SSE-Ship: A SAR Image Ship Detection Model with Expanded Detection Field of View and Enhanced Effective Feature Information

Liping Zheng[1], Liang Tan[1], Liangjun Zhao[1*], Feng Ning[1], Bo Xiao[1], Yang Ye[2]

[1]College of Computer Science and Engineering, Sichuan University of Light Chemical Industry, Zigong, China
[2]College of Resources, Sichuan Agricultural University, Ya'an, China
Email: *149189602@qq.com

Abstract

In this paper, we propose a SAR image ship detection model SSE-Ship that combines image context to extend the detection field of view domain and effectively enhance feature extraction information. This method aims to solve the problem of low detection rate in SAR images with ship combination and ship fusion scenes. Firstly, we propose STCSPB network to solve the problem of ship and non-ship object fusion by combining image contextual feature information to distinguish ship and non-ship objects. Secondly, we combine SE Attention to enhance the effective feature information and effectively improve the detection accuracy in combined ship driving scenes. Finally, we conducted extensive experiments on two standard base datasets, SAR-Ship and SSDD, to verify the effectiveness and stability of our proposed method. The experimental results show that the SSE-Ship model has $P = 0.950$, $R = 0.946$, $mAP\_0.5:0.95 = 0.656$ and $FPS = 50$ on the SAR-Ship dataset and $mAP\_0.5 = 0.964$ and $R = 0.940$ on the SSDD dataset.

Keywords

Ship Detection, SSE-Ship, STCSPB, SE Attention

1. Introduction

SAR image ship detection is an important but challenging task in maritime target detection, which requires networks to predict ships in SAR images. Ship detection can benefit many applications, for example, in the field of maritime disaster relief and marine safety monitoring to quickly and effectively target suspicious targets and take appropriate measures. Benefiting from the effective feature

representation of convolutional neural networks in deep learning, many methods [1] [2] [3] [4] [5] have achieved better results. However there are still some challenges to accurate ship inspections. As shown in **Figure 1(a)**, ships and non-ship objects have different semantics, but they have similar features (e.g., white light dots). It is difficult to distinguish them without a better combination of image contextual information. On the other hand, since the small target ships in SAR images have a single feature and large-size ships have more local features, it is difficult to detect all ships accurately if there is multi-size ship information in the image and no enhancement of effective feature information, as shown in **Figure 1(b)**. Therefore, operations to expand the detection field of view area and enhance the effective feature information are necessary in solving the problems of ship combination and ship and non-ship fusion. Existing ship detection methods either perform fast detection of small ship targets only [5]-[12] or build deeper networks for accurate detection of medium and large size ships [4] [13]-[18], but they do not discuss the above two types of problems, which leads to inaccurate detection performance in different scenarios.

Other common problems in ship detection are multi-ship combination movements and dock interference. As shown in **Figure 1(c)**, the high degree of integration of the ship with the quay causes the ship to be difficult to be identified. And in **Figure 1(d)**, multiple ships are combined together causing the overall structure to lose the typical features of a ship.
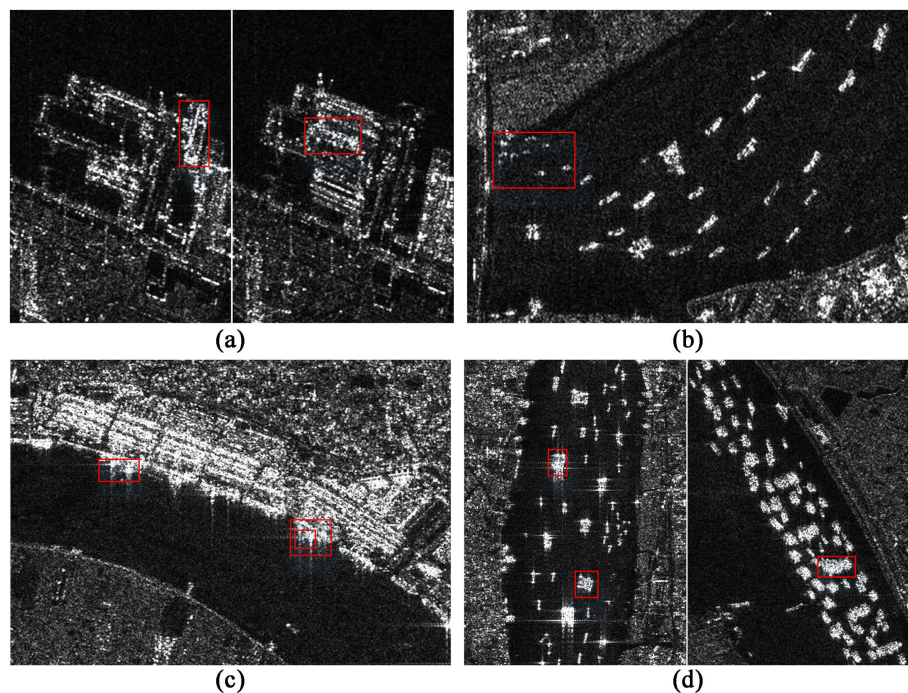
Figure 1. Case illustration of ship detection. (a) Detection results without combining image contextual information. Ships in the red boxes will be missed due to loss of connection to context. (b) Detection results without enhanced feature information. It is easy to miss the detection of small-sized ships in the red box. (c) A scene where the ship and the coastal pier are fully integrated. (d) The case of multiple ships traveling in combination.

To overcome these drawbacks, the SSE-Ship model is proposed in this paper. A new ship target detector that combines image context and enhances the effective feature information of the feature map. The model effectively enhances the effective feature information while modeling through the context of the image range. Specifically, an image containing one or more ships is used as input. First, the image features with different depths are extracted from the SAR ship images using CNN backbone networks. Then, the STCSPB network proposed in this paper is used to generate a new global memory feature map by combining the relationships of each feature layer. Secondly, the spatial attention mechanism SE Attention is introduced to enhance the effective feature information in the feature map to generate a predictable feature map. Finally, multi-task loss functions, classification loss and regression loss are constructed. The results show that the SSE-Ship detection model largely outperforms the existing methods. Specifically, P = 0.944, R = 0.940, and mAP_0.5:0.95 = 0.647 on the SSDD [19] dataset, mAP_0.5:0.95 = 0.656 and FPS = 50 on the SAR-Ship [20] dataset, mAP_0.5 = 0.978 and mAP_0.5: 0.95 = 0.667 on the HRSID [21] dataset.

2. Related Work

The SSDD dataset was first proposed by Li *et al.* [19], and provides the corresponding ship real frame and label information. The SAR-Ship dataset was first proposed by Wang *et al.* [20], and this data and widely used in the training process of maritime target detection models. The HRSID [21] dataset is a novel dataset for ship detection, semantic segmentation and instance segmentation tasks, first proposed by the University of Electronic Science and Technology in January 2020.

Unlike simple small target detection [22] [23] [24] [25] and deep network construction [18] [21] [26] [27], the goal of SSE-Ship detection is to accurately detect and locate all ships in SAR images by linking contextual modeling and enhancing effective feature information. [28] modeled global semantic information using a Swin Transformer-based model. [29] proposed a multiple attention mechanism interaction and scale enhancement network for SAR ship instance segmentation.

In this paper, YOLOv7 is used as the key component. YOLOv7 [30] is the seventh version of the YOLO algorithm, which is also improved based on the algorithm idea of YOLOv5. According to the network width and depth difference, YOLOv7 is further subdivided into YOLO7m, YOLOv7x, Yolov7-Tiny, and other versions. First, the algorithm resizes the input image to 640 × 640 and inputs it to the Backbone network. Second, three-layer feature maps with different sizes are chosen as the output through the Head layer network. Finally, the prediction result is obtained.

3. Method

3.1. Network Architecture

Figure 2 illustrates the overall design architecture of the SSE-Ship model
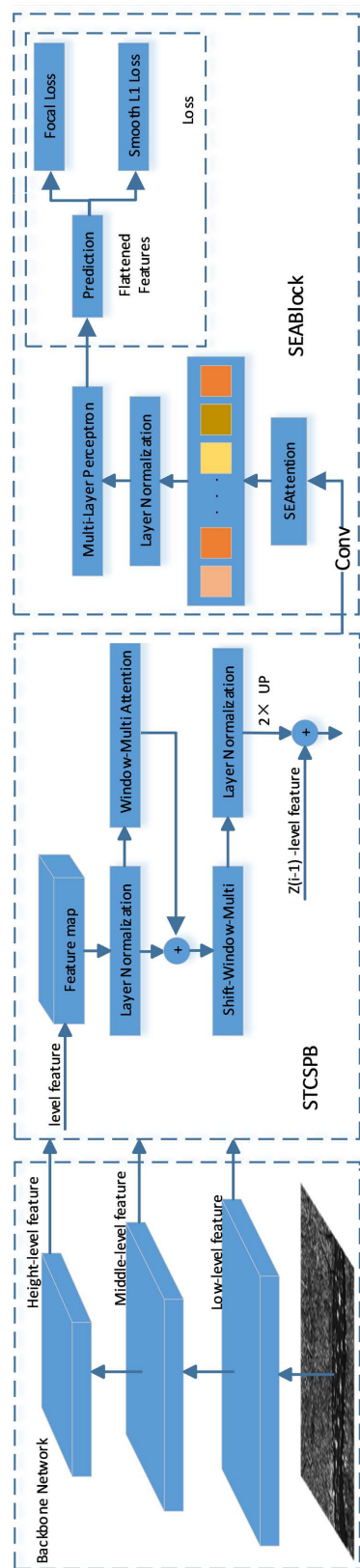
Figure 2. Overall architecture of the SSE-Ship model. It consists of 4 main components: backbone network, image contextual feature information fuser, feature enhancement block and multitasking loss function.

proposed in this paper. SSE-Ship consists of four main components: 1) a backbone network for extracting feature maps of different depths from the input image 2) a global fuser STCSPB network for generating global memory feature maps, 3) a feature enhancement block SEA Block for enhancing the effective feature information in the feature maps, and 4) a multi-task loss function for computing classification and regression errors.

The backbone network part, which feeds the input image $x \in \mathbb{R}^{3 \times H \times W}$ into the backbone network with YOLOv7 [30] as the key component, finally outputs three feature maps of different depths.

3.2. STCSPB Module

Swin Transformer [31] employs a layered Transformer solution to improve efficiency by confining the self-attention computing to non-overlapping local Windows while still allowing cross-window wiring. There are four stages in Swin Transformer, each containing a Block. This paper employs a simple data set of SAR images, which does not require too much computational attention. Therefore, this paper refers to a block (Swin-T Block) in Swin Transformer as the main content of the STCSPB module. The STCSPB structure is shown in **Figure 3**.

Figure 4 shows the Swin-T Block structure. The first part comprises the two-Layer Normalization (LN), a Window-based Multi-head Self-Attention (W-MSA), and a Mul-ti-Layer Perceptron (MLP). The W-MSA module divides the image into non-coincident windows to reduce the model's calculation amount. In the second part, to solve the cross-window information interaction problem, W-MSA in the first part is modified to the Shift-Window based Multi-head Self-attention (SW-MSA), and the rest of the part employs the LN and MLP for residual connection.
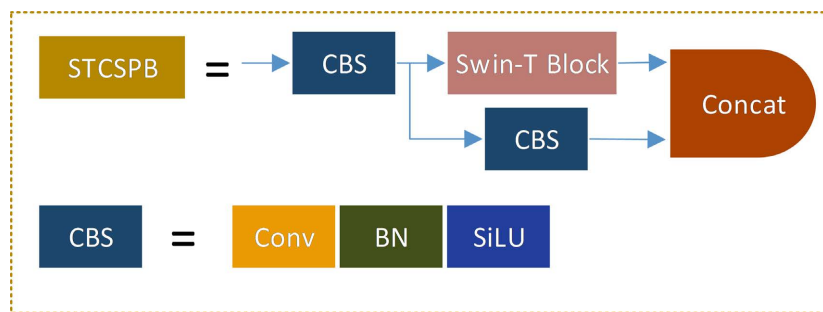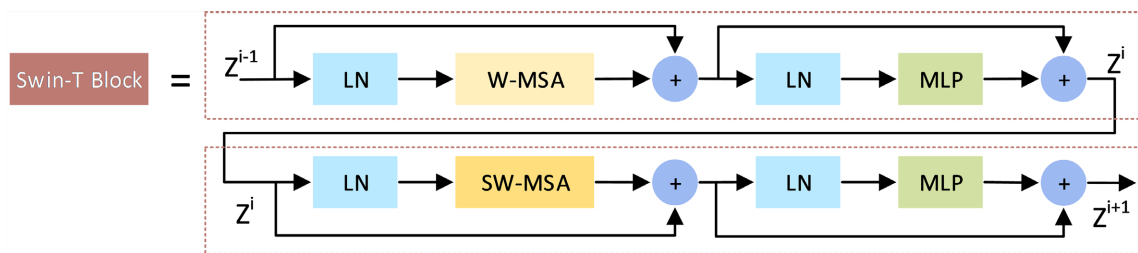

Figure 3. STCSPB structure diagram.

Figure 4. Swin-T Block structure diagram.

3.3. Network Structure Improvement

The SE (Squeeze and Excitation) [32] module first squeezes the feature graph obtained by the convolution to extract the channel-level global features. Then, the global features are subject to the Exception operation, and the weight of different channels is obtained by learning their relationship. Finally, the final feature is obtained by multiplying the original feature map. **Figure 5** shows the SE Attention structure.

$F_{tr} : X \rightarrow U, X \in R^{h' \times w' \times C'}, U \in R^{h \times w \times C}$ in **Figure 5** represents the convolution operation. The input convolution kernel is $V = [v_1, v_2, \cdots, v_c]$ where v_c represents the c-th convolution kernel. The output is $U = [u_1, u_2, \cdots, u_c]$. u_c is described with Formula (1), where * Represents the convolution operator, and v_c^s represents the 2-D kernel convolution of the s channel.

$$u_c = v_c * X, X = \sum_{s=1}^{C'} v_c^s * x^s \tag{1}$$

F_{sq} is a Squeeze operation, as the global average pooling method. It can encode the entire spatial feature on a channel into a global feature. The Squeeze operation is shown in Formula (2).

$$z_c = F_{sq}(u_c) = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} u_c(i,j), z \in R^c \tag{2}$$

The Squeeze operation obtains the global description characteristics. Next, the Exception operation is utilized to capture the relationship between channels. F_{ex} mainly adopts the sigmoid method. The Exception operation is shown in Formula (3).

$$s = F_{ex}(z, w) = \sigma(g(z,w)) = \sigma(W_2 ReLU(W_1 z)) \tag{3}$$

where $W_1 \in R^{\frac{C}{r} \times C}, W_2 \in R^{C \times \frac{C}{r}}$. In order to reduce the complexity of the model and improve its generalization ability, it also contains two fully connected layer structures and employs the ReLU activation.

Finally, each learned channel's sigmoid activation value (0~1) is multiplied by the original feature on *U*, as shown in formula (4).

$$\overline{x}c = Fscale(u_c, s_c) = s_c \cdot u_c \tag{4}$$

3.4. Loss Function

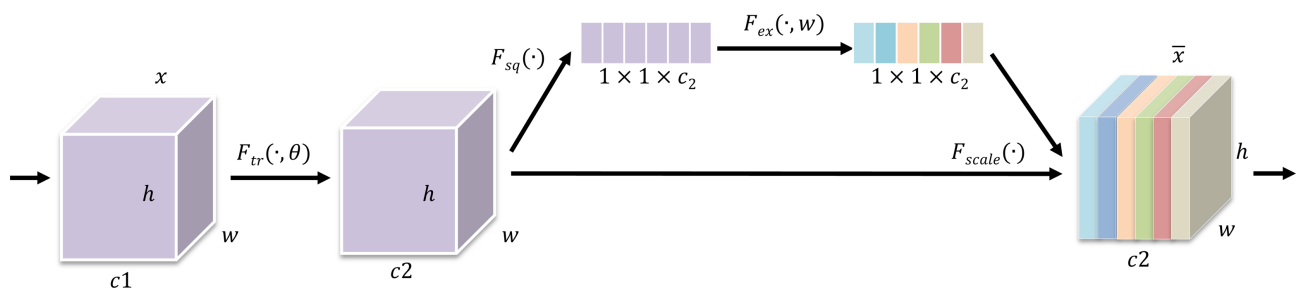The loss calculation consists of two parts: the classification loss L_{cls} between

Figure 5. SE Attention structure diagram.

the ship target prediction and the real target, and the regression loss L_{box} calculation of the ship target detection frame. The F_S loss function in this paper is composed of cross entropy classification loss and Smooth L1 regression loss, as shown in formula 5.

$$F\_S = L_{Focal} + L_{Smooth\_L1} \tag{5}$$

This paper uses cross entropy loss to calculate the classification loss of the model. The classification cross entropy loss function formula is shown in formula 6.

$$L_{Focal} = F_{Loss}\left(X_i, Y_i\right) = -\sum_{j=1}^{c} y_{ij} * \left(1 - p_{ij}\right)^{\gamma} * \log\left(p_{ij}\right) \tag{6}$$

γ is a parameter in the range of [0, 5], and when γ is 0, it becomes the initial CE loss function. c represents the number of categories, y_{ij} indicates whether the i sample belongs to category j, if it belongs to $y_{ij} = 1$, otherwise $y_{ij} = 0$. p_{ij} represents the probability that the i sample belongs to category j. This paper uses the SoftMax function to obtain the probability p_{ij} of samples belonging to each category.

For the regression loss of the prediction box, this paper uses the Smooth L1 loss function. The ship detection in this paper belongs to a single sample. If x is defined as the difference between the predicted value and the true value, the corresponding Smooth L1 loss function can be expressed as Formula 7.

$$L_{Smooth\_L1} = Smooth_{L1}\left(x\right) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \tag{7}$$

4. Experiment

4.1. Implementation Details

The experiments were conducted on the YOLOv7 [30] backbone network. Firstly, the initialized network was trained using the COCO format dataset. Secondly, the model was trained for 30 rounds using an SGD optimizer with a batch-size of 16. Where the initial learning rate of the backbone network is set to 0.02, the kinetic energy is 0.9, and the normalized mean value of the dataset images is [0.1559097, 0.15591368, 0.15588938] and the variance is [0.10875329, 0.10876005, 0.10869534]. All experiments in this paper were conducted on an NVIDIA GeForce RTX 3060 GPU.

4.2. Datasets

In this paper, ship detection models are trained and tested on SSDD [19] dataset and SAR-Ship [20] dataset. HRSID [21] is used as the experimental dataset for quantitative analysis of the models. In order to meet the same format of the three datasets, the image labels are uniformly set to COCO data format in this paper.

For the SAR-Ship dataset, the dataset comprises the 102-view GF-3 satellite data provided by the China Re-sources Satellite Application Center and the 108-view

Sentinel-1 satellite data provided by the ESA. The Institute of Aerospace Information Innovation, Chinese Academy of Sciences research team, provides the labeled data.

The dataset first processes the original 16-bit complex data into an 8-bit digital image by performing amplitude value generation, bit depth quantization, and grayscale stretching processing on the source data. Then, a ship slice with a pixel size of 256 × 256 is constructed by cropping and filtering. Finally, the LabelImg target annotation software generates the corresponding ship label box information text for each ship slice.

The dataset contains data obtained by SAR under different environmental conditions and background complexities, including 20,000 images. Among them, the multi-dimensional characteristic signs of ships include spectrum, shape, size, and spatial distribution. The ship appears gray-white in the remote sensing image and is similar to the color of many shore buildings. The typical shapes of small and medium-sized ships are point-shaped, I-shaped, and patch-shaped when photographing by remote sensing satellites. Ships occupy a small proportion of pixels in satellite image datasets. As shown in Figure 6, the spatial distribution of ships is sparse but denser at the wharf.

4.3. Evaluation Metric

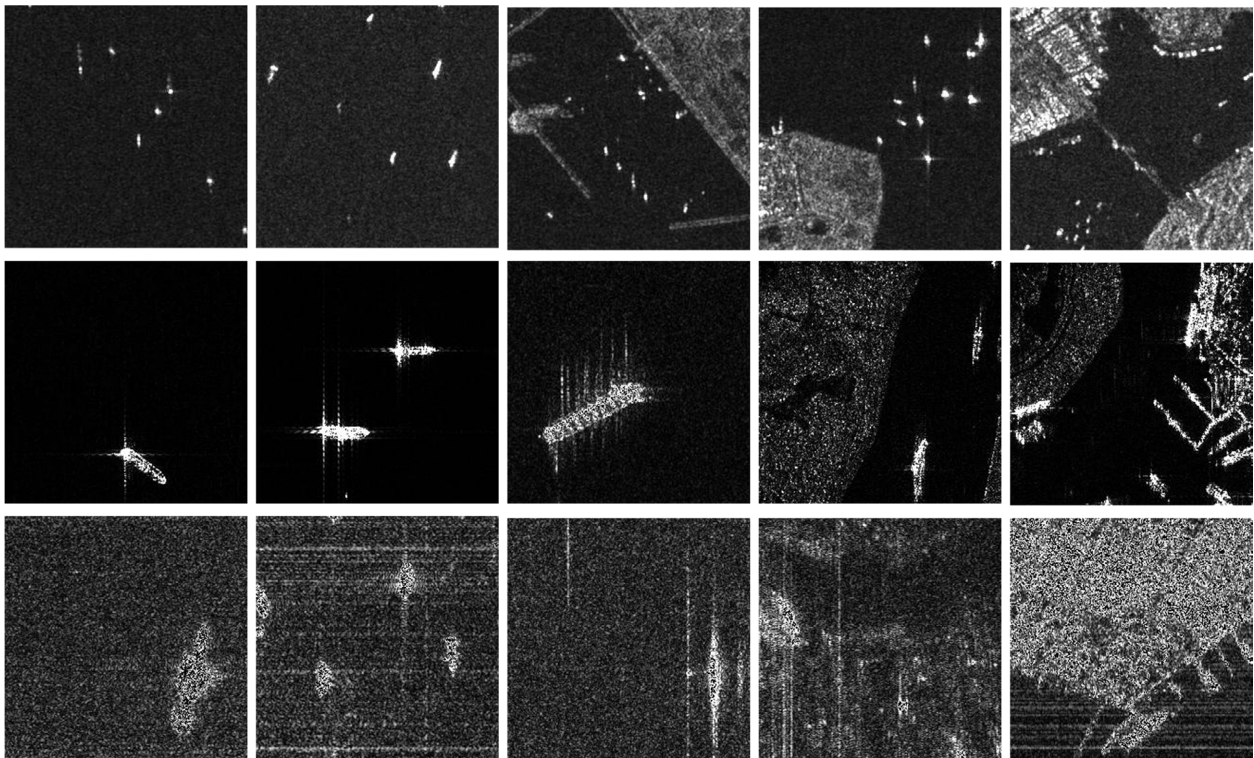In order to evaluate the algorithm's performance based on the validation dataset,

Figure 6. Typical samples of the training set. From right to left, the background complexity of the ship data gradually increases. From top to bottom, the ship data are obtained under larger, smaller, and severe environmental disturbances. It also includes data on different ship shapes (point, I, and plaque).

this paper employs the precision rate (P), the recall rate (R), and the average precision (AP) as evaluation indicators.

The basic parameters that construct the target detection evaluation index are TP (True Positive), FP (False Positive), and FN (False Negative).

TP represents the number of predicted positive targets and actually positive targets. FP represents the number of predicted positive targets but actually negative targets. FN represents the number of predicted negative targets but actually positive targets.

1) P (Precision) represents the proportion correctly identified in the prediction result of the ship, as shown in formula (8).

$$P = \frac{TP}{TP + FP} \tag{8}$$

2) R (Recall) represents the proportion correctly identified in all ground-truth marker boxes of the ship, as shown in formula (9).

$$R = \frac{TP}{TP + FN} \tag{9}$$

3) AP (Average precision) is an essential indicator for evaluating the model performance, as shown in formula (10).

$$AP = \int_0^1 P \mathrm{d}R \tag{10}$$

5. Discussion and Analysis

5.1. Comparison to State-of-the-Art

This paper first shows in Table 1 the main quantitative comparisons of SSE-Ship with SAR image ship detection methods in each detection type. Since the detection mechanisms of the major classes of target detection algorithms differ, a classification comparison is made in this paper.

It can be seen in Table 1 that SSE-Ship performs well on both SSDD and SAR-Ship datasets compared to existing algorithms. In the SSDD dataset, SSE-Ship improves 3.6% on AP_0.5:0.95 compared to CRAS YOLO [40] and 1.5% on AP_0.5:0.95 compared to CRTransSAR [38]. In addition, SSE-Ship performs better on the SAR-Ship dataset. This is attributed to two main reasons: 1) SAR-Ship is a larger dataset than SSDD, which is important for the training of the STCSPB module. 2) The SAR-Ship dataset contains information about ships in more scenarios, which effectively improves the generalization and robustness of the model.

5.2. Ablation Study

In this section, we conduct a large number of experiments to validate the effectiveness of our proposed SSE-Ship. The ablation experiments are performed using the backbone model of YOLOv7, and the results are reported on the SAR-Ship dataset.

5.2.1. STCSPB Ablation Study

STCSPB, as the core part of the context fuser, has the ability to correlate long-range contextual information. To verify its effectiveness, we use STCSPB for comparative analysis with feature fusion networks of other detection algorithms, as shown in Table 2.

As can be seen from Table 2, the STCSPB module proposed in this paper has outstanding performance. Although it is slightly lower than the fuser FPN [45] of the two-stage algorithm on mAP_0.5, there is a 2.2% gain in mAP_0.5:0.95 compared to FPN [45].

Table 1. Quantitative comparison of SSDD and SAR-Ship sets.

| Method | SSDD | | | | SAR-Ship | | | | Params (M) | Memory (Byte) |
| | mAP | | | R | mAP | | | R | | |
	P	AP_0.5:0.95	AP_0.5		P	AP_0.5:0.95	AP_0.5			
Anchor-free										
FCOS Net [33]	0.834	0.363	0.742	0.861	0.846	0.452	0.893	0.860	31.84	3002
CP FCOS [16]	0.900	-	0.911	0.858	0.889	-	0.923	0.882	-	-
Corner Net [34]	0.816	0.375	0.839	0.905	0.850	0.420	0.878	0.834	200.95	6851
LPEDet [35]	0.930	0.644	0.897	0.92	-	-	-	-	-	-
Two stage										
Faster Rcnn [36]	0.940	0.61	**0.971**	0.931	0.949	0.613	0.972	0.904	41.12	3788
Cascade Rcnn [37]	0.928	0.625	**0.971**	**0.941**	0.931	0.624	**0.973**	0.924	68.93	4104
CRTransSAR [38]	-	0.632	0.97	0.937	-	-	-	-	96	-
PVT SAR [39]	-	-	0.935	-	-	-	0.908	-	121	-
One stage										
CRAS YOLO [40]	0.931	0.611	0.987	-	0.934	0.618	0.921	-	-	-
SSD [41]	0.887	0.512	0.89	0.922	0.903	0.524	0.928	0.877	141	4123
FASC Net [42]	0.911	0.61	0.873	0.933	-	-	0.961	-	64	2505
TWC Net [43]	-	0.603	-	-	-	-	-	-	-	-
FEPS Net [44]	-	0.599	0.96	-	-	-	-	-	-	-
SSE-Ship (ours)	**0.944**	**0.647**	0.964	0.940	**0.950**	**0.656**	0.968	**0.946**	54	3275

Table 2. Fuser comparison.

| Models | Accuracy | P | R | mAP | |
				mAP_0.5	mAP_0.5:0.95
FPN [45]	98.8184	0.938	0.924	**0.971**	0.634
CARAFE-FPN [46]	98.9648	0944	0.936	0.966	0.640
PAFPN [40]	**99.1016**	0.947	0.930	0.964	0.645
STCSPB(ours)	99.0576	**0.950**	**0.946**	0.968	**0.656**

5.2.2. SE Attention

For SE Attention we use a heat map for comparison to verify the effectiveness of introducing spatial attention, as shown in Figure 7. Figure 7 uses typical large ships, docked ships and combination ships as validation images. The model with SE Attention can extract more effective characteristic information than the model without SE Attention.

5.3. Loss Function

In order to verify the effectiveness of F_S loss function on SSE-Ship model, this paper conducted three sets of comparative experiments as shown in Figure 8. It can be seen from Figure 8(a) that Focal loss has significant advantages in classifying losses, and the model converges quickly. Figure 8(b) shows that Smooth L1 loss gives full play to its advantages in model training. In Figure 8(c), CE loss and Focal loss are combined with Smooth L1 loss, respectively. The effect of combining Focal loss with Smooth L1 loss is better than the total loss of combining CE loss with Smooth L1 loss is similar.

5.4. Model Inference

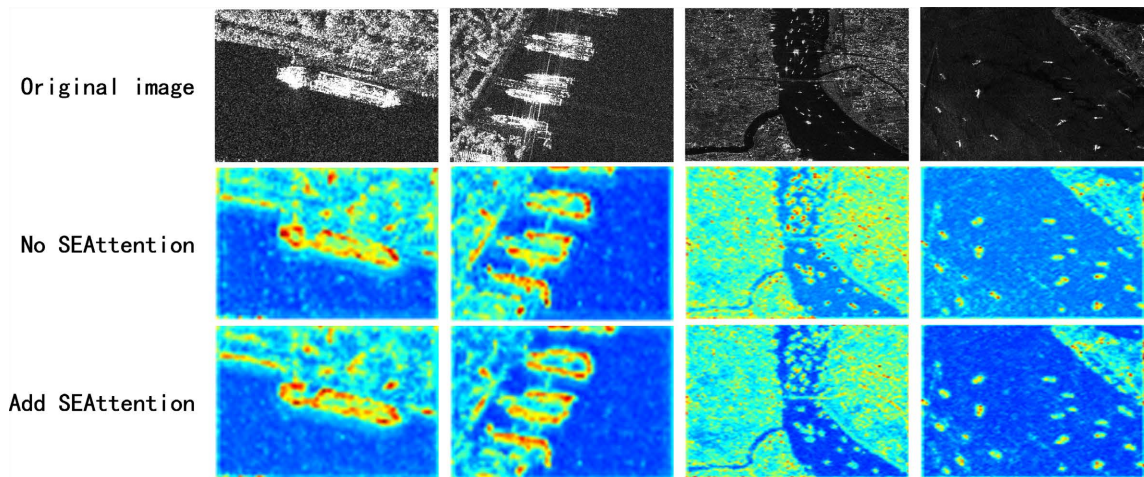In this section, the robustness and efficiency of the model are analyzed in detail

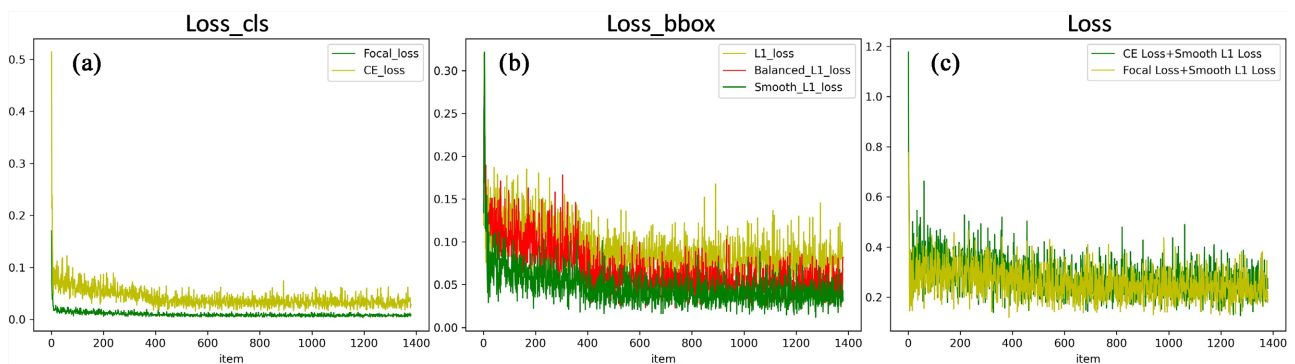**Figure 7.** Heatmap comparison.

Figure 8. Comparative experimental diagram of loss function

through the model reasoning process and results.

5.4.1. Model Robustness and Generalization

In order to illustrate the robustness of the SSE-Ship algorithm, the ship detection results of part of the SAR images are shown in Figure 9. Figure 9(a) shows the detection results under dense small targets. YOLOv7 may miss detection. As shown in Figure 9(b), YOLOv7 in SAR dataset is unsuitable for detecting large target ships. Figures 9(c)-(e) show that the YOLOv7 algorithm will have false detection under interference from nearby objects and environmental clutter. The SSE-Ship algorithm has fewer false detections and missed detections when facing interference from nearby objects, environmental clutter, dense small targets, large targets, and other ship detection situations. To sum up, the SSE-Ship has good generalization ability and robustness. Besides, the problem of the significant difference in ship size and uneven distribution of ship space is well solved.

5.4.2. Model Size and Efficiency

At present, the indicators commonly used to evaluate the size and efficiency of the model include: Memory, Parameters (Params) and Frames Per Second (FPS), etc. Memory is the size of the unit bytes that the model needs to access, which indicates the model's demand for storage unit bandwidth. Params is the sum of the parameters in the model, which is used to evaluate the size of the model volume. FPS refers to the number of pictures reasoned by the model per second, which is used to evaluate the overall efficiency of the model.

In this section, the SSE-Ship detection model is analyzed in detail using different model efficiency evaluation indicators, as shown in Table 3. CRTransSar
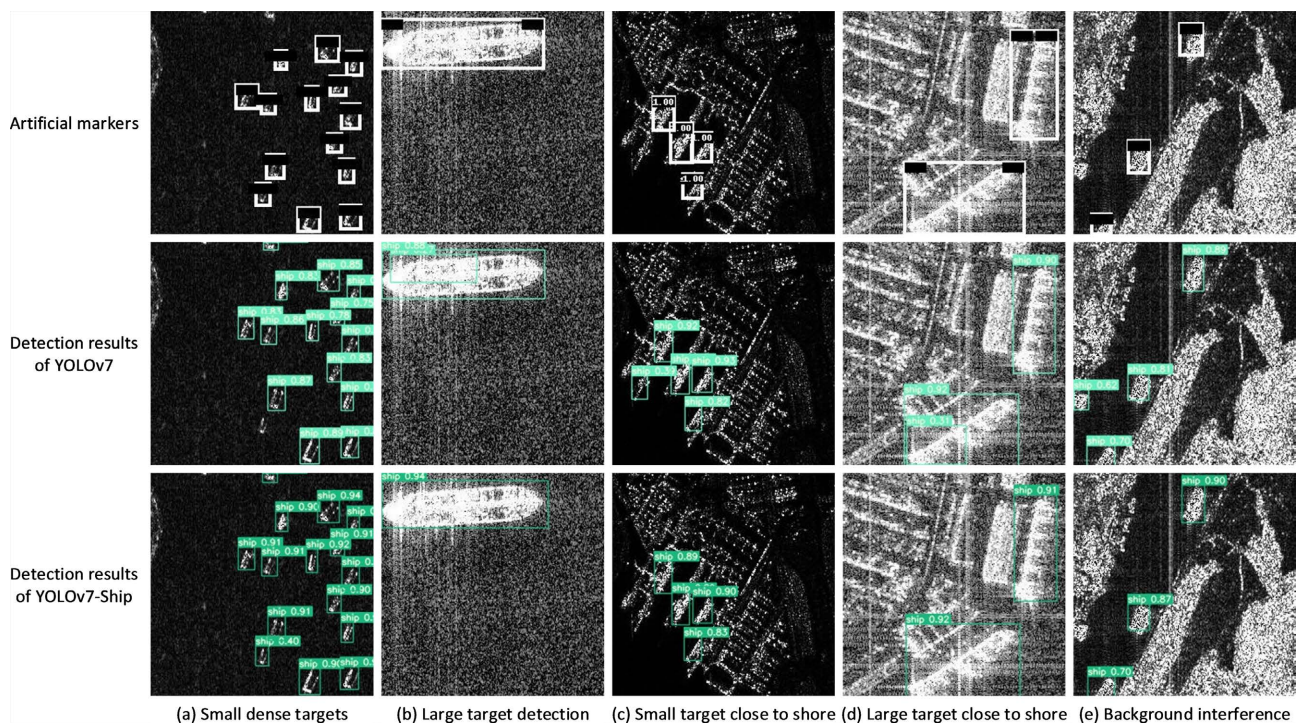

(a) Small dense targets (b) Large target detection (c) Small target close to shore (d) Large target close to shore (e) Background interference

Figure 9. Comparison chart of detection results.

Table 3. Evaluation table of model performance index.

Model Name	FPS (img/s)	Params (M)	Memory (Byte)
FCOS [33]	16.1	31.84	3002
CornerNet [34]	6.3	200.95	6851
PVT-SAR [39]	9.1	121	-
FASC-Net [42]	60	64	2505
CRTransSar [38]	7.5	96	-
Faster Rcnn [36]	14.6	41.12	3788
Cascade Rcnn [37]	10.7	68.93	4104
YOLOv5	60	45	2254
YOLOv7	58	38	2536
SSE-Ship (ours)	50	62.65	2667

and PVT-SAR, as improved two-stage detection algorithms, have improved the overall detection accuracy (Table 1), but their model parameter quantity has significantly increased and the model reasoning speed has also significantly decreased. As a first-stage algorithm with significant advantages of fast detection, FASC-Net has outstanding performance in small target detection, but the accuracy of large target detection is low (Table 1). In the performance comparison of each target detection model, the FPS index shows that the reasoning speed of the model is well controlled, and Params shows that the size of the model is slightly higher, and Memory shows that the bandwidth requirement of the model for the storage unit is normal.

To sum up, the indicators of SSE-Ship compared with other detection algorithms are within the feasible range. SSE-Ship increases the detection accuracy of medium and large objects while maintaining the detection efficiency of small objects, and there is no large consumption in the inference performance and model size of the model.

5.5. Quantitative Analysis

To evaluate the performance of the SSE-Ship model in practical applications, this paper uses HRSID as a quantitative data set. The dataset contains the ship characteristics in multiple scenes such as real clouds, rain, building interference and different SAR shooting scales. All detection models use the pre-training weight of SAR-Ship dataset to train the HRSID dataset. The quantitative results are shown in Table 4. The SSE-Ship model in this paper has significant advantages over other methods.

6. Conclusions

The method proposed in this paper is used to detect and locate ships at sea. As a ship-centered detection task, it can be related to the realistic maritime livelihood

Table 4. Quantitative analysis result.

Model Name	P	R	mAP_0.5:0.95	mAP_0.5	FPS (img/s)
CRAS-YOLO [40]	-	-	0.618	0.921	32
FASC-Net [42]	0.960	-	-	0.961	60.4
Faster Rcnn [36]	0.954	0.931	0.597	0.942	18
Cascade Rcnn [37]	0.955	0.942	0.613	0.942	13
PVT-SAR [39]	-	0.946	-	0.908	-
SSE-Ship (ours)	0.963	0.944	0.667	0.978	52

safety and ship monitoring in the no navigation zone.

Meanwhile, it is reasonable to use SSE-Ship model to detect ships at sea in SAR images. Because in practical applications, the size and noise of ships can vary greatly due to the different distances and environments of SAR shots, the use of SSE-Ship can not only identify complex types of ships according to the context, but also effectively detect multi-scale ships, *i.e.*, ensure high detection rate of small targets and improve the detection accuracy of large targets.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zhao, Y., Zhao, L., Xiong, B. and Kuang, G. (2020) Attention Receptive Pyramid Network for Ship Detection in SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**, 2738-2756. https://doi.org/10.1109/JSTARS.2020.2997081

[2] Zhang, T.W., Zhang, X., Shi, J. and Wei, S. (2019) Depthwise Separable Convolution Neural Network for High-Speed SAR Ship Detection. *Remote Sensing*, **11**, Article 2483. https://doi.org/10.3390/rs11212483

[3] Wang, Y.Y., Wang, C., Zhang, H., Dong, Y. and Wei, S. (2019) Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery. *Remote Sensing*, **11**, Article 531. https://doi.org/10.3390/rs11050531

[4] Li, X.F., *et al.* (2020) Deep-Learning-Based Information Mining from Ocean Remote-Sensing Imagery. *National Science Review*, **7**, 1584-1605. https://doi.org/10.1093/nsr/nwaa047

[5] Chang, Y.L., *et al.* (2019) Ship Detection Based on YOLOv2 for SAR Imagery. *Remote Sensing*, **11**, Article 786. https://doi.org/10.3390/rs11070786

[6] Zhang, X.H., *et al.* (2019) A Lightweight Feature Optimizing Network for Ship Detection in SAR Image. *IEEE Access*, **7**, 141662-141678.

---

https://doi.org/10.1109/ACCESS.2019.2943241

[7] Xu, X.W., Zhang, X.L. and Zhang, T.W. (2022) Lite-YOLOv5: A Lightweight Deep Learning Detector for On-Board Ship Detection in Large-Scene Sentinel-1 SAR Images. *Remote Sensing*, **14**, Article 1018. https://doi.org/10.3390/rs14041018

[8] Xu, P., *et al.* (2021) On-Board Real-Time Ship Detection in HISEA-1 SAR Images Based on CFAR and Lightweight Deep Learning. *Remote Sensing*, **13**, Article 1995. https://doi.org/10.3390/rs13101995

[9] Tang, G., *et al.* (2021) N-YOLO: A SAR Ship Detection Using Noise-Classifying and Complete-Target Extraction. *Remote Sensing*, **13**, Article 871. https://doi.org/10.3390/rs13050871

[10] Sun, Z.Z., Leng, X., Lei, Y., Xiong, B., Ji, K. and Kuang, G. (2021) BiFA-YOLO: A Novel YOLO-Based Method for Arbitrary-Oriented Ship Detection in High-Resolution SAR Images. *Remote Sensing*, **13**, Article 4209. https://doi.org/10.3390/rs13214209

[11] Jiang, J.H., *et al.* (2021) High-Speed Lightweight Ship Detection Algorithm Based on YOLO-V4 for Three-Channels RGB SAR Image. *Remote Sensing*, **13**, Article 1909. https://doi.org/10.3390/rs13101909

[12] Hong, Z.H., *et al.* (2021) Multi-Scale Ship Detection from SAR and Optical Imagery via a more Accurate YOLOv3. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**, 6083-6101. https://doi.org/10.1109/JSTARS.2021.3087555

[13] Zhang, T.W., Zhang, X.L. and Ke, X. (2021) Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection. *Remote Sensing*, **13**, Article 2771. https://doi.org/10.3390/rs13142771

[14] Zhang, S.M., Wu, R., Xu, K., Wang, J. and Sun, W. (2019) R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery. *Remote Sensing*, **11**, Article 631. https://doi.org/10.3390/rs11060631

[15] Yang, R., *et al.* (2021) 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*, **14**, 1938-1958. https://doi.org/10.1109/JSTARS.2021.3049851

[16] Sun, Z.Z., *et al.* (2021) An Anchor-Free Detection Method for Ship Targets in High-Resolution SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**, 7799-7816. https://doi.org/10.1109/JSTARS.2021.3099483

[17] Gao, F., He, Y., Wang, J., Hussain, A. and Zhou, H. (2020) Anchor-Free Convolutional Network with Dense Attention Feature Aggregation for Ship Detection in SAR Images. *Remote Sensing*, **12**, Article 2619. https://doi.org/10.3390/rs12162619

[18] Chen, C., He, C., Hu, C., Pei, H. and Jiao, L. (2019) A Deep Neural Network Based on an Attention Mechanism for SAR Ship Detection in Multiscale and Complex Scenarios. *IEEE Access*, **7**, 104848-104863. https://doi.org/10.1109/ACCESS.2019.2930939

[19] Zhang, T.W., *et al.* (2021) SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote Sensing*, **13**, Article 3690. https://doi.org/10.3390/rs13183690

[20] Wang, Y.Y., Wang, C., Zhang, H., Dong Y. and Wei, S. (2019) A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds. *Remote Sensing*, **11**, Article 765. https://doi.org/10.3390/rs11070765

[21] Wei, S.J., Zeng, X., Qu, Q., Wang, M., Su, H. and Shi, J. (2020) HRSID: A

High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Access*, **8**, 120234-120254.
https://doi.org/10.1109/ACCESS.2020.3005861

[22] Zou, L.C., Zhang, H., Wang, C., Wu, F. and Gu, F. (2020) MW-ACGAN: Generating Multiscale High-Resolution SAR Images for Ship Detection. *Sensors*, **20**, Article 6673. https://doi.org/10.3390/s20226673

[23] Zhang, T.W., *et al.* (2020) LS-SSDD-v1.0: A Deep Learning Dataset Dedicated to Small Ship Detection from Large-Scale Sentinel-1 SAR Images. *Remote Sensing*, **12**, Article 2997. https://doi.org/10.3390/rs12182997

[24] Zhang, T.W. and Zhang, X.L. (2021) ShipDeNet-20: An Only 20 Convolution Layers and < 1-MB Lightweight SAR Ship Detector. *IEEE Geoscience and Remote Sensing Letters*, **18**, 1234-1238. https://doi.org/10.1109/LGRS.2020.2993899

[25] Zhang, T.W. and Zhang, X.L. (2019) High-Speed Ship Detection in SAR Images Based on a Grid Convolutional Neural Network. *Remote Sensing*, **11**, Article 1206. https://doi.org/10.3390/rs11101206

[26] Zhang, T.W., Zhang, X.L. and Ke, X. (2021) Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection. *Remote Sensing*, **13**, Article 2771. https://doi.org/10.3390/rs13142771

[27] Wei, S.J., *et al.* (2020) Precise and Robust Ship Detection for High-Resolution SAR Imagery Based on HR-SDNet. *Remote Sensing*, **12**, Article 167. https://doi.org/10.3390/rs12010167

[28] Gao, L., *et al.* (2021) STransFuse: Fusing Swin Transformer and Convolutional Neural Network for Remote Sensing Image Semantic Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**, 10990-11003. https://doi.org/10.1109/JSTARS.2021.3119654

[29] Zhang, T.W. and Zhang, X.L. (2022) A Mask Attention Interaction and Scale Enhancement Network for SAR Ship Instance Segmentation. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5. https://doi.org/10.1109/LGRS.2022.3189961

[30] Wang, C.-Y., Bochkovskiy, A. and Liao, H.-Y.M. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022.

[31] Liu, Z., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. https://doi.org/10.1109/ICCV48922.2021.00986

[32] Hu, J., *et al.* (2017) Squeeze-and-Excitation Networks. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[33] Tian, Z., *et al.* (2019) FCOS: Fully Convolutional One-Stage Object Detection. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 9626-9635. https://doi.org/10.1109/ICCV.2019.00972

[34] Law, H. and Deng, J. (2018) CornerNet: Detecting Objects as Paired Keypoints. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *ECCV* 2018: *Computer Vision—ECCV* 2018, *Lecture Notes in Computer Science*, Vol. 11218, Springer, Cham, 765-781. https://doi.org/10.1007/978-3-030-01264-9_45

[35] Feng, Y., *et al.* (2022) A Lightweight Position-Enhanced Anchor-Free Algorithm for SAR Ship Detection. *Remote Sensing*, **14**, Article 1908.

https://doi.org/10.3390/rs14081908

[36] Ren, S., He, K., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Proceedings of the* 28*th International Conference on Neural Information Processing Systems* (*NIPS'* 15), Montreal, 7-12 December 2015, 91-99.

[37] Cai, Z.W. and Vasconcelos, N. (2021) Cascade R-CNN: High Quality Object Detection and Instance Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 1483-1498. https://doi.org/10.1109/TPAMI.2019.2956516

[38] Xia, R.F., *et al.* (2022) CRTransSar: A Visual Transformer Based on Contextual Joint Representation Learning for SAR Ship Detection. *Remote Sensing*, **14**, Article 1488. https://doi.org/10.3390/rs14061488

[39] Zhou, Y., *et al.* (2023) PVT-SAR: An Arbitrarily Oriented SAR Ship Detector with Pyramid Vision Transformer. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **16**, 291-305. https://doi.org/10.1109/JSTARS.2022.3221784

[40] Zhao, W.X., Syafrudin, M. and Fitriyani, N.L. (2023) CRAS-YOLO: A Novel Multi-Category Vessel Detection and Classification Model Based on YOLOv5s Algorithm. *IEEE Access*, **11**, 11463-11478. https://doi.org/10.1109/ACCESS.2023.3241630

[41] Liu, W., *et al.* (2015) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., eds., *ECCV* 2016: *Computer Vision—ECCV* 2016, *Lecture Notes in Computer Science*, Vol. 9905, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[42] Yu, J.M., *et al.* (2022) A Fast and Lightweight Detection Network for Multi-Scale SAR Ship Detection under Complex Backgrounds. *Remote Sensing*, **14**, Article 31. https://doi.org/10.3390/rs14010031

[43] Yu, L., *et al.* (2021) TWC-Net: A SAR Ship Detection Using Two-Way Convolution and Multiscale Feature Mapping. *Remote Sensing*, **13**, Article 2558. https://doi.org/10.3390/rs13132558

[44] Bai, L., *et al.* (2023) Feature Enhancement Pyramid and Shallow Feature Reconstruction Network for SAR Ship Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **16**, 1042-1056. https://doi.org/10.1109/JSTARS.2022.3230859

[45] Lin, T.-Y., *et al.* (2016) Feature Pyramid Networks for Object Detection. *Proceedings of* 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 936-944. https://doi.org/10.1109/CVPR.2017.106

[46] Wang, J., *et al.* (2019) CARAFE: Content-Aware ReAssembly of FEatures. *Proceedings of* 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October 2019-2 November 2019, 3007-3016. https://doi.org/10.1109/ICCV.2019.00310