SAR-SHIPNET: SAR-SHIP DETECTION NEURAL NETWORK VIA BIDIRECTIONAL COORDINATE ATTENTION AND MULTI-RESOLUTION FEATURE FUSION

Yuwen Deng *Donghai Guan*⋆ *Yanyu Chen* *Weiwei Yuan* *Jiemin Ji* *Mingqiang Wei*

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics
Collaborative Innovation Center of Novel Software Technology and Industrialization

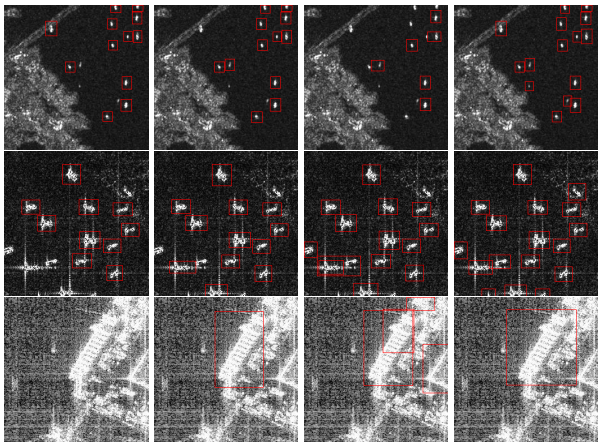
ABSTRACT

This paper studies a practically meaningful ship detection problem from synthetic aperture radar (SAR) images by the neural network. We broadly extract different types of SAR image features and raise the intriguing question that whether these extracted features are beneficial to (1) suppress data variations (e.g., complex land-sea backgrounds, scattered noise) of real-world SAR images, and (2) enhance the features of ships that are small objects and have different aspect (length-width) ratios, therefore resulting in the improvement of ship detection. To answer this question, we propose a SAR-ship detection neural network (call SAR-ShipNet for short), by newly developing Bidirectional Coordinate Attention (BCA) and Multi-resolution Feature Fusion (MRF) based on CenterNet. Moreover, considering the varying length-width ratio of arbitrary ships, we adopt elliptical Gaussian probability distribution in CenterNet to improve the performance of base detector models. Experimental results on the public SAR-Ship dataset show that our SAR-ShipNet achieves competitive advantages in both speed and accuracy.

Index Terms— SAR-ShipNet, Ship detection, Bidirectional coordinate attention, Multi-resolution feature fusion

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is an active microwave imaging sensor with long-distance observation capability in all-day and all-weather conditions and has good adaptability to monitoring the ocean. In ocean SAR images, ships are the most critical yet small targets to detect when developing a SAR search and tracking system. SAR-Ship detection aims to find the pre-defined ship objects in a given SAR scene by generating accurate 2D bounding boxes to locate them. Although many efforts have been explored to the SAR-ship detection task, it is still not completely and effectively solved, due to the non-trivial SAR imaging mechanism, where various ships are very small and blurred, and even submerged in extremely complicated backgrounds.



(a) CenterNet (b) YOLOV4 (c) EfficientDet (d) Ours

Fig. 1. Ships are often small targets and submerged in extremely complicated backgrounds. Meanwhile, SAR images inevitably contain speckle noise. These adverse factors heavily hinder accurate SAR-Ship detection. When designing a neural network model, it is natural to suppress the extracted features from the adverse factors of surroundings while enhancing the beneficial features from the ship targets. The proposed SAR-ShipNet can deal with the aforementioned problems, therefore leading to better detection results than SOTAs.

Traditional SAR target detection methods are mainly based on contrast information, geometric, texture features, and statistics. They are implemented by the hand-crafted feature extractors and classifiers. However, these methods are not only time-consuming but also lead to inaccurate detection results in complicated sea-and-land scenarios. Constant false alarm rate detectors (CFAR) [1], is one of the most commonly used techniques. [2] considers practical application situation and tries to strike a good balance between estimation accuracy and speed. [3, 4] introduce a bilateral CFAR algorithm for ship detection and reduced the influence of synthetic aperture radar ambiguity and ocean clutter.

With the development of deep learning, CNN-based detection models have emerged in multitude, which can auto-

This work was supported by the Key Research and Development Program of Jiangsu Province (BE2019012), and Joint Fund of National Natural Science Foundation of China and Civil Aviation Administration of China (U2033202). Corresponding author: D. Guan (dhguan@nuaa.edu.cn).

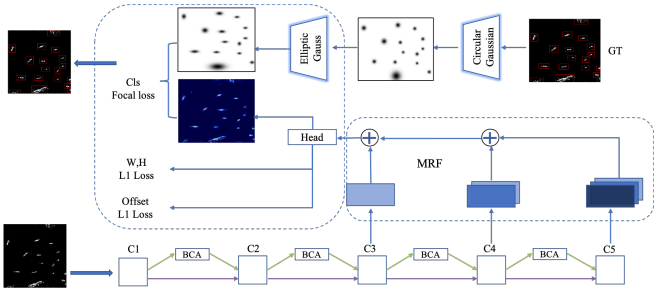


Fig. 2. Overview of SAR-ShipNet structure. SAR-ShipNet is composed of three modules: the feature extraction network that the backbone adds the attention mechanism, feature fusion: MRF, and elliptic gauss.

matically extract features and get rid of the shortcomings of manually designed features [5] for SAR-ship detection. Thus, many researchers begin to use deep learning for SAR ship detection. [6] integrates the feature pyramid networks with the convolutional block attention module. [7] introduces significant information into the network so that the detector can pay more attention to the object region. [8] proposes an anchor-free network for ship detection, using a balancing pyramid composed of attention-guided and using different levels of features to generate appropriate pyramids. [9] improves CenterNet++ and enhanced ship feature through multi-scale feature fusion and head enhancement. These detectors have achieved great results in SAR-ship detection, there are still many problems with these detectors. These problems include misclassification caused by the high similarity of ships and islands in the complex sea and land scenes, omissions in the detection of small target ships under long-distance satellite observation, and scattering noise in the SAR imaging process.

Figure 1 shows these three types of SAR-ship detection challenges, where the local regions similar to small ship targets spread over the whole background. Thus, exploring the interaction information amongst SAR image features in large-range dependencies to amplify the difference between the ship target and its background is crucial for robust detection. However, cutting-edge learning models are limited by the locality of CNNs, which behave poorly to capture large-range dependencies.

To solve these challenges, we design a high-speed and effective detector called SAR-ShipNet. We propose a new attention mechanism, i.e., bidirectional coordinated attention (BCA), to solve the effects of complex background noise and islands on ship detection. Next, we generate high-resolution feature maps in different feature layers instead of the previous solution of only generating one feature map. This can solve the problem of small ship targets and shallow pixels caused by long-distance detection and scattered noise. Finally, considering the change of detection effect caused by the aspect ratio of ships, we adopt an elliptical Gaussian probability distribution scheme to replace the circular Gaussian probability distribution scheme in CenterNet, which significantly improves the detection effect of the detector without any consumption.

2. METHODOLOGY

Motivation. SAR-ship detection encounters many challenges. Ships in SAR images are small, while backgrounds are usually complex. As a result, the small ship is easily submerged in the complex background, with a low Signal-to-Clutter Ratio (SCR). Besides, the number of ship pixels is much fewer than background pixels. That means the ship and background pixels in an image are of extreme imbalance. Meanwhile, SAR images inevitably contain speckle noise. These factors make SAR-ship detection slightly different from other detection tasks. To develop a high-precision ship detector, one should suppress the extracted features from the adverse factors of backgrounds while enhancing the beneficial features from the ship targets themselves. By completely considering both the adverse and beneficial features of SAR images with ships in them, we broadly extract different types of SAR image features, and 1) suppress data variations (e.g., complex land-sea backgrounds, scattered noise) of SAR images, and 2) enhance the features of ships that are small objects and have different aspect (length-width) ratios, therefore resulting in the improvement of ship detection. We propose a SAR-ship detection neural network (call SAR-ShipNet for short), by newly developing Bidirectional Coordinate Attention (BCA) and Multi-resolution Feature Fusion (MRF) based on CenterNet. SAR-ShipNet is composed of three modules, as shown in Figure 2. The first module is the feature extraction network that a backbone adds the attention mechanism: BCA. The second module is feature fusion: MRF. The third module is elliptic Gauss.

2.1. Bidirectional Coordinate Attention

Complicated background islands and other scattered noise affect the effectiveness of ship detection. Inspired by the coordinate attention mechanism (CA) [10], we propose a Bidirectional Coordinate Attention mechanism (BCA). CA aggregates information in two directions through Avgpooling and then encodes the generated feature maps into a pair of direction-aware and position-sensitive attention maps, which are complementarily applied to the input feature maps to enhance the representation of the object of interest. But there is a lot of noise redundancy in SAR pictures. Only using average pooling to aggregate information must have noise features to be extracted. It is necessary to ignore unnecessary redundant noise information in SAR pictures. Max pooling information is equally important, thus we propose a BCA mechanism that combines Avg and Max pooling (see Figure 3).

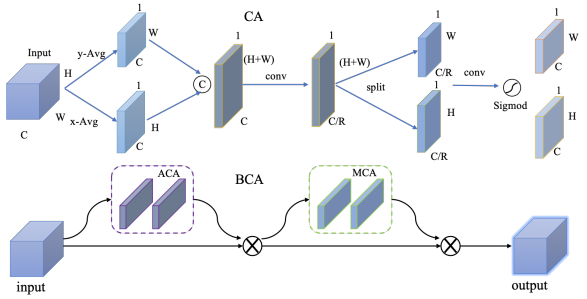


Fig. 3. Bidirectional Coordinate Attention mechanism (BCA mechanism). ACA uses Avgpooling to aggregate features, and MCA uses Maxpooling to aggregate features.

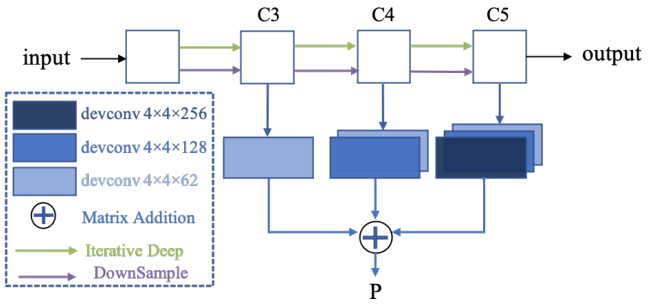


Fig. 4. MRF module. C3, C4, C5 are the feature layers output by the last three stages of Resnet50, Devconv is the deconvolution operation.

BCA is formulated as follows:

$$\begin{aligned} f_a &= \delta\left(F_1\left[\text{avgpool}\left(x_c^h\right), \text{avgpool}\left(x_c^w\right)\right]\right) \\ g^h, g^w &= \sigma\left(F_h\left(f_a^h\right)\right), \sigma\left(F_w\left(f_a^w\right)\right) \\ y_c(i,j) &= x_c(i,j) \times g_c^h(i) \times g_c^w(j) \end{aligned} \tag{1}$$

$$\begin{aligned} f_m &= \delta\left(F_2\left[\text{maxpool}\left(y_c^h\right), \text{maxpool}\left(y_c^w\right)\right]\right) \\ z^h, z^w &= \sigma\left(F_{h2}\left(f_m^h\right)\right), \sigma\left(F_{w2}\left(f_m^w\right)\right) \\ \text{output}\left(x_c(i,j)\right) &= x_c(i,j) \times g_c^h(i) \times g_c^w(j) \\ &\times z_c^h(i) \times z_c^w(j) \end{aligned} \tag{2}$$

where $x \in \mathrm{R}^{\mathrm{C}\times\mathrm{W}\times\mathrm{H}}$ is the feature map, c represents the channel index, $avgpool\left(x_c^h\right)$ and $avgpool\left(x_c^w\right)$ represents the average pooled output of the c-th channel with height h in the horizontal direction and width w in the vertical direction. [] represents the splicing operation of the feature map. F_1 represents the 1×1 convolution. δ is the non-linear activation function, $f_a \in \mathrm{R}^{\frac{C}{\mathrm{r}}\times(\mathrm{W}+\mathrm{H})\times1}$ is the intermediate feature. $f_a^h \in \mathrm{R}^{\frac{\mathrm{C}}{\mathrm{r}}\times\mathrm{H}\times1}$ and $f_a^w \in \mathrm{R}^{\frac{\mathrm{C}}{\mathrm{r}}\times\mathrm{W}\times1}$ are two vectors obtained by decomposing f_a, F_h and F_w are two 1×1 convolutions. σ is the sigmoid activation function. $g^h \in \mathrm{R}^{\mathrm{C}\times\mathrm{H}\times1}$ and $g^w \in \mathrm{R}^{\mathrm{C}\times\mathrm{W}\times1}$ are two attention weights respectively. $y_c(i,j)$ is the feature point output after avgpooling attention. Similarly, the process of using the maxpooling attention mechanism is consistent with the avgpooling attention mechanism. $\text{output}\left(x_c(i,j)\right)$ is the last output of attention through BCA. BCA makes full use of the captured position information through two different information aggregation methods so that the region of interest can be accurately captured.

2.2. Multi-resolution Feature Fusion

The Multi-resolution Feature Fusion module (MRF) is used to enhance the detailed information of small-scale ships to solve the problem of small ship targets and huge differences in surface morphology. In the deep network, if only the last feature layer is used to generate a high-resolution feature map, it is easy to lose the spatial position information of the ship, so we propose an MRF module to enhance ship features. The output of the last three stages of ResNet-50 is defined as {C3, C4, C5}. The MRF module uses three feature layers to generate three feature maps of the same size. Figure 4 shows the implementation details of the MRF module. By deconvolution of {C3, C4, C5} multiple times. Finally, we merge the three high-resolution feature maps to enhance the detailed features of the ship. The process can be defined as:

$$P = \text{dev\_}3\left(C_5\right) + \text{dev\_}2\left(C_4\right) + \text{dev\_}1\left(C_3\right) \tag{3}$$

where dev_i is the deconvolution [11] operation, i is deconvolution times. P is the total feature after fusion. After the feature fusion of the MRF module, it can significantly enhance the feature extraction of ships, reduce the detection interference caused by complex backgrounds, and improve the generalization ability of the model.

2.3. Elliptic Gauss

In the original CenterNet, the center point of the object needs to be mapped to the heatmap to form a circular Gaussian distribution. This distribution is used to measure the discrete distribution of the center point. For each GT, the key point $p \in R^2$ corresponding to category c, then calculate the key points after down sampling $\tilde{p} = \left\lfloor \frac{p}{R} \right\rfloor$. CenterNet splat all ground truth keypoints onto a heatmap $Y \in [0,1]^{\frac{W}{R}\times\frac{H}{R}\times C}$ using a Gaussian kernel $Y_{xyc} = \exp\left(-\frac{(x-\tilde{p}_x)^2+(y-\tilde{p}_y)^2}{2\sigma_p^2}\right)$, where σ_p is an object size-adaptive standard deviation. The Gaussian kernel generated by this method is a circular distribution. The parameter σ_p in the Gaussian kernel is only related to the area of GT, and the aspect ratio of GT is not fully considered. Ships in real life usually have a large aspect ratio. To fully consider the aspect ratio of GT, we are inspired by the elliptic Gaussian method in TtfNet [20]. When the key point $\tilde{p} = \left\lfloor \frac{p}{R} \right\rfloor$ is dispersed on the heatmap, the 2D Gaussian kernel Y_{xyc} is:

$$\mathbf{Y}_{x,y,c} = \exp\left(-\frac{(x-\tilde{p}_x)^2}{2\sigma_x^2} - \frac{(y-\tilde{p}_y)^2}{2\sigma_y^2}\right) \tag{4}$$

Table 1. Experimental results of SAR-ShipNet and other different SAR ship detectors.

Method	Backbone	SAR-Ship				SSDD				FPS	Parameter	Input-size
		Precision	Recall	F1	AP0.5	Precision	Recall	F1	AP0.5			
YOLOV3 [12]	DarkNet53	92.62	70.12	80	87.24	90.67	67.61	77	79.06	83	234MB	416×416
YOLOV4 [13]	DarkNet53	94.46	70.36	81	88.76	**96.94**	75.65	85	88.46	70	245MB	416×416
YOLOX [14]	DarkNet53	93.65	67.51	78	88.21	90.78	71.36	80	85.31	50	97MB	640×640
SSD300[15]	VGG16	87.79	72.48	79	82.90	93.83	33.04	49	74.07	142	91MB	300×300
SSD512 [15]	VGG16	87.48	**74.58**	81	84.42	90.07	55.22	68	70.04	80	92MB	512×512
RetinaNet[16]	ResNet50	91.52	73.24	81	88.37	39.34	51.74	45	37.53	49	145MB	600×600
CenterNet[17]	ResNet50	94.66	60.02	74	87.44	92.57	59.57	72	78.86	127	124MB	512×512
FR-CNN [18]	ResNet50	75.68	70.95	73	75.19	67.51	75	71	71.9	15	108MB	600×600
EfficientDet[19]	EfficientNet	89.48	71.77	80	85.20	94.33	39.78	56	68.27	45	15MB	512×512
SAR-ShipNet(ours)	ResNet50	**94.85**	71.31	**81**	**90.20**	95.12	**76.30**	**85**	**89.08**	82	134MB	512×512

Table 2. Ablation experiments on the SAR-Ship dataset.

CenterNet		MRF	EGS	Precision	Recall	F1	AP0.5
CA	BCA						
×	×	×	×	94.66	60.20	74	87.44
√	×	×	×	96.95	51.80	68	88.56
×	√	×	×	96.76	57.71	72	89.10
×	√	√	×	**97.06**	50.19	66	89.40
×	√	√	√	94.85	**71.31**	**81**	**90.20**

where $\sigma_x = \frac{\alpha w}{6}$, $\sigma_y = \frac{\alpha h}{6}$, α is a super parameter, w and h are the width and height of GT respectively.

2.4. Loss Function

Our training loss function consists of three parts:

$$\begin{aligned} \text{Loss} = & \frac{1}{N_{pos}} \sum_{xyc} FL\left(\hat{p}, p\right) + \frac{\lambda_1}{N_{pos}} \sum_{i} L_1\left(\hat{L}_{wh}, L_{wh}\right) \\ & + \frac{\lambda_2}{N_{pos}} \sum_{i} L_1(\hat{s}, s) \end{aligned} \tag{5}$$

where $\hat{p}$ is the confidence of classification prediction, p is the ground-truth category label, FL is Focal loss[16]. $\hat{L}_{wh}$ are the width and height of the predicted bounding box, L_{wh} are the width and height of the ground-truth bounding box. s is the offset $(\sigma x_i, \sigma y_i)$ generated by the center point (x_i, y_i) of the down-sampling process, $\hat{s}$ is the offset predicted value. N_{pos} is the number of positive samples, λ_1 and λ_2 are the weight parameters. We set $\lambda_1 = 0.1$ and $\lambda_2 = 1$.

3. EXPERIMENTS

3.1. Experimental Dataset

We directly evaluate the SAR-ShipNet model on the SAR-Ship [21] and SSDD[22] dataset. The SAR-ship dataset contains ship slices (43819) and the number of ships (59535) and the size of the all ship slices is fixed at 256×256 pixels. The SSDD data set has a total of 1160 images and 2456 ships. We randomly divide the data set into the training set, validation set, and test set at a ratio of 7:1:2.

3.2. Experimental results

We evaluate our SAR-ShipNet on 4 evaluation metrics and compare it with other methods. Table 1 shows the quantitative results of all the methods in two datasets. Compared with other detectors SAR-ShipNet achieves the best F_1, and AP on two datasets, indicating that our model has the best overall performance. This is because SAR-ShipNet uses the attention mechanism to pay more attention to ship features, and uses feature fusion to strengthen small targets and fully consider the aspect ratio of the ship. Experiments show that our model can achieve the best comprehensive performance on both the large dataset SARShip and the small dataset SSDD. Table 2 shows the ablation experimental results. It can be found that CA, BCA, MRF, and elliptic Gauss can increase the detection performance of the model. In particular, after adding the attention mechanism, the precision and AP have been improved, which shows that our model reduces the misclassification of islands and backgrounds into ships. Table 3 shows the experimental results of the effect of hyper-parameter α on SAR-Ship dataset. When α = 0.8, we get the best AP (90.20).

Table 3. SAR-ShipNet test results of different α.

Parameter	Precision	Recall	F1	AP0.5
α = 0.1	96.16	6.12	12	87.40
α = 0.2	97.84	7.2	13	88.16
α = 0.3	**98.12**	24.38	39	88.81
α = 0.4	97.58	42.11	59	89.48
α = 0.5	95.86	63.12	76	89.80
α = 0.6	97.36	55.82	71	90.03
α = 0.7	96.61	61.58	75	89.85
α = 0.8	94.85	**71.31**	**81**	**90.20**
α = 0.9	95.88	65.2	78	90.16
Circular Gaussian	97.06	50.19	66	89.40

4. CONCLUSION

In this paper, we propose an effective SAR-ShipNet for SAR-ship detection. SAR-ShipNet mainly has three modules: the BCA mechanism, the MRF module, and the elliptic Gaussian module. BCA mechanism is used to solve the problem of ship detection in complex backgrounds. It can make the model pay attention to ship features as much as possible while ignoring background noise. The MRF module is used to solve the problems of small ship sizes and shallower pixels in long-distance observation. Elliptical Gauss fully considers the influence of ship aspect ratio detection. Experimental results show that our SAR-ShipNet achieves a competitive detection performance.

5. REFERENCES

[1] Gui Gao, Li Liu, Lingjun Zhao, Gongtao Shi, and Gangyao Kuang, "An adaptive and fast cfar algorithm based on automatic censoring for target detection in high-resolution sar images," *IEEE transactions on geoscience and remote sensing*, vol. 47, no. 6, pp. 1685–1697, 2008.

[2] Gui Gao, Kewei Ouyang, Yongbo Luo, Sheng Liang, and Shilin Zhou, "Scheme of parameter estimation for generalized gamma distribution and its application to ship detection in sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 3, pp. 1812–1832, 2016.

[3] Xiangguang Leng, Kefeng Ji, Kai Yang, and Huanxin Zou, "A bilateral cfar algorithm for ship detection in sar images," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1536–1540, 2015.

[4] Zhenwei Shi, Xinran Yu, Zhiguo Jiang, and Bo Li, "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4511–4523, 2013.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[6] Zongyong Cui, Qi Li, Zongjie Cao, and Nengyuan Liu, "Dense attention pyramid networks for multi-scale ship detection in sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8983–8997, 2019.

[7] Lan Du, Lu Li, Di Wei, and Jiashun Mao, "Saliency-guided single shot multibox detector for target detection in sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3366–3376, 2019.

[8] Jiamei Fu, Xian Sun, Zhirui Wang, and Kun Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in sar images," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[9] Haoyuan Guo, Xi Yang, Nannan Wang, and Xinbo Gao, "A centernet++ model for ship detection in sar images," *Pattern Recognition*, vol. 112, pp. 107787, 2021.

[10] Qibin Hou, Daquan Zhou, and Jiashi Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13713–13722.

[11] R. Fergus, Graham William Taylor, and MD Zeiler, "Adaptive deconvolutional networks for mid and high level feature learning," in *International Conference on Computer Vision*, 2011.

[12] Joseph Redmon and Ali Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[13] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[14] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[16] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[17] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl, "Objects as points," *arXiv preprint arXiv:1904.07850*, 2019.

[18] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, 2015.

[19] Mingxing Tan, Ruoming Pang, and Quoc V Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10781–10790.

[20] Z. Liu, T. Zheng, G. Xu, Z. Yang, H. Liu, and D. Cai, "Training-time-friendly network for real-time object detection," 2019.

[21] Yuanyuan Wang, Chao Wang, Hong Zhang, Yingbo Dong, and Sisi Wei, "A sar dataset of ship detection for deep learning under complex backgrounds," *remote sensing*, vol. 11, no. 7, pp. 765, 2019.

[22] Jianwei Li, Changwen Qu, and Jiaqi Shao, "Ship detection in sar images based on an improved faster r-cnn," in *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*. IEEE, 2017, pp. 1–6.