SAR Ship Detection from Complex Background Based on Dynamic Shrinkage Attention Mechanism

Sihan Shao
School of Optoelectronic Engineering, Changchun University of Science and Technology
Changchun, China
sihan_shao0511@163.com

Haoran Li
Department of electronic Engineering, Chongqing University of Post and Telecommunications
Chongqing, China
460795057@qq.com

Shifeng Wang
School of Optoelectronic Engineering, Changchun University of Science and Technology
Changchun, China
sf.wang@cust.edu.cn

2021 SAR in Big Data Era (BIGSARDATA) | 978-1-6654-0123-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIGSARDATA53212.2021.9574155

***Abstract*—With the development of synthetic aperture radar (SAR) satellite, more high-resolution SAR imageries can be obtained, which have been widely used for ship detection. SAR ship detection based on convolution neural network has shown more significant potential. However, several challenges due to its imaging characteristic remain to be addressed: 1) SAR images are severely polluted by speckles under normal conditions. 2) Complex background such as inshore building and harbor may cause background to be brighter than targets. These problems result in the blurred edges of objects in SAR images and the inefficiency of ship feature extraction, which limit the performance of marine ship detection. To solve these issues, we first analyzed the impact of these issues on the feature level and proposed a feature-level denoising method called dynamic shrinkage attention (DSA). This method can adaptively suppress the irrelevant spatial response. Experiments on the SSDCB dataset show the efficiency and feasibility of our algorithm for SAR marine ship detection in complex background.**

***Index Terms*—synthetic aperture radar; ship detection; attention mechansim; feature denoise; deep learning**

I. Introduction

Ship detection is a significant technique for marine surveillance in fields such as illegitimate fishing, maritime traffic control, and oil spill detection. Synthetic Aperture Radar (SAR) as an active microwave remote sensor is most suitable for ship detection due to its all-weather and all-time working characteristic and constant resolution even when far away from the observed targets. Whereas, unlike natural images from the optical sensor, Synthetic aperture radar as a coherent imaging system may result in inherent characteristics such as speckle noise and backscatter effect. Ships will be easily submerged in background clutter or heavy noise. Apart from the imaging mechanism, other targets in complex backgrounds like offshore cays, inshore buildings, or a harbor have similar backscattering mechanisms with ships. All these factors may harm detection performance and produce high false-positive rates.

With rapid development of deep learning, many research works have paid more attention to the deep learning-based ship detection methods. Li et al. [2] presented an improved Faster R-CNN with several techniques such as transfer learning, feature fusion and hard negative mining. Kang et al. [3] proposed cross-level fusion between deep semantic and shallow spatial features to provide complementary information for detection. Lin et al. [4] designed squeeze-and-excitation mechanisms and adopted SE block which firstly introduced attention mechanism to SAR ship detection. Although previous research has developed several improved network, these methods do not gave specific improvement projects to deal with SAR-specific imaging defects mentioned above.

To conquer the aforementioned limitation, we firstly explore the influence of complex background noise in neural network and how these noise effects network's performance in feature-level. Then to alleviate the effect of noise, we present a new method called dynamic shrinkage attention module motivated by soft thresholding function [5] and dynamic ReLU [6]. Our method as a lightweight unit which can easily inserted in any network architectures. It can effectively eliminate the noise-related features during the training process and boost feature representation ability at the same time. Experimental analyses on SAR ship detection in complex background dataset (SSDCB) illustrate that our proposed algorithm achieves better performance on detection precision with minimal additional computational cost.

The remainder of this paper is organized as follows. We will introduce our proposed methods in section II. Section III describes experiments and analyzes experimental results with SAR images. The conclusion is given in section IV.

II. Methodology

In this section, we introduce a feature-level denoising module called dynamic shrinkage attention that is lightweight and can easily embed in any detectors.

A. Dynamic Shrinkage Attention Mechanism

In classic signal denoising methods, the soft thresholding function has been generally utilized as a significant step [5]. In general, the raw image is transferred to a domain (e.g. feature map domain) in which the features of targets are highlighted and the noise-related features tend to be near-zero numbers that are unimportant. However, the non-targets that are a similar backscattering effect with ship and speckle noise in SAR images may result in false feature extraction for

978-1-6654-0123-4/21/$31.00 ©2021 IEEE

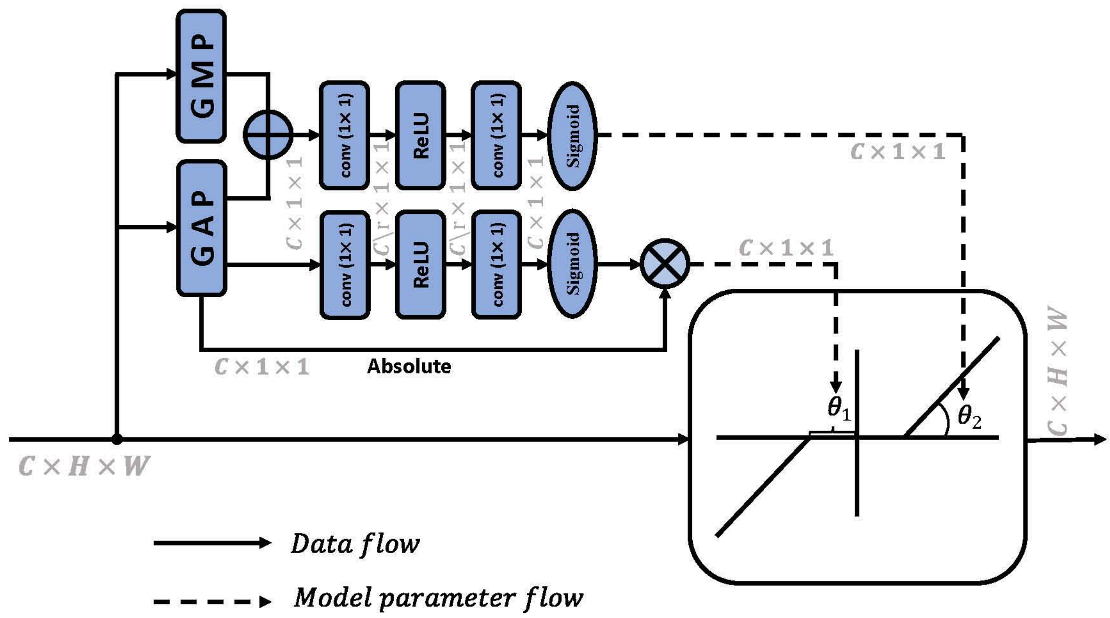

Fig. 1. The architecture of dynamic shrinkage attention.

convolution neural networks. Therefore, the soft threshold is to design a filter that can keep favorable feature information, and transform noise-related feature information to zero. Nevertheless, how to design a proper threshold in classical signal denoising algorithms has always been a challenging issue. Aiming at this question, with the help of deep learning, the filter may enable to be learned automatically using a gradient descent algorithm to avoid the trouble of artificial operation. The soft thresholding function can be indicated by:

$$\boldsymbol{x} = sign(\boldsymbol{y}) \cdot max\{|\boldsymbol{y}| - \boldsymbol{\theta_1}, 0\} = \begin{cases} \boldsymbol{y} - \theta, \boldsymbol{y} > \theta \\ \;0, \quad |\boldsymbol{y}| < \theta \\ \boldsymbol{y} + \theta, \boldsymbol{y} < \theta \end{cases} \tag{1}$$

where y is the input feature map, x is the denoising feature map, and θ is soft threshold. For the network to automatically obtain the appropriate thresholds, we build a non-linear transformation layer into the building unit $\boldsymbol{\theta_1(y)}$. Additionally, the value of threshold can be learned in this unit, which is introduced below:

$$\boldsymbol{\theta_1} = \boldsymbol{\theta_1(y)} = \sigma(g(\mathbf{GAP}(\mathbf{y}))) \cdot |\mathbf{GAP}(\mathbf{y})| \tag{2}$$

Global average pooling (GAP) aggregates spatial information and the non-linear layer is used to build up the nonlinear mapping between contextual information of feature maps and scaling factor of thresholds. σ refers to the sigmoid function that scales the parameters to the range of $(0, 1)$ and g means a non-linear layer consisted of two connection layers with the ReLU function. At the end of the processing, the corresponding scaling factors are multiplied by the absolute average value of $\boldsymbol{y}$ to get the threshold. After this step, the thresholds can be positive and keep in an acceptable range, thereby the output features are avoided to become all zeros.

Furthermore, dynamic ReLU [6] indicates that adapting a piecewise linear activation function dynamically for each input can also improve the representation ability of the neural network. Therefore, we further improve the soft thresholding function, which also can be regarded as a piecewise linear function, from static to dynamic. The modified soft thresholding can be illustrated as follows:

$$\boldsymbol{x} = (1 + \boldsymbol{\theta_2}) \cdot sign(\boldsymbol{y}) \cdot max\{|\boldsymbol{y}| - \boldsymbol{\theta_1}, 0\} \tag{3}$$

where the default α is 1.0. Compared to its static counterpart, our method is named dynamic shrinkage attention has a negligible extra computational cost, but significantly more representation capability. To compute dynamic parameters efficiently, we use not only average-pooling but also max-pooling simultaneously referring to CBAM [7]. The process of dynamic parameters is shown in Eq.(4).

$$\boldsymbol{\theta_2} = \sigma(g(\mathbf{GAP}(\mathbf{y}) + \mathbf{GMP}(\mathbf{y}))) \tag{4}$$

The overall architecture of the dynamic shrinkage attention (DSA) module is shown in Fig. 2. Several DSA units are stacked in the neural network so that discriminative features can be learned and irrelevant complex background and scattering noise in feature map are converted to zero values to alleviate the noise-related features during multiple iterations of training.

III. EXPERIMENTS AND ANALYSIS

In this section, we assess the performance of our proposed method on SAR ship detection dataset. All Experiments are implemented based on mmdetection.

TABLE I
QUANTITATIVE EVALUATIONS OF THE MARINE SHIP DETECTION IN SSDCB DATASET BASED ON TWO REPRESENTATIVE DETECTORS. DSANET50 (RESNET50 WITH DSA MODEL) OUTPERFORMS THE BASELINE METHOD.

Datasets	Detector	Backbone	Params(M)	mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
SSDCB	Faster R-CNN	ResNet-50+FPN	41.35	58.2	93.6	66.6	53.7	64.3	61.3
		DSANet-50+FPN	41.67	**59.0**	**94.1**	**68.1**	**54.2**	**65.4**	**61.9**
SSDCB	Cascade R-CNN	ResNet-50+FPN	69.15	61.4	93.4	71.8	56.2	68.0	69.4
		DSANet-50+FPN	69.47	**62.0**	**93.8**	**73.1**	**56.8**	**68.7**	**70.4**

TABLE II
COMPARISON EVALUATIONS WITH OTHER FOUR ATTENTION MECHANISMS. THE PERFORMANCE OF THIS ALGORITHM IS BASED ON FASTER R-CNN. THE SIZE OF INPUTS IS 416 $\times$ 416. THE SUPERIOR RESULTS ARE MARKED AS **BOLD**.

Backbone	Params(M)	mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
ResNet-50	41.35	58.2	93.6	66.6	53.7	64.3	61.3
SE+ResNet-50	43.88	$57.1_{(-1.1)}$	$93.3_{(-0.3)}$	$63.9_{(-2.7)}$	$52.6_{(-2.7)}$	$63.0_{(-1.3)}$	$59.1_{(-2.2)}$
ECA+ResNet-50	41.35	$57.1_{(-1.1)}$	$93.3_{(-0.3)}$	$64.1_{(-2.5)}$	$52.6_{(-2.7)}$	$63.2_{(-1.1)}$	$58.8_{(-2.5)}$
CBAM+ResNet-50	43.88	$57.0_{(-1.2)}$	$93.3_{(-0.3)}$	$63.8_{(-2.8)}$	$52.7_{(-2.6)}$	$62.9_{(-1.4)}$	$59.2_{(-2.1)}$
NL+ResNet-50	55.00	$58.6_{(+0.4)}$	$94.0_{(+0.4)}$	$66.9_{(+0.3)}$	$54.1_{(+0.4)}$	$64.7_{(+0.4)}$	$\mathbf{63.6_{(+2.3)}}$
DSA+ResNet50	41.67	$\mathbf{59.0_{(+0.8)}}$	$\mathbf{94.1_{(+0.5)}}$	$\mathbf{68.1_{(+1.5)}}$	$\mathbf{54.2_{(+0.5)}}$	$\mathbf{65.4_{(+1.1)}}$	$61.9_{(+0.6)}$

A. Dataset Introduction

To prove the validity of our proposed method, we assess it with a SAR ship dataset collected by Wang [1]. 108 Sentinel-1 images and 102 Chinese Gaofen-3 images are collected and then labeled by SAR experts in this dataset. This dataset contains in 43,819 ship chips of 256 pixels in both length and width. These chips mainly include various scales and complex backgrounds. They diversify in the field of resolution, polarization, imaging mode, incidence angle. For the SAR ship detection, to make a reasonable evaluation of our proposed framework, we use the standard COCO metrics including mAP, AP_{50}, AP_{75}, AP_s, AP_m, AP_l which are defined in Table III.

TABLE III
THE COCO FORM OBJECT DETECTION EVALUATION METRICS.

Metrics	**Metrics Meaning**
mAP	AP at IoU=0.50: 0.05: 0.95
AP_{50}	AP at IoU=0.50
AP_{75}	AP at IoU=0.75
AP_s	AP for small objects: are $a < 32^2$
AP_m	AP for medium objects: 32^2 <area $< 96^2$
AP_l	AP for large objects:area $> 96^2$

B. Experiment Results and Comparison of proposed method

To examine the effectiveness of our dynamic shrinkage attention module, we insert the DSA unit into two popular detection frameworks called Faster R-CNN and Cascade R-CNN which use ResNet50 and FPN as the backbone. For a equitable comparison, only the pretrained backbone model on ImageNet is displaced while keeping the other components in the entire detector intact.

Table I, presented the performance of the modified backbone with the DSA module on two advanced detectors on the SSDCB dataset. Comparing with our baseline algorithms, DSANet50+FPN gains 0.8% and 0.6% achievements in terms of mAP for the SSDCB. The results also prove that our proposed method boost ship detection performance and gets higher precision. Moreover, the value of AP_{75} for detectors with DSANet50 on SSDCB is 68.1% and 73.1%, respectively, which gains improvements equivalent to 1.5% and 1.3%, respectively. The value of AP_{50} also show a slight growth with 0.5% and 0.4% gains. The reason why AP_{50} improves slightly is that the coarse regression accuracy has almost reached the upper limit. The results also show our DSA module can boost more refined bounding box regression efficiently.

Here three typical scenes form SAR images are picked as the input of the benchmarking algorithm and our proposed method, and the results are indicated in Fig. 2. The first case is offshore ships with complex backgrounds. The second case is containing clusters small size ships. The third case is inshore ships and land buildings with similar backscattering effects. Comparing the three detection results in Fig. 2, it shows that the detector with the DSA module can accurately locate the ships in different scenes and avoid a high false alarm rate due to complex background and noise.

C. Comparison of Other Attention Mechanism

Next, we select a typical two-stage detector Faster R-CNN to compare DSA with other competitive attention mechanisms. The primary backbones are replaced with the corresponding attention embedded ResNet50, which are pretrained on ImageNet, for a reasonable comparison. The results presented in Table II show that the DSA unit greatly improves the values of the whole metrics. This indicates that the DSA module are able to retain the feature representation of the refined spatial information and suppress the noise-related information such as speckle noise and complex background. Therefore, the robustness of SAR ship detection can be improved.

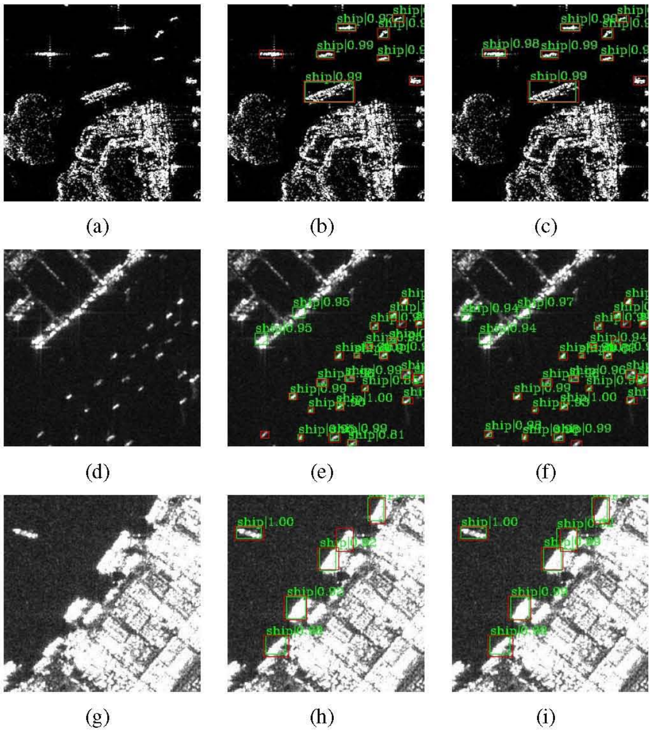

Fig. 2. The detection results in three images. Green and red rectangles are predicted boxes and ground truths, respectively. The first row demonstrates original images, the second row shows detection results by baseline, and the third row shows results by Faster R-CNN with DSANet50.

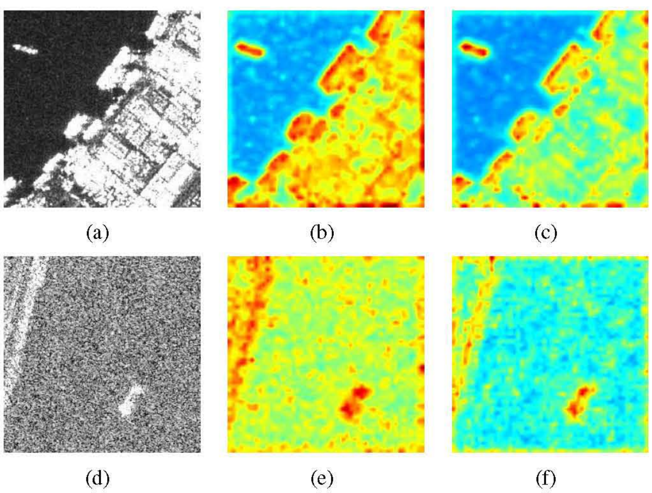

Fig. 3. Visualization of the feature maps. The second and third rows display feature maps of the second stage of baseline and our method (DSANet50), respectively.

Compared with other attention mechanisms, we find that not all attention modules proposed for optical images are suitable for detection tasks in SAR images. For instance, SE [4] and ECA [8] module, as typical channel attention modules, are efficient for many tasks of optical image. However, the performance of them was dropped slightly comparing with the benchmark algorithm. The same performance degradation also occurred on the CBAM [7], which is known as a mixed spatial and channel attention. The Non-local attention [9] (NL) module is an approach that combines the attention mechanism and non-local mean operation. We compare NL module with DSA module and find out that the NL theoretically has a feature denoising function due to non-local mean operation. Therefore, the NL module also enhances the ability of feature denoising with slight improvement. To conclude, the attention modules proposed for the optical image are not necessarily suitable for the SAR marine ship detection, and the DSA module can recover the discriminative feature map and alleviate the affect of noise and complex background in feature map.

D. *Visualization of the feature maps*

In this section, we further visualize the feature maps in neural networks and observe what important spatial information can be extracted. As shown in Fig. 3, compared with our baseline, the response of noise and complex background becomes very low because of the DSA block. In addition, the detailed features of ship are emphasized. The validity of our proposed approach is confirmed by feature map visualization.

IV. CONCLUSIONS

In this paper, we proposed the dynamic shrinkage attention module (DSA) to boost the representational capability of the network by denoising and recalibrating feature maps dynamically. The DSA module is a lightweight architectural block that can easily insert into any existing detector framework. Through comprehensive experimental investigation, we showed the feasibility and robustness of the DSA module across the SSDCB dataset in SAR image ship detection. In addition, our paper also clarifies the limitation of precedent architectures to alleviate the interference of noise-related spatial response. The proposed approach in our paper develops the accuracy of marine ship detection in SAR image with complex background and speckle noise. In the future, we hope this insight may prove useful for other vision tasks with high noise and redundant information.

ACKNOWLEDGMENT

This research is funded by the Natural Science Foundation of Jilin Province (20150101047JC) and is also supported by Artificial Intelligent and Photoelectric Perception Lab., School of Optoelectronic Engineering.

REFERENCES

[1] Wang Y, Wang C, Zhang H, et al. A SAR dataset of ship detection for deep learning under complex backgrounds[J]. remote sensing, 2019, 11(7): 765.

[2] Li J, Qu C, Shao J. "Ship detection in SAR images based on an improved faster R-CNN", Sar in Big Data Era: Models, Methods & Application, 2017.

[3] Kang M, Ji K, Leng X, et al. "Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection", Remote Sensing, Vol.9, no.8, pp.860-876, 2017.

[4] Z. Lin, K. Ji, X. Leng, et al., "Squeeze and excitation rank faster r-cnn for ship detection in sar images," IEEE Geoscience and Remote Sensing Letters 16(5), 751–755 (2018).

[5] D. L. Donoho, "De-noising by soft-thresholding," IEEE Transactions on Information Theory 41(3), 613–627 (2002).

[6] Y. Chen, X. Dai, M. Liu, et al., "Dynamic relu," in European Conference on Computer Vision,351–367, Springer (2020).

[7] S. Woo, J. Park, J.-Y. Lee, et al., “Cbam: Convolutional block attention module,” in Proceedings of the European conference on computer vision (ECCV), 3–19 (2018).
[8] W. Qilong, W. Banggu, Z. Pengfei, et al., “Eca-net: Efficient channel attention for deep convolutional neural networks.,” (2020).
[9] X. Wang, R. Girshick, A. Gupta, et al., “Non-local neural networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 7794–7803 (2018).