IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10283453

GL-NET:GAUSSIAN LEADING NETWORK FOR SAR SHIP DETECTION

Zenghao Chen[1], *Yanan You*[1*], *Wenli Zhou*[1], *Gang Meng*[2]

[1] School of Artificial Intelligence, Beijing University of Posts and Telecommunications
[2] Beijing Institute of Remote Sensing Information
Beijing, China

ABSTRACT

Ship detection in synthetic aperture radar (SAR) images is a basic and challenging task in marine monitoring. There has been made remarkable achievements in recent years. However, the existing detection methods still have the problem of ambiguous location information. It leads to the lack of model pertinence, and the poor discrimination of foreground and neighboring background. In this paper, we propose a ship detection method based on FSAF. In particular, we design a Gaussian Leading Block (GLB) capable to eliminate the interference of neighboring background information in the ground truth. Experiments on the HRSID dataset show that the proposed method achieves a 3.6% Average Precision (AP) improvement over the baseline at a low cost.

***Index Terms*—** Synthetic aperture radar (SAR), ship detection, remote sensing, adaptive Gaussian Mask.

1. INTRODUCTION

SAR is an active radar with all-day and all-weather characteristics and an indispensable part of the remote sensing field. Therefore, compared with widely known optical images, SAR plays a more important stable role in maritime target monitoring, maritime rescue and port security tasks. Internationally, the application of deep learning architecture in the field of remote sensing has been widely recognized. However, the ship detection task in SAR is still full of challenges due to the lack of detailed information of SAR image targets and the confusion of complex offshore environment. Those are represented by the easy loss of dense small targets and serious false alarms on the bank in ship detection [1].

With the wave of deep learning, feature extraction based on CNN has gradually replaced CFAR and become the mainstream SAR detection algorithm. In recent years, the anchor-free algorithm has been favored by many scholars for its lightweight and efficient characteristics. In [2], the K-nearest neighbor method is used to further extract the features of the candidate frame and the scattering point feature is used as the pointcut to realize target recognition. It greatly improves the accuracy of anchor-free detection. In [3], Context-based deformable (CBD) module is proposed to combine deformable convolution and context attention with anchor-free algorithm. It makes full use of the advantages of anchor-free algorithm for better regression effect on objects with multiscale targets. Moreover, [4] and [5] employ novel fusion mode between layers to enhance the fusion mode of different feature levels.

At the same time, various FPN-based context frameworks are proven to be effective for the SAR detection of multi-scale targets [6, 7, 8]. [9] further integrates the upper and lower layers of FPN through the self-attention mechanism to make up for the loss of upper spatial information and lower feature semantic information. [10] enhances the top layer features by adding dense connection and self-attention branch roads to make up for the lack of information in the high layer of FPN. Pre-detection is added before each feature fusion in [11] as an auxiliary prediction to overcome the semantic gap between layers. However, there is a drawback in FPN that the sample allocation mechanism of scale-guide is adopted for samples. Simply taking the size of the target as the only standard for sample allocation will introduce serious human interference and cause irreparable damage to the model. In this regard, FSAF [12] and SAPD [13] propose to adopt the sample allocation strategy with a loss-guide mechanism instead. In FSAF, the target is sent to all layers indistinguishably. Afterwards,The most suitable layer is selected by the comparison of loss, so as to improve the model performance greatly. However, in FSAF, the model treats every point in the target and surrounding areas in the labeling box without discrimination. As a result, the model cannot track the characteristic information of the target entity appropriately and the labeling background information may be an unstable threat misleading the model.

In this paper, we propose Gaussian Leading Network based on FSAF. It adaptively distinguishes labeling targets and labeling background in the labeling boxes through the Gaussian Masks as an attention allocator. At the samt time, the different contribution of neighboring and distant background can be preserved. We provide three optional Gaussian Masks corresponding to the different orientation information for each target and use loss as the guide to teach our model to select the correct template automatically.

*Corresponding to youyanan@bupt.edu.cn.

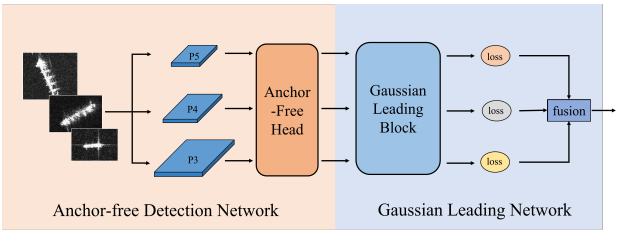

Fig. 1. Structure of Gaussian Leading Network. Each target is uniformly input to all levels in FPN. Anchor-Free Head ouputs the initial loss of each level. GL-Net outputs losses of each target and are fused as final loss

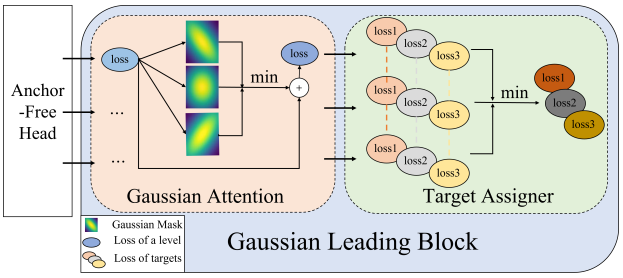

Fig. 2. Structure of Gaussian Leading Block. The loss of every target is refined by Gaussian Attention module and finally assigned to single level by Target Assigner module.

2. PROPOSAL METHODS

2.1. Overview of the Proposed Method

In this section, we present the details of the proposed network, whose structure is shown in Fig.1. It includes two parts, namely the anchor-free detection network and the Gaussian Leading Network. In the network, the anchor-free detection network carries out undifferentiated detection on all targets, outputs initial loss from the detector and inputs it into the GLB for loss-level attention guidance and sample allocation.

2.2. Gaussian Leading Block

GLB consists of two main parts, the Gaussian Attention (GA) and the Target Assigner (TA) as is shown in Fig.2. GA aims to allocate attention towards the labeling target of horizontal box, so as to compensate the ground-truth misleading inherent in the labeling method. It tells the network to pay more attention to the labeling target within the labeling area and appropriately suppress the labeling background. We adopt the horizontal box as our labeling method, which is a classic annotation method with the two-point annotation format. Since the minimal outer rectangle, a mix of target information and background information, is outlined as the ground truth, the horizontal box tends to be a large box with a small target. This may lead to the ground-truth misleading. On the other hand, there is a noticeable issue in SAR images that a lot of target details are hidden and it is necessary to enrich the features. Therefore, we propose to use adaptive Gaussian Masks to help the model correctly handle the horizontal box. In this way, the model can recognize accurate foreground information and take the background information around as auxiliary information to improve the detection performance of the model. In each feature layer, we first obtain the initial loss heatmap by the anchor-free detection network. Then, we provide three candidate Gaussian Masks for each target that are adaptive to fit the size of the ground truth, so as to accurately extract the direction information of the target. The Gaussian Masks are generated by $f(x, y)$:

$$f(x,y) = \lambda(2\pi\delta_1\delta_2\sqrt{1-\rho^2})^{-1} exp[-\frac{1}{2(1-\rho^2)} * (\frac{(x-\mu_1)^2}{\delta_1^2} + \frac{(y-\mu_2)^2}{\delta_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\delta_1\delta_2})] \quad (1)$$

$$\mu_1 = w/2; \mu_2 = h/2; \quad (2)$$

$$\delta_1 = w/r; \delta_2 = h/r; \quad (3)$$

where $\lambda = w * h$ is the impact factor, used to adjust the influence of the mask size. r=3 is the variance coefficient, adjusting the aggregation effect of Gaussian Masks. ρ, representing the correlation coefficient, is the dependent variable of Gaussian Masks changes, were set to 0.7, 0.7 and 0. μ and δ are associated with the wide and high of ground truth, making the Gaussian Masks automatically adapt the target box.

Furthermore, inspired by the loss-guide mechanism, we use the loss to guide the choice of Gaussian Masks. We utilize 3 masks to get the reconstruction loss respectively and select the minimum reconstruction loss plus a skip-connection operation to update the loss. Experiments show that GA can effectively generate the Gaussian Masks that best fits the truth value target, and improve the recognition ability of the target to a considerable extent. In addition, we add a skip-connection operation as a insurance of original information.

The TA module is responsible for assigning targets to appropriate feature layers through a loss-guide mechanism, which avoids the artificial scale-guide allocation strategy. This improves the flexibility and adaptability of the model. In the FPN framework, the idea of multi-layer dive-and-conquer according to scale is adopted to deal with the problem of multi-scale objectives. As a result, each single object is initially limited to a certain level based on its size, and this "one-shot" pattern deprives the object of the possibility that it may be accurately detected by other levels. Therefore, TA proposed a sample allocation strategy based on loss, in o

to ensure that every layer of FPN has the opportunity to try detecting the target.

As shown in Fig.2, we input the reconstruction loss refined by GA into the TA module. TA selects the smallest reconstruction $loss_(t_i)$ as the final output of the network to perform gradient feedback operations:

$$loss(t_i) = \min_j loss_{p_j}(t_i) \tag{4}$$

where t_i represents a target, and $loss_(p_j)$ represents the loss of t_i in the j-th feature layer.

2.3. Training Strategy

In the proposed network, we choose FocalLoss[14] as classification loss and IoULoss as positioning loss. During training, we used GL-Net to refine and optimize the target loss. Target loss is composed of elements in the corresponding position of the initial loss map output from the anchor-free detection head. The final loss is obtained by the combination of all reconstruction loss and initial loss. Finally, the combination of all target reconstruction loss results in the final loss as the final training loss and Stochastic Gradient Descent is used for model parameter iteration

3. EXPERIMENTS

3.1. Dataset and Implementation

We choose the HRSID dataset as the benchmark to evaluate the performance of the model. HRSID was proposed by Wei[15], including 5604 SAR images with 800×800 pixels. The dataset includes 16,951 ships with a variety of resolutions ranging from 0.5 m to 3 m. The ships in HRSID are distributed in a dispersed manner, part of which are regarded as small targets with high detection difficulty and representativeness. In terms of dataset partitioning, we selected 3642 images for training and 1962 images for testing. In addition, we selected Avergae Precision(AP) as the evaluation metric of the model.

3.2. Implementation Details

We use the pre-trained Resnet-50 as the feature extraction network. We conduct the experiments using Pytorch on the Ubuntu platform, and 2 NVIDIA Tesla T4 GPUs are used. The initial learning rate was set to 0.001, warmup iterations to 500 with linear warmup mode adopted and a total of 24 epochs were trained.

3.3. Results and Discussion

Table 1 shows the performance of our model and other comparison models in the data set. It can be seen from the table that our model has been significantly improved after adding the GLB module on the basis of the FSAF, and has achieved the optimal test effect and gained a 3% increase compared to baseline, indicating that it is necessary to further refine the target information in the ground truth area, and our Gaussian Leading Block has had a remarkable effect on it. Moreover, according to the visual results shown in Fig.3, it's obviously that the inclusion of GL-Net is beneficial for model to identify the target more accurately, achieve the purpose of increasing the precision and reducing false alarm rate.

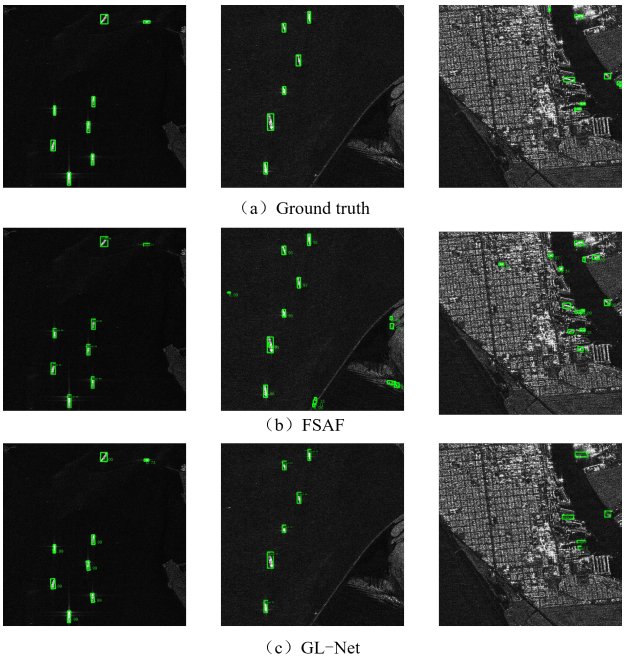

Fig. 3. Visual detection results of the proposed method. The proposed method has lower false alarm rate and higher precision.

Table 1. Experimental results on HRSID. All of the values denote the accuracy in percentage.

Method	Backbone	AP(%)	$AP_{50}(\%)$	$AP_{75}(\%)$
Faster RCNN	Resnet50	60.0	80.6	63.8
FCOS	Resnet50	63.7	82.7	62.4
FSAF	Resnet50	64.3	88.6	74.8
ours	Resnet50	**67.9**	**90.2**	**78.2**

4. CONCLUSION

In this paper,we propose a Gaussian Attention anchor-free detection method based on FSAF, aiming at solving the problem of unclear target definition and poor detection performance caused by ground-truth misleading and missing target direction information in traditional horizonta labeling method. After a detailed introduction of the proposed method structure

and process, we conduct a scientific simulation experiment on the HRSID data set. Experimental results show that this method can achieve higher detection accuracy and further improve the effectiveness of ship detection in SAR images. Our future work will focus on how to differentiate and fuse the profile features and textural features of targets for feature extension to make up for the shortage of detailed information in SAR images.

5. ACKNOWLEDGEMENT

This research was supported in part by the National Natural Science Foundation of China under Grant 62101060; and in part by the Beijing Natural Science Foundation, China, under Grant 4214058.(author: Yanan You.)

6. REFERENCES

[1] Q. Hu, S. Hu, and S. Liu, "Banet: A balance attention network for anchor-free ship detection in sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[2] X. Ma, S. Hou, Y. Wang, J. Wang, and H. Wang, "Multiscale and dense ship detection in sar images based on key-point estimation and attention mechanism," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[3] X. Sun, Y. Liu, Z. Yan, P. Wang, W. Diao, and K. Fu, "Sraf-net: Shape robust anchor-free network for garbage dumps in remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6154–6168, 2021.

[4] Z. Wang, J. Guo, L. Zeng, C. Zhang, and B. Wang, "Mlffnet: Multilevel feature fusion network for object detection in sonar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2022.

[5] X. Wang, S. Wang, C. Ning, and H. Zhou, "Enhanced feature pyramid network with deep semantic embedding for remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7918–7932, 2021.

[6] T. Zhang, X. Zhang, and X. Ke, "Quad-fpn: A novel quad feature pyramid network for sar ship detection," *Remote Sensing*, vol. 13, no. 14, p. 2771, 2021.

[7] P. Chen, Y. Li, H. Zhou, B. Liu, and P. Liu, "Detection of small ship objects using anchor boxes cluster and feature pyramid network model for sar imagery," *Journal of Marine Science and Engineering*, vol. 8, no. 2, p. 112, 2020.

[8] K. Zhou, M. Zhang, H. Wang, and J. Tan, "Ship detection in sar images based on multi-scale feature extraction and adaptive feature fusion," *Remote Sensing*, vol. 14, no. 3, p. 755, 2022.

[9] D. Zhang, H. Zhang, J. Tang, M. Wang, X. Hua, and Q. Sun, "Feature pyramid transformer," in *European conference on computer vision*. Springer, 2020, pp. 323–339.

[10] J. Cao, Q. Chen, J. Guo, and R. Shi, "Attention-guided context feature pyramid network for object detection," *arXiv preprint arXiv:2005.11475*, 2020.

[11] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "Augfpn: Improving multi-scale feature learning for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 595–12 604.

[12] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 840–849.

[13] C. Zhu, F. Chen, Z. Shen, and M. Savvides, "Soft anchor-point object detection," in *European conference on computer vision*. Springer, 2020, pp. 91–107.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[15] 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. 120 234–120 254, 2020.