolution SAR images based on feature analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 119–123, Jan. 2014.

[9] L. Zhai, Y. Li, and Y. Su, "Inshore ship detection via saliency and context information in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1870–1874, Dec. 2016.

[10] X. Zhang, B. Xiong, G. Dong, and G. Kuang, "Ship segmentation in SAR images by improved nonlocal active contour model," *Sensors*, vol. 18, no. 12, p. 4220, 2018.

[11] L. Zhu, G. Xiong, D. Guo, and W. Yu, "Ship target detection and segmentation method based on multi-fractal analysis," *J. Eng.*, vol. 2019, no. 21, pp. 7876–7879, Nov. 2019.

[12] Y.-Y. Nie, S.-C. Fan, and P.-L. Shui, "Fast ship contour extraction in SAR images," *J. Eng.*, vol. 2019, no. 19, pp. 5885–5888, Oct. 2019.

[13] Y. Liang, K. Sun, Y. Zeng, G. Li, and M. Xing, "An adaptive hierarchical detection method for ship targets in high-resolution SAR images," *Remote Sens.*, vol. 12, no. 2, p. 303, 2020.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image net classification with deep convolutional neural networks," in *Adv. neural Inf. Process. Syst.*, vol. 2012, pp. 1097–1105.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations, ICLR*, San Diego, CA, USA, May 2015.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[18] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.

[19] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*. [Online]. Available: http://arxiv.org/abs/1905.11946

[20] 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.

[21] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.

[22] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[23] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High quality object detection and instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Nov. 28, 2019, doi: 10.1109/TPAMI.2019.2956516.

[24] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring R-CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 6402–6411.

[25] 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)*, Jun. 2016, pp. 779–788.

[26] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[27] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[29] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9627–9636.

[30] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, 2019.

[31] S. Xian, W. Zhirui, S. Yuanrui, D. Wenhui, Z. Yue, and F. Kun, "AIR-SAR-Ship-1.0: High-resolution SAR ship detection dataset," *J. Radars*, vol. 8, no. 6, p. 852–862, 2019.

[32] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era, Models, Methods Appl. (BIGSARDATA)*, Nov. 2017, pp. 1–6.

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

[34] J. Ai, R. Tian, Q. Luo, J. Jin, and B. Tang, "Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10070–10087, Dec. 2019.

[35] R. Wang, F. Xu, J. Pei, C. Wang, Y. Huang, J. Yang, and J. Wu, "An improved faster R-CNN based on MSER decision criterion for SAR image ship detection in harbor," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, Aug. 2019, pp. 1322–1325.

[36] Y. Liu, M.-H. Zhang, P. Xu, and Z.-W. Guo, "SAR ship detection using sea-land segmentation-based convolutional neural network," in *Proc. Int. Workshop Remote Sens. with Intell. Process. (RSIP)*, May 2017, pp. 1–4.

[37] 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.

[38] K. Fu, Z. Chang, Y. Zhang, G. Xu, K. Zhang, and X. Sun, "Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 161, pp. 294–308, Mar. 2020.

[39] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019.

[40] J. Jiao, Y. Zhang, H. Sun, X. Yang, X. Gao, W. Hong, K. Fu, and X. Sun, "A densely connected End-to-End neural network for multiscale and multiscene SAR ship detection," *IEEE Access*, vol. 6, pp. 20881–20892, 2018.

[41] J. Zhao, W. Guo, Z. Zhang, and W. Yu, "A coupled convolutional neural network for small and densely clustered ship detection in SAR images," *Sci. China Inf. Sci.*, vol. 62, no. 4, p. 42301, Apr. 2019.

[42] S. Zhang, R. Wu, K. Xu, J. Wang, and W. Sun, "R-CNN-Based ship detection from high resolution remote sensing imagery," *Remote Sens.*, vol. 11, no. 6, p. 631, 2019.

[43] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.

[44] J. Zhao, Z. Zhang, W. Yu, and T.-K. Truong, "A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," *IEEE Access*, vol. 6, pp. 50693–50708, 2018.

[45] Q. Li, R. Min, Z. Cui, Y. Pi, and Z. Xu, "Multiscale ship detection based on dense attention pyramid network in SAR images," in *Proc. IGARSS - IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 5–8.

[46] C. Chen, C. Hu, C. He, H. Pei, Z. Pang, and T. Zhao, "SAR ship detection under complex background based on attention mechanism," in *Proc. Chin. Conf. Image Graph. Technol.*, Singapore: Springer, 2019, pp. 565–578.

[47] Y. Qian, Q. Liu, H. Zhu, H. Fan, B. Du, and S. Liu, "Mask R-CNN for object detection in multitemporal SAR images," in *Proc. 10th Int. Workshop Anal. Multitemporal Remote Sens. Images (MultiTemp)*, Aug. 2019, pp. 1–4.

[48] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.

[49] Y. Li, Z. Ding, C. Zhang, Y. Wang, and J. Chen, "SAR ship detection based on resnet and transfer learning," in *Proc. IGARSS - IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 1188–1191.

[50] Q. Fan, F. Chen, M. Cheng, S. Lou, R. Xiao, B. Zhang, C. Wang, and J. Li, "Remote sensing ship detection using a fully convolutional network with compact polarimetric SAR images," *Remote Sens.*, vol. 11, no. 18, p. 2171, 2019.

[51] T. Zhang, X. Zhang, J. Shi, and S. Wei, "Depthwise separable convolution neural network for high-speed SAR ship detection," *Remote Sens.*, vol. 11, no. 21, p. 2483, 2019.

[52] S. Wei, H. Su, J. Ming, C. Wang, M. Yan, D. Kumar, J. Shi, and X. Zhang, "Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet," *Remote Sens.*, vol. 12, no. 1, p. 167, 2020.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351. New York, NY, USA: Springer-Verlag, 2015, pp. 234–241.

[54] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944.

[55] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS—Improving object detection with one line of code," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5562–5570.

[56] E. Xie, P. Sun, X. Song, W. Wang, D. Liang, C. Shen, and P. Luo, "PolarMask: Single shot instance segmentation with polar representation," 2019, *arXiv:1909.13226*. [Online]. Available: http://arxiv.org/abs/1909.13226

[57] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2015.

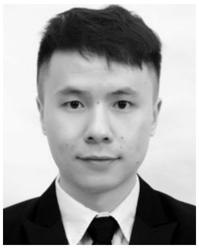
[58] W. Liu, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[59] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," 2019, *arXiv:1912.02424*. [Online]. Available: http://arxiv.org/abs/1912.02424

[60] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

[61] B. Steiner, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.

[62] K. Chen *et al.*, "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*. [Online]. Available: http://arxiv.org/abs/1906.07155



YUXING MAO (Student Member, IEEE) received the B.S. degree from Army Logistics University, Wuhan, China, in 2014, and the M.S. degree from Naval University of Engineering, Wuhan, in 2016. He is currently pursuing the Ph.D. degree in armament science and technology with Space Engineering University, Beijing, China.

His research interests include computer vision and remote sensing image processing, especially on objection detection and instance segmentation.

YUQIN YANG received the B.E. degree from the Kunming University of Science and Technology, Kunming, China, in 2018. She is currently pursuing the M.S. degree with the School of Architecture and Planning, Yunnan University, Kunming.

Her current research interest is the application of artificial intelligence in geographic information systems and remote sensing.

ZIYUAN MA (Member, IEEE) received the B.E. degree from the Shanghai University of Engineering and Technology, Shanghai, China, in 2015. He is currently pursuing the M.S. degree with the Nanjing University of Aeronautics and Astronautics, Nanjing, China.

His research interests include detection, identification and tracking of UAV ground targets, and automatic control systems.

MINGZHE LI received the B.E. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2012, and the M.S. and Ph.D. degrees in armament science and technology from Space Engineering University, Beijing, China, in 2015 and 2019, respectively.

He is currently a Research Assistant with the PLA Academy of Military Science, Beijing, China, for the period of 2019–2020. His research interest includes the passive coherent location of maritime moving target.

HAO SU received the B.E. degree from the Chengdu College, University of Electronic Science and Technology of China, Chengdu, China, in 2016, where he is currently pursuing the M.S. degree with the School of Information and Communication Engineering.

His research interests include computer vision and SAR/remote sensing image processing, especially on object detection and instance segmentation.

JUN ZHANG received the B.E. degree from Tsinghua University, Beijing, China, in 1990, and the M.S. degree from Tongji University, Shanghai, China, in 1993.

He is currently the Dean of the School of Architecture and Planning and a Professor-Level Senior Engineer with Yunnan University. From 2015 to 2020, he has committed to remote sensing and geographic information systems. His current research interest is the application of artificial intelligence in geographic information systems and remote sensing.

Prof. Zhang was the owner of Luban Award in China.