# Contextual Squeeze-and-Excitation Mask R-CNN for SAR Ship Instance Segmentation

Tianwen Zhang
*School of Information and Communication Engineering*
*University of Electronic Science and Technology of China*
Chengdu, China
twzhang@std.uestc.edu.cn

Xiaoling Zhang
*School of Information and Communication Engineering*
*University of Electronic Science and Technology of China*
Chengdu, China
xlzhang@uestc.edu.cn

Jianwei Li
*Department of Electronic and Information Engineering*
*Naval Aeronautical University*
Yantai, China
lgm_jw@163.com

Jun Shi
*School of Information and Communication Engineering*
*University of Electronic Science and Technology of China*
Chengdu, China
shijun@uestc.edu.cn

***Abstract*—Ship detection and ship classification using synthetic aperture radar (SAR) have been extensively studied. Yet, SAR ship segmentation unexpectedly receives less attention. Therefore, we will supplement the blank of such study in this paper. Specifically, we present a novel contextual squeeze-and-excitation Mask R-CNN (C-SE Mask R-CNN) dedicated to ship instance segmentation in SAR images. Note that the instance segmentation simultaneously considers ship detection and ship segmentation. Intuitively, C-SE Mask R-CNN is a variant of Mask R-CNN from the computer vision community. It embeds a contextual squeeze-and-excitation module (C-SE Module) into RoIAlign of Mask R-CNN to capture prominent different levels of backgrounds' contextual information. Experimental results on the public PSeg-SSDD dataset reveal the objective accuracy progress (i.e. a 1.4% AP gain on the detection task meanwhile a 0.9% AP gain on the segmentation task) of C-SE Mask R-CNN, in contrast to the vanilla Mask R-CNN.**

***Keywords*—*Synthetic aperture radar (SAR), ship instance segmentation, contextual, squeeze-and-excitation, Mask R-CNN***

## I. INTRODUCTION

Synthetic aperture radar (SAR) is an advanced active earth observation microwave sensor. It is more suitable for observing the ocean than some optical sensors due to its capacity to work around the clock regardless of the weather. Ships, as important participants in ocean activities, are very valuable targets. Their effective monitoring is conducive to marine trade supervision, marine fishery management and marine disaster relief. Thus far, ship surveillance using SAR (including ship detection [1], that is, the positioning of ship locations, and ship recognition or classification [2], that is, the recognition of specific types) has received much attention from an increasing number of scholars [3]-[6].

Up to now, diverse SAR ship detection methods have been proposed. The most classic is the constant false alarm (CFAR) detector [7], [8] whose core is to perform sea clutter modeling. Then, a fixed false alarm probability is set to identify whether each pixel belongs to a ship pixel. CFAR detectors are in pixel-level. Yet, ocean climate conditions are variable, and different sea or wind conditions will lead to the established clutter distribution model to fail. Therefore, CFAR detectors have the disadvantage of poor multi-scene migration. Another common classical approach is based on template matching. This type of method requires manual feature library in advance. However, the establishment process is both time-consuming and labor-consuming. Furthermore, it is difficult to include all types of ships and cannot detect unknown enemy ships.

In recent years, deep learning technology has been widely used in SAR ship surveillance. Thanks to the development of convolutional neural networks (CNNs), the current end-to-end automatic object detection algorithms, e.g., Faster R-CNN [10], YOLO [11], RetinaNet [12], and so on, tremendously improve the accuracy of SAR ship detection. In the SAR community, many scholars have proposed various ship detectors based on deep learning. For example, based on Faster R-CNN, Kang *et al.* [13] fused multi-layer contextual information to reduce false alarms. They experimentally enlarged the extracted proposals of RPN by ~5 times to cover more background pixels. Yet, such straightforward enlargement results in the non-compact ship boundary box regression. Still, the defects cannot obscure the virtues, and their such work is in fact a good excitation of this paper. Based on YOLO, Huang *et al.* [14] proposed one novel noisy ship direction classification-auxiliary detection technique. This technique performed the ship pixel distribution in SAR images to classify ships into different directions, which can reduce the shape difference between bounding boxes and ground truths. Based on RetinaNet, Wang *et al.* [15] modified the backbone network to conduct the multi-resolution SAR ship detection from Gaofen-3 images. The above three open reports performed SAR ship detection in the box-level, i.e., a box represents a ship. It is same as the generic object detection in the computer vision community.

Later, inspired by the rotated text detection, e.g. R2CNN [16] and EAST [17], some scholars started to adopt a rotatable box to detect ships where the ship direction needs to be estimated. For example, Chen *et al.* [18] reported an adaptive recalibration mechanism used for multiscale and arbitrary-oriented SAR ship detection; Yang *et al.* [19] added an angle estimation branch to RetinaNet to achieve rotatable ship detection; He *et al.* [20] designed a novel polar encoding scheme enabling more accurate arbitrary-oriented ship detection. Compared with the vertical-horizontal bounding box, the rotatable one can suppress more background interferences, help for refined position regression.

This work was supported in part by the National Key R&D Program of China under Grant 2017YFB0502700 and in part by the National Natural Science Foundation of China under Grants 61571099, 61501098, and 61671113.

2022 IEEE Radar Conference (RadarConf22) | 978-1-7281-5368-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/RADARCONF2248738.2022.9764228

However, a rotatable box still does not completely eliminate the influence of the background.

Recently, some scholars have tried to adopt the segmentation mask to merely assist SAR ship detection. This can ease the interference of background pixels and supervise ship bounding box regression. For example, Mao *et al.* [21] designed a score map regression network using the segmentation mask to weight output features. Their model improved the detection accuracy of inshore ships. However, their model still stays in the detection stage and does not complete the segmentation of ship pixels. Wu *et al.* [22] studied the relationship between ship instance segmentation and object detection in detail, in which an instance segmentation assisted ship detection network (ISASDNet) is proposed to improve detection performance. However, their method process is not an end-to-end implementation, contrary to the common scheme in the deep learning community. Moreover, the true instance segmentation is in fact not achieved, that is, only the detection task is completed. Sun *et al.* [23] proposed an semantic-segmentation-style side branch to highlight ship features so as to make the feature maps more discriminative for classification. However, they did not achieve accurate image pixel classification.

Note that there are only two open reports that have achieved SAR ship instance segmentation, i.e., Su *et al.* [24] and Wei *et al.* [25]. See our investigation in [26]. Specifically, Su *et al.* [24] designed a novel high-resolution feature extraction network for object instance segmentation in remote sensing images. They evaluated the model on the optical dataset. Meanwhile, they tested the model on SAR data and achieved satisfactory ship instance segmentation results. However, their model was not specifically designed for SAR, that is, they did not specifically consider the characteristics of SAR, potentially leading to some performance degradation in more complex SAR scenarios. Moreover, Wei *et al.* [25] just in fact released a high-resolution SAR images dataset (HRSID) for ship detection and instance segmentation, but they did not exclusively study the method of SAR ship instance segmentation. In their report, only some general instance segmentation models from the computer vision community were applied directly without thinking. This cannot provide effective research ideas for subsequent scholars. Given the above, SAR ship segmentation receives less attention. We think this may be caused by the late release of public data sets. Anyway, ship segmentation is the most ideal way to realize ship detection. Note that ship instance segmentation includes ship detection and ship segmentation. Instance segmentation is in fact one of the most important research tops in the field of computer vision.

Therefore, we will supplement the blank of such study in the SAR community in this paper. We select the most classical Mask R-CNN framework [27] to conduct our experiments. Specifically, we present a contextual squeeze-and-excitation Mask R-CNN (C-SE Mask R-CNN) which is a variant of Mask R-CNN. Inspired by the work of Kang *et al.* [13], we embed a contextual squeeze-and-excitation module (C-SE Module) into RoIAlign of Mask R-CNN that can effectively capture remarkable different levels of backgrounds' contextual information. Multi-level contextual information can enable the classifier to learn more background pixels. Squeeze-and-excitation mechanism can adaptively select more significant information among different contextual levels, which suppress redundant features and enlarge useful ones. We perform experiments on the public Official-SSDD dataset [26]. Results reveal that C-SE Mask R-CNN offers a 1.4% average precision (AP) gain on the detection task and a 0.9% AP gain on the segmentation task, compared with the vanilla Mask R-CNN. Moreover, we conduct some ablation experiments to confirm the effectiveness of the squeeze-and-excitation mechanism, and to determine the value range of context.

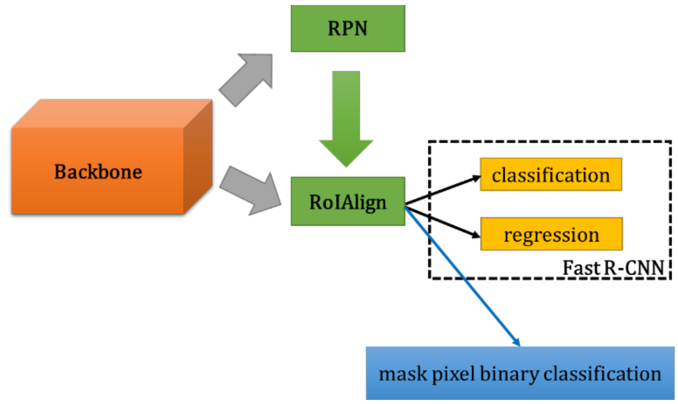



Fig. 1. Network architecture of Mask R-CNN.

## II. Methodology

### A. Mask R-CNN

Mask R-CNN is a classical and commonly-used framework for instance segmentation proposed by He *et al.* [27] in 2017. Fig. 1 is the network architecture of Mask R-CNN. Intuitively, Mask R-CNN is a complementary improvement version of Faster R-CNN. A backbone network is used to extract targets' abstract features. The extracted features are shared by a region proposal network (RPN), and a detection head, i.e. Fast R-CNN. RPN is responsible for generating some possible proposals or regions of interest (RoIs). Then, the proposals are mapped into the features maps of the backbone network so as to extract the corresponding features for the follow-up finer bounding box classification and regression in Fast-CNN [30].

However, Mask R-CNN adds an extra segmentation branch to Faster R-CNN (marked in blue in Fig. 1). This branch is used to classify the pixels in the corresponding proposal bounding box, i.e. foreground (i.e. ship) or background. Thus, Mask R-CNN follows the basic idea of detection before segmentation. Furthermore, note that Mask R-CNN proposed a RoIAlign to replace the raw RoI Pooling in Faster R-CNN. This is to solve the problem of region feature mis-alignment caused by twice quantization in RoI pooling operation. RoIAlign performs better than RoI Pooling in both detection and segmentation tasks. Especially for the pixel-sensitive segmentation task, it is able to reduce feature loss.

### B. C-SE Mask R-CNN

*1) Motivation* Mask R-CNN offers state-of-the-art instance segmentation performance on the natural optical images, i.e., the COCO dataset, but it performs poorly on SAR images. On

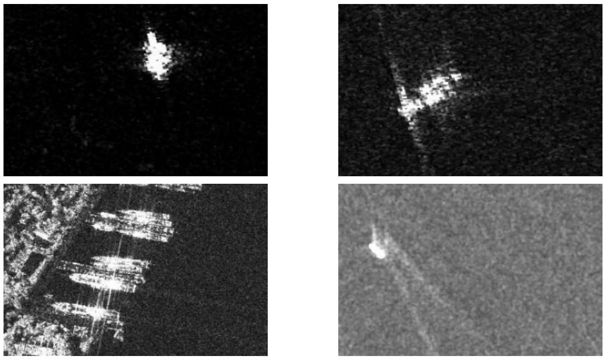


Fig. 2. Network architecture of Mask R-CNN.

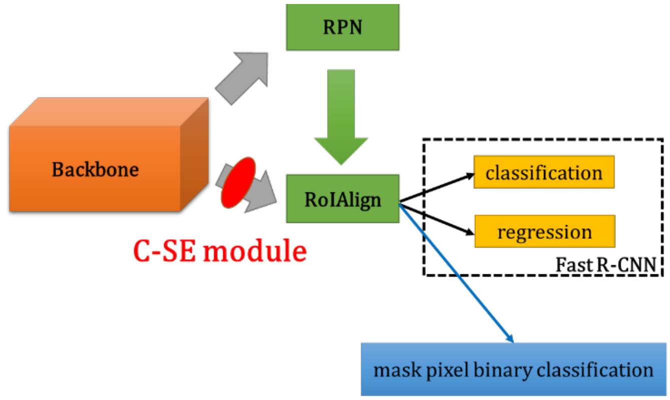
the one hand, SAR images have serious speckle noise, which poses a certain challenge to pixel classification. On the other hand, the ship in SAR image is composed of a series of discrete scattering points coming from radar scattering mechanism. This reduces the accuracy of segmentation. Moreover, the perishing cross-sidelobes will also result in inaccurate segmentation Fig. 2 shows some SAR ships. From Fig. 2, SAR ships are obviously different from optical ships. Their unclear edges increase the difficulty of segmentation.

Thus, we suggest to enlarge the scope of proposals before RoIAlign. In this way, the segmentation branch can consider more background contextual information, learn more background features, and finally search for possible ship pixels more wisely and exclude non-ship pixels. This is also in fact inspired by the work of Kang *et al.* [13].

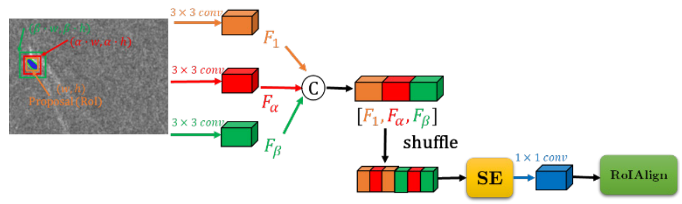
*2) Proposed C-SE Mask R-CNN* We design a contextual squeeze-and-excitation module (C-SE module) that will work before RoIAlign (mared in red), as shown in Fig.3(a). Fig. 3(b) shows the detailed implementation of C-SE module.

From Fig. 3(b), the raw size of the proposal or RoI generated by RPN is $(w, h)$ where w denotes the width and h denotes the height. We enlarge the raw proposal twice by two factors $\alpha$ and $\beta$ where $\beta > \alpha > 1$. In our work, we set $\alpha = 2.0$ and $\beta = 3.0$ experimentally and empirically. More parameter settings can be further studied in the future. Some speckle noise can be observed by the red medium-sized proposal meanwhile the ship wake can be observed by the green large-sized one. This multi-level contextual information can more fully learn the features of various background pixels.

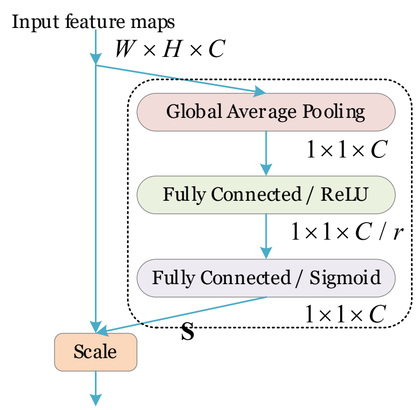
Then, the three regions are processed by $3 \times 3$ convolution so as to refine their features, i.e., obtaining $F_1$, $F_\alpha$ and $F_\beta$. Here, $F_1$, $F_\alpha$ and $F_\beta$ all have 256 channels based on the basic FPN architecture. Afterwards, the three kinds of feature maps are concatenated into a big feature map with 768 channels (i.e., $256 \times 3$) which contain all types of contextual information. Moreover, note that to avoid the collaborative consistency among each convolution so as to transmit more important information, we shuffle their channel orders randomly. Finally, in order to suppress useless contextual information and highlight salient one further, the shuffle output is processed by SE. SE [31] is able to model channel correlation to promote the flow of information among the channel.




(a)




(b)




(c)

Fig. 3. Network architecture of C-SE Mask R-CNN. (a) The overall C-SE Mask R-CNN architecture. (b) The implementation of the C-SE module. (c) The implementation of the squeeze-and-excitation (SE).

Furthermore, to facilitate input into the raw Fast R-CNN, a $1 \times 1$ convolution is used to achieve channel dimensionality reduction, i.e., from 768 to 256. This is also in fact to trade-off the calculation cost.

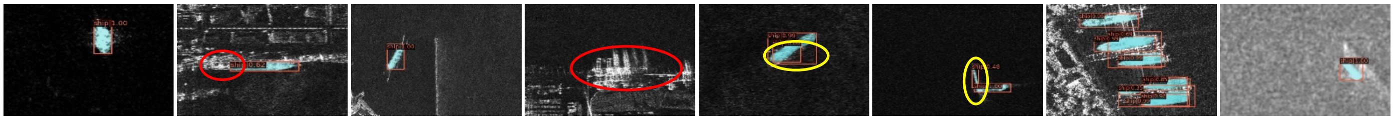
Given the above, the whole process can be described by

$$F = Conv_{1\times1}\left\{SE\left\{shuffle\begin{bmatrix} Conv_{3\times3}(P) \\ ©Conv_{3\times3}(\alpha \bullet P) \\ ©Conv_{3\times3}(\beta \bullet P)\end{bmatrix}\right\}\right\} \qquad (1)$$

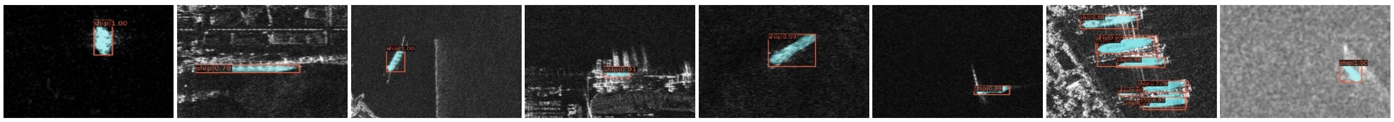
where © denotes the concatenation operation, and $P$ denotes the raw proposal region.

TABLE I. QUANTITATIVE RESULTS OF SAR SHIP INSTANCE SEGMENTATION.

| Method | Detection Task | | | | | | Instance Task | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $AP$ | $AP_{0.50}$ | $AP_{0.75}$ | $AP_s$ | $AP_m$ | $AP_l$ | $AP$ | $AP_{0.50}$ | $AP_{0.75}$ | $AP_s$ | $AP_m$ | $AP_l$ |
| Mask R-CNN | 61.5 | 90.1 | 74.0 | 61.8 | 63.0 | 8.5 | 57.7 | 88.0 | **72.7** | 57.4 | 60.4 | 21.8 |
| **C-SE Mask R-CNN (Ours)** | **62.9** | **90.2** | **74.1** | **62.8** | **65.4** | **22.6** | **58.6** | **89.2** | 71.2 | **58.3** | **60.7** | **26.7** |



(a)



(b)

Fig. 4. Qualitative results of SAR ship instance segmentation. (a) Mask R-CNN. (b) C-SE Mask R-CNN.

Fig. 3(c) shows the implementation process of SE. In Fig. 3(c), the channel compression ratio $r$ is set to 3 (i.e. from 768 to 256). $C$ denotes the channel number, i.e., $C = 768$. More details about SE can be found in our previous work in [32] and the work of Hu *et al.* [31].

Note that different from the work of Kang et al. [13] who only set a single context, we arrange three kinds of contextual ranges because it can ensure more sufficient background feature learning.

## III. EXPERIMENTS

### A. Experimental Details

Our experiments are run on a personal computer with a NVIDA RTX 3090 GPU and an Intel i9-10900KF CPU. The software framework is Pytorch and MMDetection [28]. Training optimizer is the stochastic gradient descent (SGD). The learning rate is 0.008. The batch size is set to 4. SAR images are resized to 512 pixels. ImageNet pretrained weights ResNet-50 is used to extract features. We train C-SE Mask R-CNN and Mask R-CNN by 12 epoch. At the 8-th and the 11-th epoch, the learning rate is reduced by 1 time. Feature pyramid network (FPN) architecture is selected to detect multi-scale ships. The ratio of positive and negative samples is set to 1:3. The classification loss is the cross entropy, and the regression loss is the smooth L1. Their definitions can be found in [10]. Moreover, the non-maximum suppression (NMS) is used to restrain some duplicate detections, whose intersection-over-union (IoU) threshold is set to 0.5. Others are in line with the original reports of Mask R-CNN.

### B. Dataset

Official-SSDD (PSeg-SSDD) is used in our work which is the first open dataset for SAR ship detection and segmentation. It is an official released version based on the initial version SSDD in [29]. The ship segmentation mask training labels have been provided in our previous work [26]. There are 928 training SAR images and 232 test SAR images in Official-SSDD. They are collected from RadarSat-2, TerraSAR-X, and Sentinel-1 satellites. Moreover, their polarization modes include VV, VH, HH, and HV. Their resolutions range from 1m to 10m. There are 2456 ships in Official-SSDD with different sizes and scenes.

### C. Evaluation Criteria

The COCO evaluation criteria is adopted where the average precision with an IoU threshold of $T$ is defined by

$$AP_T = \int_0^1 P_T(R_T)dR_T \tag{2}$$

where $R_T$ denotes the recall with an IoU threshold of $T$ and $P_T$ denotes the precision. In this paper, we will exhibit $AP_{0.50}$ and $AP_{0.75}$.

AP across the IoU threshold list $[0.50{:}\,0.05{:}\,0.95]$ serve as the core accuracy criteria, which is defined by their mean value, that is,

$$AP = \frac{\sum_{T\epsilon\{0.50,0.55,\cdots,0.85,0.95\}} AP_T}{10} \tag{3}$$

Moreover, the detection accuracies of small, medium, and large ships refer to the AP of the ship area $< 32^2$, $> 32^2$ but $< 96^2$, and $> 96^2$.

## IV. RESULTS

### A. Quantitative Results

TABLE I shows the quantitative results of ship instance segmentation. We evaluated the accuracy of detection task and segmentation task respectively. From TABLE I, our proposed C-SE Mask R-CNN achieves a 1.4% AP gain on the detection task, and a 0.9% AP on the segmentation one, in contrast to the vanilla Mask R-CNN. This confirms the effectiveness of C-SE Mask R-CNN.

### B. Qualitative Results

Fig. 4 presents the qualitative results of SAR ship instance segmentation. From Fig. 4, C-SE Mask R-CNN outperforms the

TABLE II. ABLATION RESULTS OF SE.

| | SE | Detection $AP$ | Segmentation $AP$ |
|---|---|---|---|
| C-SE Mask R-CNN | × | 62.4 | 58.0 |
| | ✓ | **62.9** | **58.6** |

vanilla Mask R-CNN. Some ships are not detected by Mask R-CNN, but they are detected by C-SE Mask R-CNN. Moreover, C-SE Mask R-CNN offers finer segmentation results.

### C. Ablation Experiments

Furthermore, we conduct another one experiment to confirm the effectiveness of SE. TABLE II shows the ablation results of SE. From TABLE II, SE can improve the detection $AP$ by 0.5%, and the segmentation $AP$ by 0.6%. Therefore, it is necessary to balance the channel importance after shuffle operation.

## V. CONCLUSIONS

We report a novel C-SE Mask R-CNN for SAR ship instance segmentation. Specifically, we extract the features of a proposal with three contextual information. Then, we adopt SE to model the channel relationship of three levels of contextual information. The above is named C-SE Module which will be embedded into RoIAlign of the raw Mask R-CNN so as to capture prominent different levels of backgrounds' contextual information. Finally, C-SE Mask R-CNN improves ship detection and segmentation performance in SAR images, that is, a 1.4% AP gain on the detection task meanwhile a 0.9% AP gain on the segmentation task Our experiments are conducted on the open PSeg-SSDD dataset. Moreover, we also confirm the effectiveness of SE.

## REFERENCES


[1] F. Gao, Y. He, J. Wang, A. Hussain, and H. Zhou, "Anchor-free Convolutional Network with Dense Attention Feature Aggregation for Ship Detection in SAR Images," *Remote Sensing*, vol. 12, no. 16, pp. 2619, 2020.

[2] B. Meng, M. J. Min, Z. Xi, and L. G. Wang, "A High Resolution SAR Ship Sample Database and Ship Type Classification," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020, pp. 1774-1777.

[3] T. Zhang, and X. Zhang, "Injection of Traditional Hand-Crafted Features into Modern CNN-Based Models for SAR Ship Classification: What, Why, Where, and How," *Remote Sensing*, vol. 13, no. 11, pp. 2091, 2021.

[4] J. Wang, C. Lu, and W. Jiang, "Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression," *Sensors*, vol. 18, no. 9, pp. 2851, 2018.

[5] L. Jin, and G. Liu, "An Approach on Image Processing of Deep Learning Based on Improved SSD," *Symmetry*, vol. 13, no. 3, pp. 495, 2021.

[6] T. Zhang, X. Zhang, J. Shi, S. Wei, J. Wang, J. Li, H. Su, and Y. Zhou, "Balance Scene Learning Mechanism for Offshore and Inshore Ship Detection in SAR Images," *IEEE Geoscience and Remote Sensing Letters*, pp. 1-5, early access, 2020.

[7] 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.

[8] G. Gao, "Statistical modeling of SAR images: a survey," *Sensors*, vol. 10, no. 1, pp. 775-795, 2010-01-21, 2010.

[9] J. Zhu, X. Qiu, Z. Pan, Y. Zhang, and B. Lei, "Projection Shape Template-Based Ship Target Recognition in TerraSAR-X Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 2, pp. 222-226, 2017.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788.

[12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 2020.

[13] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection," Remote Sensing, vol. 9, no. 8, pp. 14, 2017.

[14] J. Huang, Y. Niu, and W. Gan, "Ship Detection based on SAR Imaging Supervised by Noisy Ship Direction," in *International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, 2021, pp. 372-377.

[15] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery," *Remote Sensing*, vol. 11, no. 5, pp. 531, 2019.

[16] Y. Jiang, X. Zhu, X. Wang, S. Yang, W. Li, H. Wang, P. Fu, and Z. Luo, "R2 CNN: Rotational Region CNN for Arbitrarily-Oriented Scene Text Detection," in *International Conference on Pattern Recognition (ICPR)*, 2018 pp. 3610-15.

[17] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "EAST: An Efficient and Accurate Scene Text Detector," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2642-2651.

[18] C. Chen, C. He, C. Hu, H. Pei, and L. Jiao, "MSARN: A Deep Neural Network Based on an Adaptive Recalibration Mechanism for Multiscale and Arbitrary-Oriented SAR Ship Detection," *IEEE Access*, vol. 7, pp. 159262-159283, 2019.

[19] R. Yang, Z. Pan, X. Jia, L. Zhang, and Y. Deng, "A Novel CNN-Based Detector for Ship Detection Based on Rotatable Bounding Box in SAR Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1938-1958, 2021-01-08, 2021.

[20] Y. He, F. Gao, J. Wang, A. Hussain, E. Yang, and H. Zhou, "Learning Polar Encodings for Arbitrary-Oriented Ship Detection in SAR Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3846-3859, 2021.

[21] Y. Mao, Y. Yang, Z. Ma, M. Li, H. Su, and J. Zhang, "Efficient Low-Cost Ship Detection for SAR Imagery Based on Simplified U-Net," *IEEE Access*, vol. 8, pp. 69742-69753, 2020.

[22] Z. Wu, B. Hou, B. Ren, Z. Ren, S. Wang, and L. Jiao, "A Deep Detection Network Based on Interaction of Instance Segmentation and Object Detection for SAR Images," *Remote Sensing*, vol. 13, no. 13, pp. 2582, 2021.

[23] W. Sun, and X. Huang, "Semantic attention-based network for inshore SAR ship detection," in *International Conference on Digital Image Processing (ICDIP)*, Singapore, 2021, pp. 11878.

[24] H. Su, S. Wei, S. Liu, J. Liang, C. Wang, J. Shi, and X. Zhang, "HQ-ISNet: High-Quality Instance Segmentation for Remote Sensing Imagery," *Remote Sensing*, vol. 12, no. 6, pp. 989, 2020.

[25] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation," *IEEE Access*, vol. 8, pp. 120234-120254, 2020.

[26] T. Zhang, X. Zhang, J. Li, X. Xu, B. Wang, X. Zhan, Y. Xu, X. Ke, T. Zeng, H. Su, I. Ahmad, D. Pan, C. Liu, Y. Zhou, J. Shi, and S. Wei, "SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis," Remote Sensing, vol. 13, no. 18, pp. 3690, 2021.

[27] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980-2988.

[28] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. Change Loy, and D. Lin, "MMDetection: Open MMLab Detection Toolbox and Benchmark," 2019, arXiv:1906.07155.

[29] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *SAR in Big Data Era: Models, Methods and*

*Applications (BIGSARDATA)*, Beijing, China, 2017, pp. 1-6.

[30] R. Girshick, “Fast R-CNN,” in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440-1448.

[31] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011-2023, 2020.

[32] T. Zhang, and X. Zhang, “Squeeze-and-Excitation Laplacian Pyramid Network with Dual-Polarization Feature Fusion for Ship Classification in SAR Images,” IEEE Geoscience and Remote Sensing Letters, pp. 1-1, early access, 2021.