Engineering Applications of Artificial Intelligence 123 (2023) 106444

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

SAR ship localization method with denoising and feature refinement

Cheng Zha [a], Weidong Min [a,b,c,*], Qing Han [a,b,c], Wei Li [b,c,d], Xin Xiong [b,e], Qi Wang [a,b,c], Meng Zhu [a]

[a] *School of Mathematics and Computer Science, Nanchang University, Nanchang 330031, China*
[b] *Institute of Metaverse, Nanchang University, Nanchang 330031, China*
[c] *Jiangxi Key Laboratory of Smart City, Nanchang 330031, China*
[d] *School of Software, Nanchang University, Nanchang 330047, China*
[e] *Information Department, First Affiliated Hospital of Nanchang University, Nanchang 330006, China*

ARTICLE INFO

Dataset link: https://github.com/Ezrealzc/DFR

Keywords:
Synthetic Aperture Radar
Ship detection
Denoising
Hierarchical Feature Fusion
Feature refinement

ABSTRACT

Synthetic Aperture Radar (SAR) ship detection is greatly important to marine transportation monitoring and fishery resource management. To improve the detection accuracy of small ships, an SAR ship localization method with Denoising and Feature Refinement (DFR) is proposed in this paper. It consists of three parts. The first part is the denoising module, which uses non-local mean to suppress the speckle noise of the SAR image. The second part is Hierarchical Feature Fusion (HFF) module. It can integrate more low-level features by adding skip connections. This prevents the low-level spatial position information of the fused features from being diluted by high-level semantic information, therefore it is beneficial to the detection of small ships. The third part is a center-based ship predictor with Feature Refinement (FR). The FR module is proposed to refine the features and reduce the background interference, which is conducive to locate ships more accurately. Extensive experiments are conducted. The experimental results show that after adding the denoising and FR modules, the value of $\mathrm{AP}_{0.5}$ is increased by 1.7% and 2.3%, respectively, which proves the effectiveness of these two modules. In inshore and offshore scenarios, the $\mathrm{AP}_{0.5}$ values of DFR are 0.884 and 0.966, respectively, achieving the best results. The proposed method can also be generalized to mark lesion locations in medical images and detect offshore oil production platforms.

1. Introduction

Synthetic Aperture Radar (SAR) is an active microwave imaging sensor, which has the characteristics of all-day work, all-weather imaging, and long operating distance. It can penetrate through clouds and fog, and still obtain high-resolution two-dimensional images under extremely low-visibility meteorological conditions. Unlike optical images, SAR images have longer wavelengths and more complex imaging mechanisms. It is more difficult to interpret intuitively. The most prominent thing is that the object is discontinuous in the SAR images. It is composed of multiple discrete irregular scattering center bright spots, and the key semantic information of object is hidden among these scattering centers. Due to the advantages of SAR all-weather imaging, it has been widely used in many fields (Zanchetta and Zecchetto, 2020; Elyouncha et al., 2021; Oloumi et al., 2020). The task of SAR object detection is to effectively locate the interesting objects from the SAR images with a large scene, and its essence is to separate the objects from the clutter background. Assume that there are q ships in the given image, and then the detection results of the mathematical formulation could be expressed as Eq. (1).

$$output = \left\{ \left(x^1_{center}, x^1_{size}, x^1_{offset}\right), \ldots, \left(x^i_{center}, x^i_{size}, x^i_{offset}\right), \ldots, \left(x^q_{center}, x^q_{size}, x^q_{offset}\right) \right\} \tag{1}$$

Here, x^i_{center}, x^i_{size} and x^i_{offset} represent the center point, size and offset of the ith ship, respectively. Then according to these three attributes, the position and size of the ith ship could be directly determined.

Speckle noise is a feature of SAR imaging system, which is derived from random scattering of objects in a basic resolution unit. Speckle noise manifests as small speckles related to the signal in an image and can obscure the structure of the image; thus, the grayscale resolution and clarity of the SAR image become significantly worse. As a result, the image sometimes does not reflect the scattering characteristics of the object accurately. It seriously affects the SAR image feature extraction (Sinha et al., 2020), object recognition (Li et al., 2020b) and other subsequent processing and application (Zhang et al., 2020).

With the development of artificial intelligence technology (Fortino et al., 2021), many machine learning methods are widely used in the fields of crack prediction and identification (Nguyen-Le et al., 2020; Khatir et al., 2020), damage detection (Tran-Nogc et al., 2020, 2021), automatic laser profile recognition (Wang et al., 2020), pedestrian re-recognition (Wang et al., 2021b), fire detection (Huang et al., 2022), and microexpression recognition (Yang et al., 2021). In recent years,

---

* Corresponding author at: School of Mathematics and Computer Science, Nanchang University, Nanchang 330031, China.
E-mail address: minweidong@ncu.edu.cn (W. Min).

https://doi.org/10.1016/j.engappai.2023.106444
Received 28 March 2022; Received in revised form 13 April 2023; Accepted 8 May 2023

many object detection methods based on deep learning (Serrano, 2022) have been proposed. You Only Look Once (YOLO) (Redmon et al., 2016) is one of classical object detection algorithms. The detection problem is transformed into a regression problem for the first time, and the convolutional neural network is used to directly complete the determination of the object category and the prediction of the boundary. YOLOv3 (Redmon and Farhadi, 2018) is proposed in the continuous improvement, which uses Darknet53 as the backbone. Skip connection and the binary cross-entropy function are introduced in this version to avoid gradient disappearance and achieve multi-label prediction, respectively. YOLOX is proposed by Ge et al. (2021), and it achieves state-of-the-art results across a large-scale range of models by introducing a decoupled head and the leading label assignment strategy.

Faster R-CNN (Ren et al., 2017) is proposed by combining the advantages of R-CNN (Girshick et al., 2014) and SPP-net (He et al., 2014). The method generates many candidate boxes through Region Proposal Networks (RPN) and introduces a multitask loss function, so it shows good performance in object detection. The imbalance between positive and negative samples reduces the convergence efficiency and detection accuracy of the network. Cai and Vasconcelos (2018) propose cascading R-CNN to continuously optimize the prediction results by cascading multiple detection networks. Different from ordinary cascades, each detection model of cascade R-CNN is trained on positive and negative samples determined by different Intersection over Union (IoU) (Zhu et al., 2021) thresholds; thus, the detector at each stage has sufficient training samples. In RetinaNet, which is proposed by Lin et al. (2017b), two independent Fully Convolutional Networks (FCN) (Sun and Wang, 2018) branches are used to complete classification and localization tasks. Then focalloss is also used to resolve category imbalance, thereby reducing the loss assigned to well-categorized samples. Considering the importance of the fusion between different features, BiFPN is proposed by Tan et al. (2020) in EfficientDet. It would set a weight for each feature and adjust these weights according to the network loss, enhancing the expression ability of the fused feature. Then, the compound scaling method is designed to balance recall and precision depending on the resolution of the input image, and the depth and width of the network.

To solve the Region of Interest (RoI) (Wang et al., 2021a) misalignment between the feature map and the original image, He et al. (2017) add a mask branch based on Faster R-CNN and use FCN to predict the segmentation mask in a pixel-to-pixel manner. RoI alignment (Bai et al., 2020) determines the value of each sampling point from the adjacent grid points on the feature map. Bilinear interpolation (Bc and Prakash, 2020) is used to complete the one-to-one correspondence between input pixels and output pixels. Subsequently, some segmentation detectors are proposed, such as mask scoring R-CNN (Huang et al., 2019) and cascade mask R-CNN (Chen et al., 2019). Cascade mask R-CNN is a hybrid of mask R-CNN and cascade R-CNN. It combines the excellent characteristics of the two detectors, and then it shows better performance in object detection and completes the instance segmentation tasks. A mask branch is added in each cascade structure to complete the instance segmentation task and generate the pixel level mask.

With the breakthrough of deep learning in the field of computer vision, researchers in the field of SAR image ship detection also begin to adopt these methods. However, most methods are originally proposed for optical images. Although they can be applied to SAR image ship detection, the characteristics of SAR images and ship objects are not considered, such as the uneven background scattering clutter and the interference of nearby objects. Speckle noise covers the details of the objects, which affects the information extraction ability of the model. In addition, the ship objects are relatively smaller than the ocean background, and the interference near the ship causes inaccurate positioning. Therefore, an SAR ship detection method with Denoising and Feature Refinement (DFR) is proposed. The contributions of this paper could be described as follows:

(1) Denoising module is introduced to reduce the impact of coherent speckle noise, which is beneficial for extracting high-quality features.

(2) Hierarchical Feature Fusion (HFF) module is proposed to accurately detect small ships in SAR images. This module could prevent low-level features from being diluted by high-level features with semantic information, preserving more spatial location information.

(3) Feature Refinement (FR) module based on deformable convolution is proposed to refine the features and reduce the impact brought by the background interference on ship detection.

(4) The proposed method achieves the best results in offshore and inshore scenarios, proving its effectiveness and superior performance.

The rest of this paper is organized as follows. In Section 2, we describe related work. The main idea of the proposed method is introduced in Section 3. Experimental results and analysis are shown in Section 4. Section 5 presents our conclusions.

2. Related work

Speckle noise in SAR images is an inherent defect of the SAR imaging system, and its existence would present a great obstacle to the subsequent interpretation and application of SAR images. Therefore, classical non-modeling methods, such as Median Filtering (MF) (Tang and Lin, 2008) and Average Filtering (AF) (Hosseini et al., 2015), are proposed to filter out coherent speckle noise in the early research. The main idea behind them is to use local statistical features, such as the median or average of the pixel values in the local smoothing window, to replace the pixel value of the central pixel. Although these methods can reduce coherent speckle noise to a certain extent, the statistical characteristics of speckle noise are not considered, resulting in the correlation characteristics between natural features being ignored. Then, the edge texture information and noise information cannot be accurately distinguished. As a result, the edge and detail information are blurred while reducing noise. To solve this problem, a filtering method based on Non-Local Mean (NLM) (Deledalle et al., 2014) is proposed. The main idea is to use the average value of similar blocks to replace the pixel value of the target block. In this way, the structural information of the image is completely used, and the edge and texture information are better preserved.

The spatial domain filtering method (Katayama et al., 2000) is based on statistics and estimation theory. The approximate relationship between real images and the corresponding pixels of the noisy images could be estimated by assuming that the coherent speckle noise conforms to a certain statistical model. The most classic method is the adaptive filter based on image statistical characteristics, which is proposed by Lee (2009). The basic idea is to take advantage of the local statistical characteristics of the image to control the output of the filter, thereby adapting to the image transformation. The idea of Lee filtering is borrowed and developed by later researchers, and many spatial adaptive filters based on local statistical characteristics are proposed, such as Kuan filtering (Klaine et al., 2005) and Frost filtering (Pan et al., 2020).

The Block Matching 3D (BM3D) combines the denoising methods in the spatial domain and transform domain to better preserve some details in images, which can improve the image denoising effect. Under the same size of similar blocks and search areas, the algorithm complexity of BM3D is about three times that of NLM. Denoising methods based on Principal Component Analysis (PCA) overcome the shortcomings brought by standard orthogonal transformation and can generate good results for image details with high-structure. But the existence of noise would inevitably affect the accuracy of principal component analysis. The K-Singular Value Decomposition (K-SVD) solves the problem that the fixed transformation matrix basis cannot adapt to the image texture information, but it requires a large amount of calculation to update the dictionary. The Discrete Cosine Transform (DCT) converts a two-dimensional image from the spatial domain to the frequency domain, but the sparse representation obtained by two-dimensional transformation is only suitable for specific image modes, and it has low robustness.

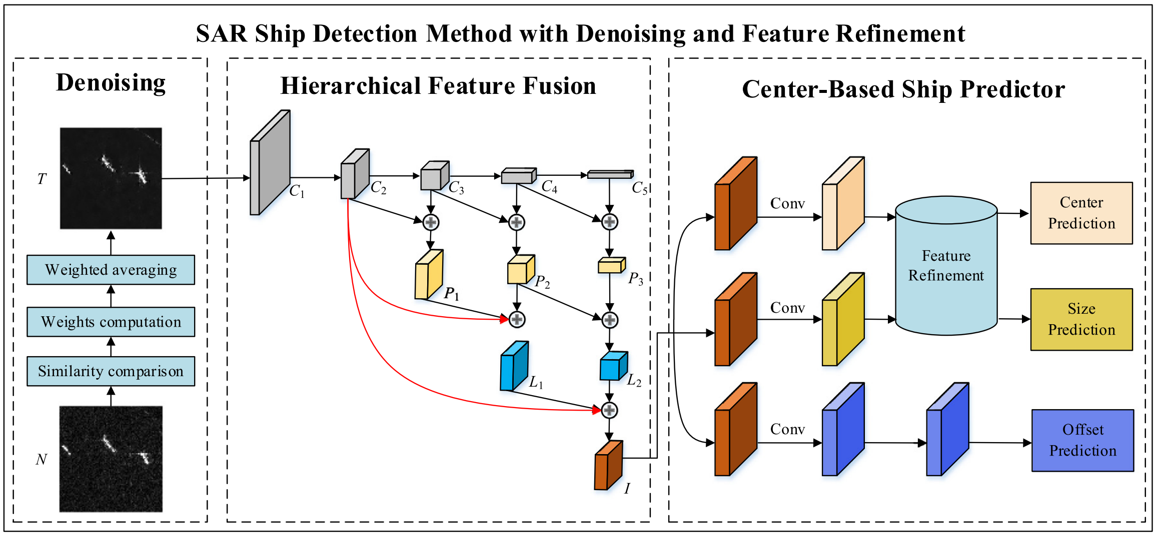


Fig. 1. Overall architecture of DFR.

Weighted Nuclear Norm Minimization (WNNM) is a low rank algorithm for constraining the singular value of a sparse matrix, increasing the weight of the nuclear norm to improve the flexibility of the nuclear norm. However, it has low accuracy in the estimation of local image block iteration parameters and needs a lot of time for block matching.

In addition, there are many noise reduction methods using self-attention deep neural networks. A novel dual-attention denoising network is proposed by Shi et al. (2021). In this method, two parallel branches with attention modules, the spatial branch and the spectral branch, are designed to process the spatial and spectral information separately. Bera and Biswas (2021) propose a new method for Low Dose CT (LDCT) denoising. It consists of three parts, a novel convolutional module, a new noise aware mean square error loss for LDCT denoising and a novel discriminator with self-attention and pixel-wise GANs for CT denoising tasks. A novel 3D Self-Attention Convolutional Neural Network (SACNN) is proposed by Li et al. (2020a). This method combines 3D self-attention and a novel self-supervised learning scheme to deal with the LDCT denoising problem.

Due to the existence of coherent speckle noise in SAR ship images, the images cannot accurately reflect the scattering characteristics of the ground object and the fine structure of the images would be hidden. This is not beneficial to extract fine features from the object. To solve this problem, we compare the noise reduction effects of various algorithms, and a method suitable for SAR ship image denoising would be selected to suppress the noise in the image, which is beneficial to the subsequent feature extraction.

The low detection accuracy of small ships is always regarded as a problem worthy of attention in SAR image ship detection. Low-level features have a higher resolution and contain more location and detail information. However, they contain less semantic information due to less convolutional layer processing. High-level features have stronger semantic information, but low resolution and poor perception of detail. This inconsistency at different levels causes the detector to focus more on salient objects with sufficient semantic information and ignore small objects. HRSDNet (Wei et al., 2020) uses the high-resolution feature pyramid network (Lin et al., 2017a) as the backbone network to obtain accurate spatial accuracy by connecting features from high resolution to low resolution in parallel. This finding indicates that the pyramid feature representation is a solution for small object detection problem. Inspired by this method, we designed a hierarchical feature fusion module, which can fuse more low-level features into output feature, preventing low-level features from being diluted by high-level features. A novel detector based on CNN is proposed by Dai et al. (2020), in which the features generated by the ROI pooling layer are merged. The constructed CNN framework combined with the feature representation strategy can obviously enhance the position and semantic information of the multi-scale ships. An effective and stable detector in a refined manner is proposed by Guo et al. (2021), and a feature pyramids fusion module is designed to generate more powerful semantic information. To address the problems of arbitrary orientation, large aspect ratio and dense arrangement in SAR ship detection task, Arbitrary-Oriented ship detection for SAR images based on Feature Separation and Feature Alignment (FSFADet) is proposed by Zhu et al. (2022). They establish a new SAR ship rotation detection dataset, and introduce a refined network and a feature alignment module to boost the SAR ship detection performance. The difference between FSFADet and our method is that many rotated anchors with different sizes should be preset to detect different features of five layers respectively in FSFADet, while anchors are unnecessary and only the fused features are detected in our method. For the convenience of reading and understanding, the above literatures about object detectionhave been summarized in Table 1.

One of the reasons for inaccurate location of some ships is background interference. The ships in SAR images often show different scattered power distribution from the background. However, some local areas of the ships have a similar scattering mechanism to some parts of the surrounding area. The semantic representation of the local area near the ship boundary is susceptible to interference from adjacent areas with similar scattered power. In this case, the features corresponding to the low semantic information are not well aligned with the object, further leading to the inaccurate boundary box. Some predicted boxes are unable to accurately cover the object due to this situation. To solve this problem, a center-based ship predictor with FR is proposed in this paper.

3. Method

The overall architecture of DFR includes three parts, as shown in Fig. 1. The first part is a Denoising module, which uses NLM to suppress speckle noise in SAR images, thereby obtaining a low-noise image. The second part is the Hierarchical Feature Fusion module, which takes advantage of convolutional neural network to extract image features and features with different scales are fused to obtain a fused feature. The third part is the center-based ship predictor with FR module. Through three prediction branches, the center point, width and height of the ship, and the coordinate offset of the center point are predicted, respectively. The results of the three branches are combined to obtain the final ship detection result.

Table 1
List of literature review about object detection.

Methods	Innovation
YOLO	Transforms the detection problem into a regression problem for the first time
YOLOv3	Shortcut connection and binary cross-entropy function are introduced to avoid gradient disappearance and achieve multi-label prediction, respectively.
YOLOX	Introduce a decoupled head and the leading label assignment strategy
R-CNN	Fine-tune is performed after pre-training
SPP-net	The entire image can be directly input without cropping by using the spatial pyramid pooling
Faster R-CNN	Design RPN network to replace selective search method and introduce multi-task loss function
Cascade R-CNN	Train multiple cascaded detectors with different IoU thresholds
RetinaNet	Focal loss is used to resolve category imbalance, thereby reducing the loss assigned to well-categorized samples.
EfficientDet	BiFPN is proposed to fuse different features. Compound scaling is designed to balance recall and precision.
Mask R-CNN	The RoIAlign is proposed to make the calculation of the spatial position of RoI more accurate
Mask Scoring R-CNN	Introduce Score strategy to refine prediction results
Cascade Mask R-CNN	A segmentation head is added to each stage to maximize the diversity of samples
HRSDNet	Connect multiple sub-networks from high resolution to low resolution in parallel, using soft non-maximum suppression to improve performance
CenterNet++	A feature pyramids fusion module is designed to generating more powerful semantic information
Dai et al	Combine with the feature representation strategy enhances the position and semantics information for the multi-scale ships
FSFADet	Achieve arbitrary-oriented ship detection and introducing the IoU-smooth L1 loss to address the boundary problem

3.1. SAR image denoising

The process of acquiring and transmitting SAR images is affected by various noises, which blur the edge texture of the image. Therefore, noise reduction pre-processing is necessary to obtain clear and high-quality SAR ship images to ensure that high-quality features can be extracted. The denoising algorithm based on NLM has become a research hotspot of SAR image denoising and achieves good denoising effect. NLM makes full use of the redundant information of the image; it can maintain the details of the image to the greatest extent whilst reducing the noise. The basic idea is that the estimated value of the current pixel is obtained by calculating the weighted average of pixels in the image with a similar neighborhood structure. Therefore, to measure the similarity between pixels in entire image, two fixed-size windows are set, namely, the search window ($D \times D$) and the neighborhood window ($d \times d$). It is convenient to calculate the similarity between each pixel and other pixels in the image. The neighborhood window slides in the search window, and the weight of pixel would be determined based on the similarity among neighborhoods.

The noisy SAR ship image is assumed to be $N = \{N(x) \mid x \in C\}$, where C represents the coordinate domain of the image. The denoised image is T, and then the value of pixel point y in T can be calculated by Eq. (2).

$$T(y) = \sum_{x \in C} N(x) * s(x, y) \tag{2}$$

where $s(x, y)$ represents the similarity between the pixel point x and the pixel point y, and its value is determined by the similarity between the image blocks $V(x)$ and $V(y)$. $V(x)$ and $V(y)$ refer to the image blocks which are centered on x and y, respectively. Euclidean distance is one of the most commonly used methods to measure the similarity of two image blocks, thus we adopt this method to calculate the similarity here. $s(x, y)$ can be expressed by Eq. (3).

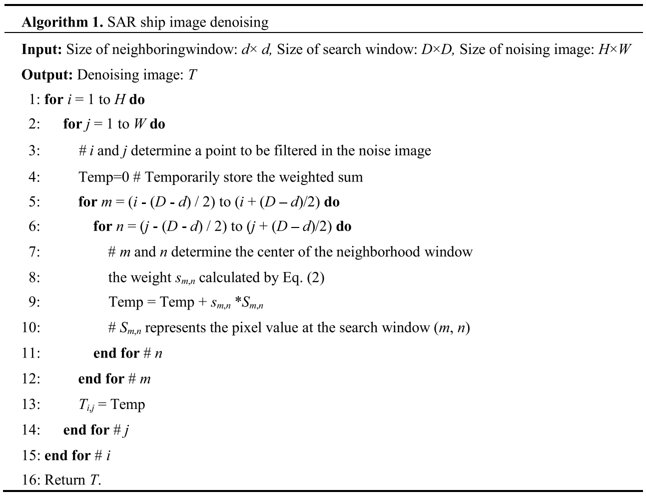
$$s(x, y) = \frac{1}{Z(x)} exp(-\frac{\|V(x) - V(y)\|^2}{h^2}) \tag{3}$$

where $Z(x)$ is the normalized coefficient, which can be expressed by Eq. (4). h is the smoothing parameter. The larger the h, the better the denoising effect will be, but at the same time the image will be more blurred. The smaller the h, the more edge details can be preserved, but the more noise points would be also remained.

$$Z(x) = \sum_{y} exp(-\frac{\|V(x) - V(y)\|^2}{h^2}) \tag{4}$$

Algorithm 1 explains the denoising process of SAR ship image based on NLM in details. Steps 1–2 determine the position of pixel points in the images. The Temp in step 4 is used to temporarily store the weighted sum of the center pixels of the neighborhood window in the search window. Steps 5–6 determine the position of the center pixel of the neighborhood window. Steps 8–9 calculate the weighted sum as the pixel value of the denoising image, and then continuously calculate the weighted sum of each position of the image to finally obtain the denoised image T.

The structural information of the image can be preserved more completely by measuring the differences between pixels according to the structural similarity between neighborhoods. At the same time, the weighted average of similar blocks is used to estimate the real images, and the noise can be effectively removed, thereby providing a guarantee for the extraction of features with more detailed information.



Algorithm 1. SAR ship image denoising

```
Input: Size of neighboringwindow: $d× d$, Size of search window: $D×D$, Size of noising image: $H×W$
Output: Denoising image: $T$
1: for $i = 1$ to $H$ do
2:     for $j = 1$ to $W$ do
3:         # $i$ and $j$ determine a point to be filtered in the noise image
4:         Temp=0 # Temporarily store the weighted sum
5:         for $m = (i - (D - d) / 2)$ to $(i + (D – d)/2)$ do
6:             for $n = (j - (D - d) / 2)$ to $(j + (D – d)/2)$ do
7:                 # $m$ and $n$ determine the center of the neighborhood window
8:                 the weight $s_{m,n}$ calculated by Eq. (2)
9:                 Temp = Temp + $s_{m,n}$ *$S_{m,n}$
10:                # $S_{m,n}$ represents the pixel value at the search window $(m, n)$
11:            end for # $n$
12:        end for # $m$
13:        $T_{i,j}$ = Temp
14:    end for # $j$
15: end for # $i$
16: Return $T$.
```

3.2. Hierarchical feature fusion

SAR ships have multi-scale features, most of which are small in size. Therefore, obtaining high-resolution features with multi-scale information is very important for SAR ship detection. The low-level features have the characteristics of high resolution and rich spatial locations, whereas the deep level features have less detailed information. However, the latter can efficiently extract rich semantic information. An effective feature fusion process should be designed to fuse the low and deep level features to fully use the semantic information of the deep level features without losing the spatial location information of the low-level features. Thus, the fused features contain rich semantic information and location information. In the feature extraction process, as the number of down-sampling increases, the spatial location information of the object is reduced. A Hierarchical Feature Fusion

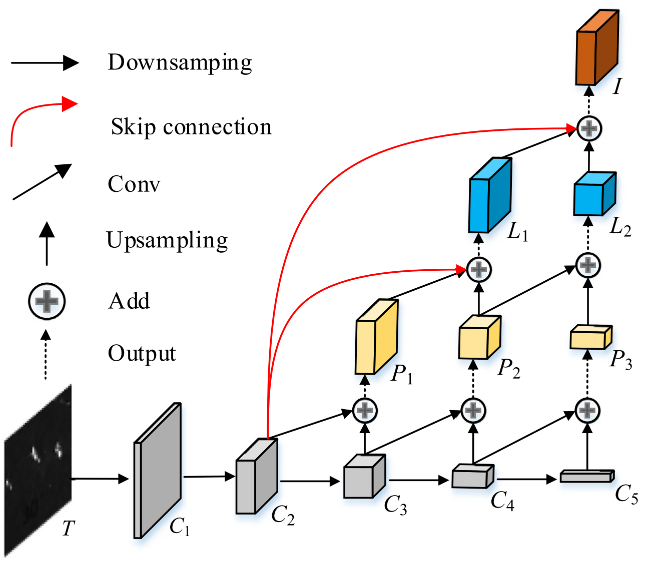


Fig. 2. The framework of hierarchical feature fusion.

module (HFF) is proposed to avoid the loss of the spatial location of the small objects during down-sampling. This module combines convolution, up-sampling, down-sampling, skip connection and fusion. Among them, down-sampling operation is similar to stage in Residual Network (ResNet). However, this module mainly focuses on feature fusion, and ResNet focuses on feature extraction. In Feature Pyramid Network (FPN), only adjacent features would be fused. But the proposed HFF module could fuse more features. Similarly, DenseNet could also fuse many features by connecting all layers. Each layer would use the outputs of all the front layers as extra input. Therefore, it requires the input and output of all blocks to have the same resolution. The proposed method could obtain features with different resolution by down- sampling and then fuse C_2 by HFF in each layer to prevent spatial location information in features from being diluted. The divergent convergent mechanism would observe the overall information of an image and then focus on the important areas. Important features would be emphasized by this mechanism. Different information between features would be fused in HFF, improving the discriminability of features.

The form of hierarchical fusion could avoid the direct aggregation of low-level features and high-level features with large differences. To alleviate the problem of gradient disappearance and over-fitting, skip connection is necessary (Huang et al., 2017). But only short connection between feature maps would easily lead to gradient disappearance problem. Thus, long skip connection is added in the module to prevent the spatial location information of the fused features from being diluted by semantic information. Long skip connection could transform features between different layers. The process of information transmission is similar to that of Long Short Term Memory (LSTM). But LSTM is a method used to extract the temporal features of objects, and the transmitted information is temporal features. After a series of operations, the final output feature I contains more low-level local detailed information and high-level global semantic information. Feature extraction and fusion consume lots of computation, which would lead to excessive model complexity. To solve this problem, we lower the model size and computation by reducing the number of channels. The framework of hierarchical feature fusion is shown in Fig. 2.

The input of hierarchical feature fusion module is the image T after denoising. Features $\{C_1, C_2, C_3, C_4, C_5\}$ are obtained by feature extractor, corresponding to features of 2, 4, 8, 16 and 32 times down-sampling, respectively. To avoid the problems of low resolution and loss of location information in features, the resolution of the final fusion feature is kept as 4 times down-sampling. And then to preserve low-level feature information, C_2 is transmitted to the feature fusion of each layer. The fusion process is divided into three layers. The first layer of the fusion process is shown in Eq. (5).

$$P = \left\{F(Conv(C_n) + U(C_{n+1}))|n = 2, 3, 4\right\} \tag{5}$$

where $Conv(\cdot)$ represents the 3×3 convolutional layer, $U(\cdot)$ represents up sampling and $F(\cdot)$ represents the fusion process. $P = \{P_1, P_2, P_3\}$ is the feature of the first level fusion. The second layer and third layer fusion process are showed in Eqs. (6) and (7).

$$L = \left\{L_1, L_2\right\} = \begin{cases} L_1 = F(C_2 + \mathrm{Conv}(P_1) + U(P_2)) \\ L_2 = F(Conv(P_2) + U(P_3)) \end{cases} \tag{6}$$

$$I = F(C_2 + Conv(L_1) + U(L_2)) \tag{7}$$

where $L = \{L_1, L_2\}$ is the fused feature of the second layer. The resolution of feature would remain unchanged after convolution, and the resolution of feature would become twice the original size after up-sampling. Assume that the size of input images is 512 × 512. The resolutions of P_1 and C_2 are 128 × 128, and the resolution of P_2 is 64 × 64. L_1 with the resolution of 128 × 128 is obtained by fusing C_2, P_1 after convolution and P_2 after up-sampling. The resolution of P_3 is 32 × 32, the resolution of P_2 is 64 × 64. L_2 with the resolution of 64 × 64 is generated by fusing P_3 after up-sampling and P_2 after convolution. Finally, output feature I with the resolution of 128 × 128 is obtained by fusing C_2, L_2 after up-sampling and L_1 after convolution. The skip connection in this module could recover the lost information during down-sampling. High level semantic information could be transmitted to the fused feature by up-sampling operation. Through hierarchical feature fusion, fused feature map I with enhanced multi-scale information can be obtained. It is beneficial to further improve the detection performance of the model.

3.3. Center-based ship predictor with feature refinement

The interference that is close to the ship boundary and has similar scattering power to the ship may affect the localization accuracy. A center-based ship predictor with FR is proposed to solve this problem. At the same time, anchor-free ship detection can be achieved by predicting the location of the center point and the size of the ship. The structure of the center-based ship predictor with FR is shown in Fig. 3. It consists of three branches, namely, the center prediction, the size prediction and the offset prediction. The first branch could achieve the higher accuracy to predict the center position of ship by introducing the FR module. The second branch uses information from the first branch to better predict the size of the predicted box. In this branch, a box more consistent with ground truth could be predicted by combining with the feature of the center point. According to the results of the first and the second branches, a predicted box could be obtained. The third branch is used to predict the offset of the center position during sampling. When the output feature is mapped to input images, the offset is used to finely adjust predicted box. Thus, only the second branch uses the information from the first branch.

The output of the center prediction branch expresses the probability that each pixel is the center of ships, which is between [0, 1]. Therefore, the output of this branch should be normalized by the sigmoid function. It is defined in Eq. (8).

$$sigmoid(x) = \frac{1}{1 + exp(-x)} \tag{8}$$

We propose a FR module based on deformable convolution (Dai et al., 2017) to improve the localization accuracy. The output of the FR module is a refined feature map G that can be used for center prediction. Each location l_0 on the output feature map G can be expressed by Eq. (9).

$$G(l_0) = \sum_{l_i \in R} w(l_i) \cdot X(l_0 + l_i + \Delta l_i) \tag{9}$$

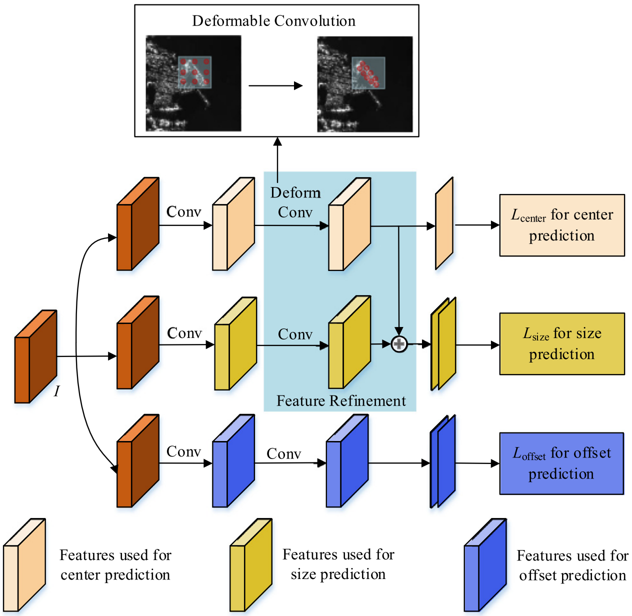


Fig. 3. The structure of center-based ship predictor.

where X is the input feature map, R is a regular grid for sampling over the features X, l_i enumerates the locations in R, Δl_i $(i = 1, 2, \ldots, |R|)$ is kernel offset. The refined feature map G is the weighted sum of the sampling values. Assuming that the corresponding coordinate of the ship center point in the denoising image T is (x_t, y_t), when the spatial resolution of the original image to heat map is reduced by z times, its coordinates on the heat map would be $(x_t/z, y_t/z)$. Then, when the center point of the ship is mapped from the heat map to the original image, the offsets in the horizontal and vertical directions are $o_x = x_t - \lfloor x_t/z \rfloor$ and $o_y = y_t - \lfloor y_t/z \rfloor$.

The output size of the size prediction branch is W/z × H/z × 2, the two channels correspond to the width and height of the predicted box. The output size of the offset prediction branch is also W/z × H/z × 2, and the two channels correspond to the horizontal and vertical offset values of the ship center mapped from the predicted heat map to the original image. During model training, the logistic regression loss function with penalty terms is used as the loss function of the center prediction branch, as shown in Eq. (10).

$$L_{center} = -\frac{1}{M}\sum_{xy}\begin{cases}(1-\hat{\gamma}_{xy})^{\alpha}\log(\hat{\gamma}_{xy}) & \gamma_{xy} = 1\\ (1-\gamma_{xy})^{\beta}(\hat{\gamma}_{xy})^{\alpha}\log(1-\hat{\gamma}_{xy}) & otherwise\end{cases} \tag{10}$$

where M refers to the number of ships in the image T. α and β are the penalty factors. The γ_{xy} represents the value of (x, y) positions in the predicted heat map.

The loss functions of the size prediction and offset prediction branches use the L1 distance loss function, which can be represented by L_{size} and L_{offset}. The calculation equations can be expressed as Eqs. (11) and (12).

$$L_{size} = \frac{1}{M}\sum_{i=1}^{M}\left(\left|\hat{w}_i - w_i\right| + \left|\hat{h}_i - h_i\right|\right) \tag{11}$$

$$L_{offset} = \frac{1}{M}\sum\left(\left|\hat{o}_x - o_x\right| + \left|\hat{o}_y - o_y\right|\right) \tag{12}$$

where w and h correspond to the width and height of the predicted box. o_x and o_y represent the horizontal and vertical predicted offset values of the center point mapped from the predicted heat map to the original map. The overall loss function of our model is the weighted sum of three loss functions, and it can be shown as Eq. (13).

$$L = L_{center} + \lambda_{size} * L_{size} + L_{offset} \tag{13}$$

where λ_{size} is the weight of L_{size}.

Table 2
Information about the SSDD.

Sensors	Polarization	Resolution	Senses
RadarSat-2 TerraSAR-X Sentinel-1	HH, VV, HV,VH	1 m–15 m	Inshore, offshore

Table 3
The hyper-parameter settings.

Hyper-parameter	Value
Learning rate	0.0005
Number of training epochs	120
Batch size	8
Optimizer	Adaptive moment estimation

4. Experimental results and evaluation

In this section, the performance of our method is evaluated using the open-source dataset named SSDD. Firstly, the SSDD dataset and the experimental environments are introduced. Then, the evaluation metrics are described. Finally, we compare the proposed method with other methods to verify the superiority of the proposed method.

4.1. Experimental data and environment

The experimental dataset is SSDD (Li et al., 2017), and its detailed information is introduced in Table 2. The original SAR images of SSDD are mainly derived from RadarSat −2, TerraSAR-X and Sentinel-1 sensors. The polarization modes of these sensors include HH, VV, HV and VH. In addition, the experimental data are divided into inshore and offshore scenes according to the complexity of the background.

The SSDD consists of 1160 SAR images; 928 SAR images are used as a training set, and 232 SAR images are used as a test set. The bounding box size distribution of each ship in the training set and the test set is shown in Fig. 4.

The keras framework is employed to implement our method. The packages used in the development of DFR include tensorflow, numpy, OpenCV2, PIL and keras, and DFR is carried out using the python language. The integrated development environment is PyCharm. All experiments are run on the personal computer with the 64-bits windows operating system. The processors include Intel(R) Xeon(R) Silver 4210 CPU @ 2.20 GHz, NVIDIA Quadro RTX 5000 and 32 GB RAM. The adaptive moment estimation with the initial learning rate of 0.0005 is adopted as an optimizer to minimize the loss function, and λ_{size} is set to 0.1. The hyper-parameter settings are shown in Table 3.

4.2. Evaluation metrics

To select a denoising method which is suitable for SAR images, the Equivalent Number of Looks (ENL) (Anfinsen et al., 2009) is used to evaluate the noise suppression performance of different methods. ENL is a common indicator for the objective quality evaluation of SAR images, which can be used to measure the degree of noise suppression. It is expressed by Eq. (14).

$$ENL = \frac{\mu^2}{\sigma} \tag{14}$$

where μ denotes the average value and σ denotes the variance value. The larger the ENL, the weaker the noise and the smoother the image would be.

Average Precision (AP) is adopted to evaluate the detection performance to quantitatively and comprehensively evaluate the effectiveness of our method. AP is the area under the precision–recall curve; thus, precision and recall should be calculated initially. Precision indicates the ratio of the ships detected accurately in all detected ships, and recall is the ratio of the ships detected in all ground truths. They are the

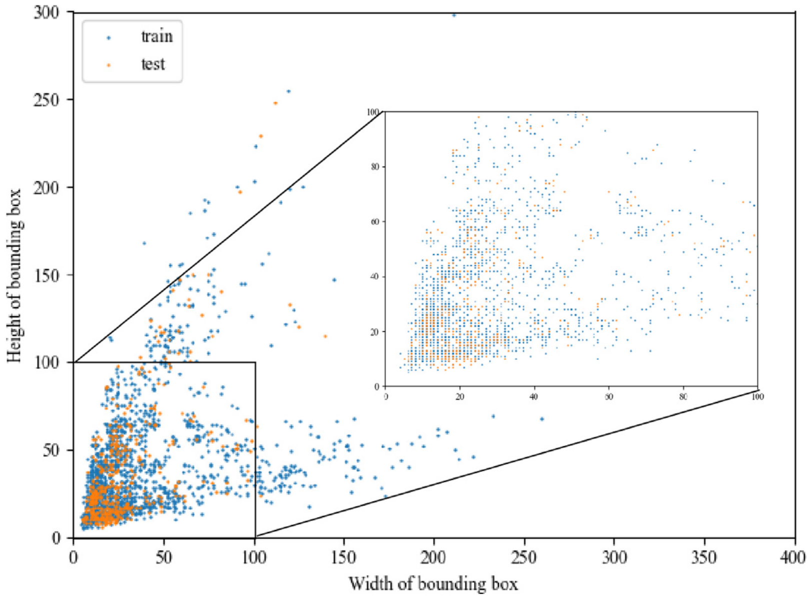


Fig. 4. Bounding box sizes distribution of SSDD.

commonly used evaluation metrics, which can be defined as Eqs. (15) and (16).

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{15}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{16}$$

where True Positive (TP) is the number of correctly detected ships, False Positive (FP) represents the number of false alarms and False Negative (FN) is the number of missed positive ships. It is difficult to accurately evaluate the performance of an object detection model by using precision or recall alone. Thus, F1 is introduced to balance precision and recall to evaluate the performance of models better. It can be expressed as Eq. (17).

$$F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{17}$$

IoU is a commonly used performance evaluation index in the field of object detection, which is the ratio of the intersection and union of the predicted box and the ground truth. It can be defined as Eq. (18).

$$IoU = \frac{B_p \cap B_{gt}}{B_p \cup B_{gt}} \tag{18}$$

where B_p represents predicted box and B_{gt} represents ground truth. The larger the IoU value, the more consistent the predicted box and the ground truth are.

In the Cartesian coordinate system, if recall is set to the horizontal coordinate and precision is set to the vertical coordinate, then the AP value is the area under the precision–recall curve, which can be calculated by Eq. (19).

$$AP = \int_0^1 P(R)dR \tag{19}$$

where P refers to precision and R denotes recall. $\mathrm{AP}_{0.5}$ refers to the AP where the IoU threshold is 0.5. That is, only predicted boxes with IoU greater than 0.5 are regarded as correct detection results.

4.3. Comparison of denoising performance

The existence of noise in SAR images would cover up the detailed information of the image and affect the quality of feature extraction. To deal with the problem, a comparison experiment of the denoising effect of various algorithms is designed to select a suitable SAR image denoising method to suppress the noise. Here, ENL is used to evaluate the denoising ability of different methods. By comparing the denoising effects of MF, AF, Frost, Lee and NLM algorithms, the best method would be selected to suppress the noise in the image, which is conducive to extracting high-quality features.

Table 4
The ENL of different methods.

	MF	AF	Frost	Lee	NLM
ENL	13.24	15.22	9.91	9.31	**18.36**

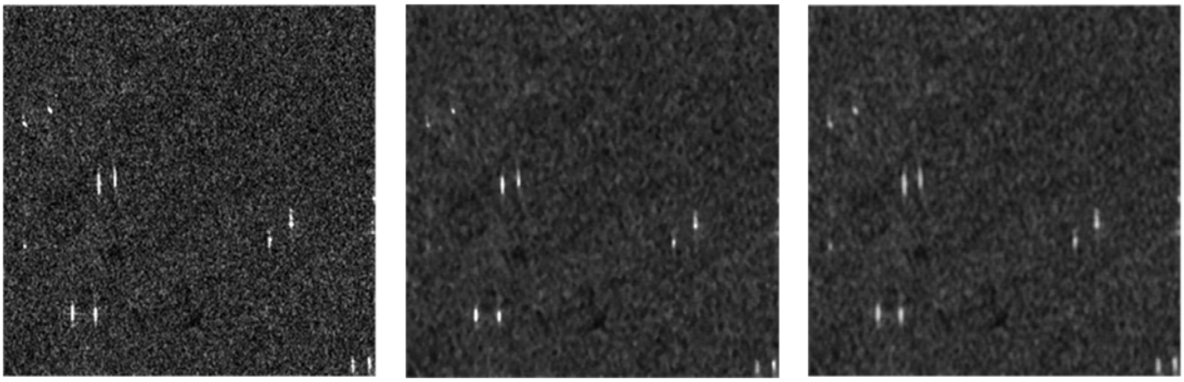
The comparison results are shown in Table 4. The ENL of AF and NLM are higher than 15, whereas the ENL of the other methods are less than 15, indicating that the denoising effect of AF and NLM is better than the other methods. And then the ENL of NLM is higher than the ENL of AF, indicating that NLM has the best denoising effect. From the table, it is clear that the ENL of NLM is 3.14 higher than the ENL of AF, and the difference in denoising effect is noticeable in the images, as shown in Fig. 5(c) and (f). When evaluating the quality of a denoising method, the numerical value is not the only criterion, but the actual effect in images should also be considered.

Fig. 5 shows that MF and AF can reduce image noise and obtain good noise smoothing capabilities, but the overall image is blurred. Frost and Lee can also reduce image noise and maintain image clarity, but the texture information of ship is seriously lost. NLM can maintain the sharpness of the image, as well as the detailed information of the ship, which would be conducive to subsequent feature extraction. Therefore, our denoising module is designed based on NLM to obtain low-noise images.

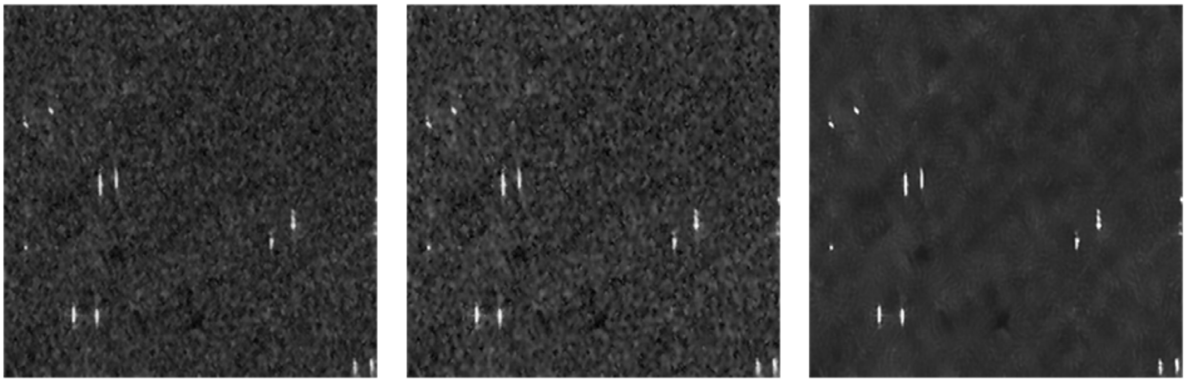
4.4. Effect of denoising and feature refinement

A denoising module and a FR module are designed to improve the accuracy of ship detection in SAR images in this paper. To prove the effectiveness of these modules, they are added in sequence to analyze the performance of model. In addition, all experiments are performed under the same settings for fair comparison.

The ablation experiment results of the denoising module and FR module are shown in Table 5. The corresponding PR curve is shown in Fig. 6. From the table, we can find that F1 and $\mathrm{AP}_{0.5}$ without the denoising module and the FR module are 0.811 and 0.911, respectively. When denoising module is added, F1 and $\mathrm{AP}_{0.5}$ are 0.880 and 0.928,



(a) Original (b) MF (c) AF



(d) Frost (e) Lee (f) NLM

Fig. 5. The denoising effect of different methods.

Table 5
Influence of denoising and FR modules in our method.

Item	Denoising	FR	F1	$AP_{0.5}$
1	×	×	0.811	0.911
2	✓	×	0.880	0.928
3	×	✓	0.925	0.934
4	✓	✓	0.930	0.951

Table 6
The detection resultsof different methods in offshore scenes.

Methods	P	R	F1	$AP_{0.5}$
RetinaNet	0.989	0.910	0.948	0.913
EfficientDet	**0.990**	0.746	0.851	0.892
Faster R-CNN	0.813	0.931	0.868	0.917
YOLOX	0.964	0.923	0.943	0.920
DFR (Ours)	0.960	**0.944**	**0.952**	**0.966**

respectively. This proved that adding denoising module could be used to improve F1 and $AP_{0.5}$. That is to say, denoising module can effectively improve the detection performance. This is because the denoising module can reduce the interference of noise on the detected object. When both the denoising module and the FR model are added, F1 and $AP_{0.5}$ are 0.930 and 0.951, respectively. By comparing the second and fourth rows in the table, it is obvious that the FR module bring an improvement in F1 and $AP_{0.5}$. The reason for this is that the FR module could refine the features to achieve a higher positioning accuracy. According to above analyses and results, it could be concluded that the denoising and FR modules can both improve the performance of the model.

In order to further prove the effectiveness of the FR module, we show the visualization results of the ablation experiment about the FR module, as shown in Fig. 7. In SAR ship detection, due the size of ships is small, many methods could not generate appropriate predicted boxes that could fit ships well, resulting in the unqualified IoU. When the IoU is lower than a threshold, the missing detection would happen. From Fig. 7, we can find that the IoU of many predicted boxes have been improved after adding the FR module. That is to say, the predicted boxes generated by the method with FR module would fit object ships better. This is because that the FR module could determine the center points of object ships accurately, and the information would be transferred to the size prediction branch. In a word, the results of Fig. 7 prove that FR module could improve the positioning accuracy by refining features. The proposed method obtains higher IoU and reduces missing detection caused by unqualified IoU by adding the FR module.

4.5. Comparison with other methods

The open sea area in SAR images has a single background and generally has fewer interference factors. But in inshore scene, there is more complicated background interference than that in offshore scene; hence, inshore ships detection is more difficult than offshore ships detection. Especially when the ship and the land are connected as a whole, the detection of the inshore ship would become more difficult. Experiments were carried out in the offshore, inshore and entire scenes to evaluate the performance of DFR. Current mainstream object detection algorithms, such as RetinaNet, EfficientDet, Faster R-CNN, and YOLOX are directly compared with our method to verify the effectiveness of the proposed method.

Table 6 shows the ship detection results of different methods in offshore scenes. The table shows that the detection results of these methods are relatively good because the offshore scene is relatively simple and the object suffers slight interference. The F1 and $AP_{0.5}$ of our method are 0.952 and 0.966, respectively; these results were the best among these methods. As for RetinaNet and EfficientDet, they

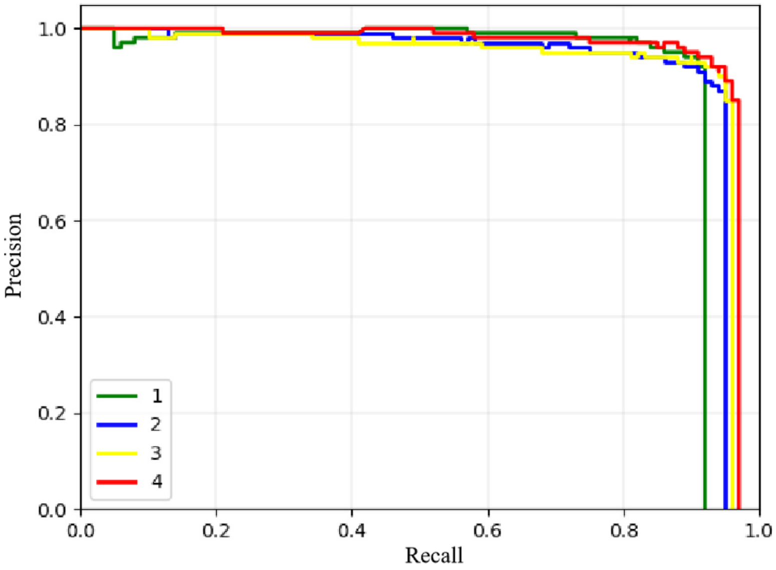


Fig. 6. The PR curves of ablation experiment.

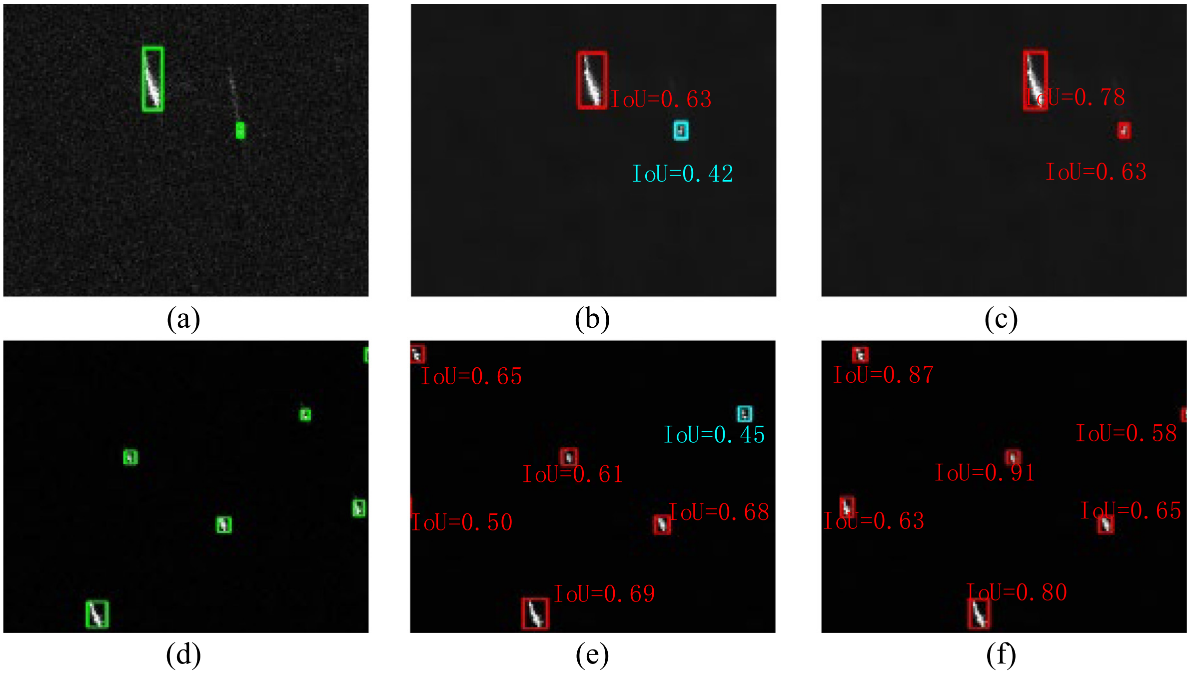


Fig. 7. Visualization results of the ablation experiment about FR module. (a) and (d) ground truth, (b) and (e) without FR, (c) and (f) with FR. The green, red, light blue boxes represent the ground truth, right results and results with unqualified IoU, respectively.

could both obtain features with different resolutions, but the anchor mechanism limits their ability of detecting small objects. The features extracted by Faster R-CNN are all single-layered and own a small resolution, making it difficult to detect small objects. However, our method could obtain fused features with different resolutions by using HFF module, resulting in a better result than that of R-CNN-based method.

The ship detection results of different methods in inshore scenes are shown in Table 7. From the table, it is clear that the evaluation indicators of the detection results generally decrease in the inshore scenes, compared with the detection results in offshore scenarios. This is because inshore scenarios own more complicated background, and the buildings on the shore cause great interference to ship detection. In this case, DFR still achieved the best test results, with F1 and $AP_{0.5}$ of 0.848 and 0.884, respectively. The detection results of our method are better than YOLOX, because the FR module of our method can refine features and then making the positioning of ships more accurate.

Table 7
The detection results of different methods in inshore scenes.

Methods	P	R	F1	$AP_{0.5}$
RetinaNet	0.814	0.686	0.745	0.669
Efficientdet	0.867	0.382	0.531	0.560
Faster R-CNN	0.472	**0.892**	0.617	0.829
YOLOX	0.815	0.863	0.838	0.824
DFR (Ours)	**0.875**	0.824	**0.848**	**0.884**

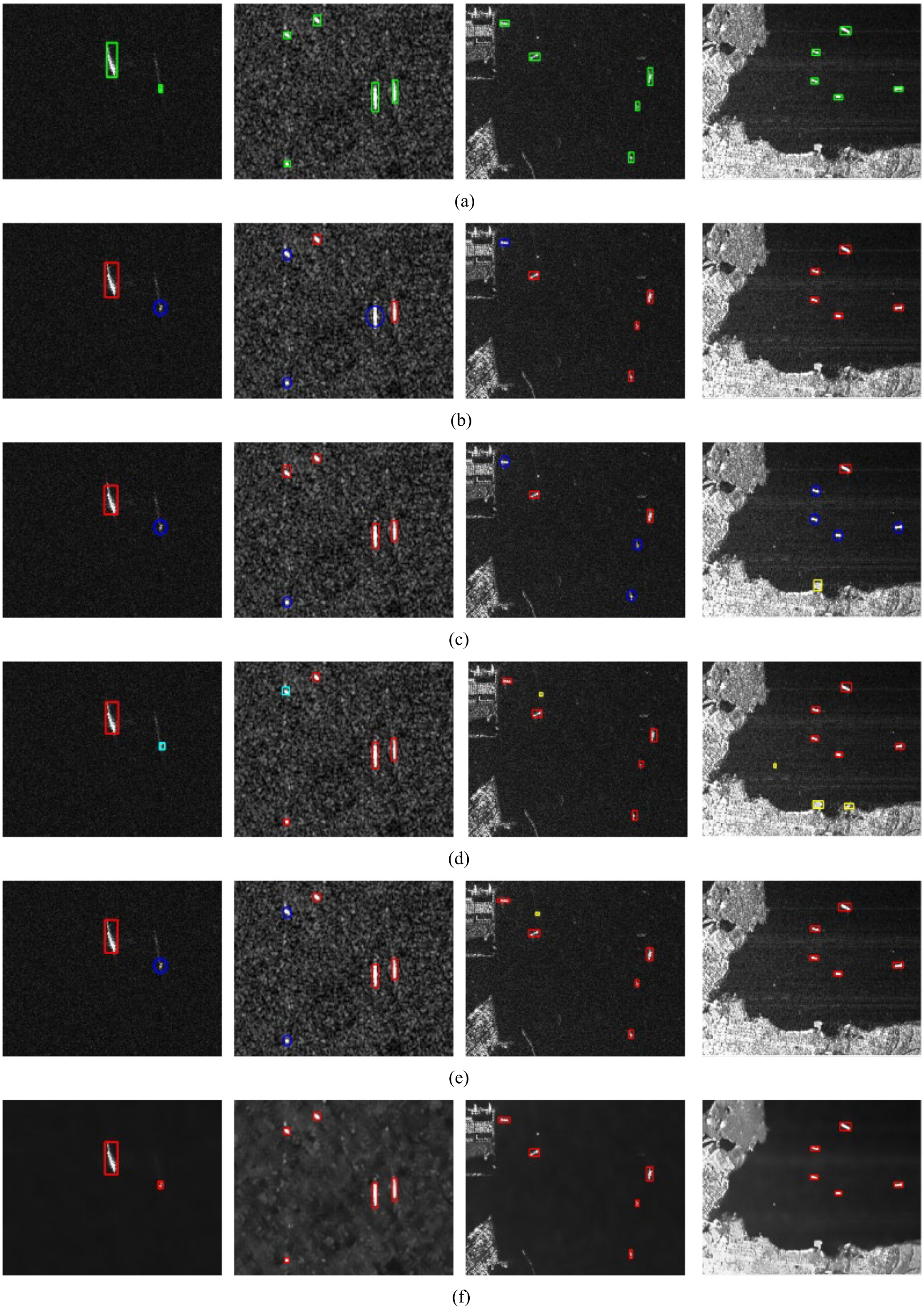

(a)
(b)
(c)
(d)
(e)
(f)

Fig. 8. Visualization results of ship detection. (a) ground truth, (b) RetinaNet, (c) EfficientDet, (d) Faster R-CNN, (e) YOLOX, (f) DFR(Ours). The green, red, yellow and light blue boxes represent the ground truth, right results, false results and results with unqualified IoU, respectively. The dark blue circle represents the missing results.

The ship detection results of different methods in entire scenes are shown in Table 8. The experimental results of HRSDNet and Cascade R-CNN come from the paper written by Zhang et al. (2021). From the table, it is found that although the F1 of HRSDNet is highest, the F1 of our method is very close to it and our method has higher $AP_{0.5}$ and lower GFlops. The $AP_{0.5}$ of DFR is 0.951 and it is higher than other methods. In addition, our method has lowest model size. As for GFlops, our method is higher than YOLOX and EfficientDet, but our method gets better results than these two methods in other evaluation metrics.

Several representative scenes for ship detection in the SSDD test dataset are selected to present the visualization results, as shown in Fig. 8. Fig. 8(a) shows the ground truth, and the detection results of

Table 8
The detection results of different methods in entire scenes.

Methods	F1	$AP_{0.5}$	GFlops	Model size
RetinaNet	0.906	0.862	68.81	416 MB
EfficientDet	0.793	0.829	**4.62**	46 MB
Faster R-CNN	0.801	0.901	226.41	108 MB
YOLOX	0.920	0.901	26.64	69 MB
Cascade R-CNN	0.924	0.905	91.05	532 MB
HRSDNet	**0.937**	0.908	131.28	694 MB
DFR (Ours)	0.930	**0.951**	47.55	**41** MB

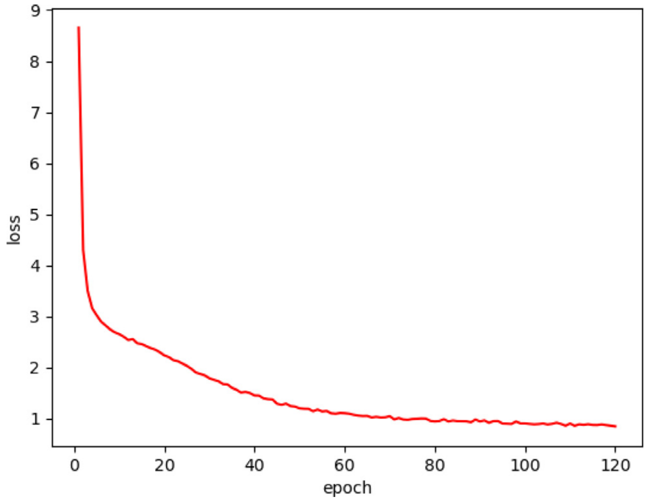


Fig. 9. The training loss curve.

RetinaNet, EfficientDet, Faster R-CNN, YOLOX and DFR are shown in Fig. 8(b)–(f), respectively. From the figure, it is clear that there are some false and miss detection results for small objects in other methods. And the fourth column images of Fig. 8 indicate that the background of inshore scene would have a big negative effect on the accuracy of detection model, which causes some false detection results. Feature refinement is beneficial to improve localization accuracy. To sum up the above descriptions, DFR not only can deal with small objects well, but also can achieve good performance by refining features.

The method proposed by us is general and could be applied in both inshore and offshore scenarios. The related researches in the offshore scenes are relatively mature, so these methods have achieved good results. As for inshore scene, there is a strong interference of the complex background. Moreover, the size of ships in SAR images is generally small, containing too few discriminative features. Thus, it is difficult to detect inshore ships in SAR images. In this case, the improvement in detection performance of inshore scene is meaningful. In addition, our method has lower GFlops and smallest model size.

On the other hand, the stability and convergence are important properties for evaluating model performance. In deep learning, when a model could reach a stable state within finite steps, it is convergent. Loss function is a common method to evaluate the stability of a model. Loss function measures the accuracy of model fitting data. If loss function gradually decreases and becomes stable in finite steps, the model is considered as convergent model. The change in training loss with increasing epoch is shown in Fig. 9. From the figure, it is clear that the loss value is relatively high in the early stage of training, and then with the increase of epoch, the loss value drops rapidly and tend to stabilize. This fully proves that our model is convergent.

5. Conclusion

The existence of speckle noise in SAR images would affect the quality of the extracted feature. Detecting small ships in large scenarios is a great challenge. In addition, background interference leads to inaccurate localization. To address these problems, an SAR ship localization method with denoising and feature refinement is proposed. Firstly, a denoising module is used to suppress the speckle noise inherent in SAR images. We analyze the denoising results of many different denoising methods and finally select the NLM with the best denoising effect as our denoising module. The quality of image and the extracted features are improved after processing by this module. Secondly, a hierarchical feature fusion module is proposed, which can integrate features of different levels. The fused features contain rich semantic information and location information, which are conducive to the detection of objects at different scales. Finally, the FR module is proposed to reduce the background interference to improve the localization accuracy. Extensive experiments are conducted to prove the effectiveness of these proposed modules. According to experimental results, it is clear that the proposed denoising and FR modules could effectively improve the performance of the detector. The proposed HFF module could further improve the detection accuracy of small ships. In numerical results, the $AP_{0.5}$ values of DFR are 0.884, 0.966 and 0.951 in inshore, offshore and entire scenes, respectively, achieving the best results. This proves the superiority of DFR.

It is very important to accurately locate ships, which are the main carriers of marine resource exploitation and economic activities. In traditional ship detection, maritime personnel are required to locate the position of ships through video observation. But our country owns a vast sea area and the number of ships is large, traditional ship detection would consume a lot of manpower. At the same time, long-term monitoring will cause visual fatigue to the monitoring personnel, meaning that detection accuracy cannot be guaranteed. However, using the method proposed in this paper to achieve ship detection cannot only accurately locate ship objects in the complex sea background, but also reduce a lot of labour costs. In future work, we would firstly consider how to apply the proposed method to other fields, such as thermal imaging detection (Glowacz, 2021a,b), marking lesion locations in medical images, and detecting offshore oil production platform, etc. Then, consider denoising within deep learning framework to obtain better denoising effect. What is more, we would also compare the proposed method with other networks after denoising process to study on how to further improve detection accuracy.

CRediT authorship contribution statement

Cheng Zha: Conceptualization, Investigation, Methodology, Writing - original draft. **Weidong Min:** Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing. **Qing Han:** Conceptualization, Methodology, Software, Supervision. **Wei Li:** Conceptualization, Validation, Writing – review & editing. **Xin Xiong:** Conceptualization, Investigation, Methodology, Validation. **Qi Wang:** Investigation, Validation, Visualization, Supervision. **Meng Zhu:** Software, Investigation, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The relevant code of our paper has been submit to Github. The website of the code is https://github.com/Ezrealzc/DFR.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62076117 and Grant 62166026, and in part by Jiangxi Key Laboratory of Smart City under Grant 20192BCD40002.

Appendix A

See Tables A.1 and A.2.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.engappai.2023.106444.

Table A.1
List of acronyms used in the paper.

Acronym	Full form
SAR	Synthetic aperture radar
DFR	Synthetic aperture radar ship localization method with denoising and feature refinement
HFF	Hierarchical feature fusion
FR	Feature refinement
YOLO	You only look once
RPN	Region proposal networks
IoU	Intersection over union
FCN	Fully convolutional networks
RoI	Region of interest
MF	Median filtering
AF	Average filtering
NLM	Non-local mean
BM3D	Block matching 3D
PCA	Principal component analysis
K-SVD	K-singular value decomposition
DCT	Discrete cosine transform
WNNM	Weighted nuclear norm minimization
LDCT	Low dose CT
SACNN	Self-attention convolutional neural network
FSFADet	Arbitrary-oriented ship detection based on feature separation and feature alignment
FPN	Feature pyramid network
SM	Strip map
ST	Staring spot-light transmitter
HS	High-resolution spot-light
ENL	Equivalent number of looks
AP	Average precision
TP	True positive
FP	False positive
FN	False negative

Table A.2
List of variables.

Variables	Paraphrase
D	The side length of the search window
d	The side length of the neighborhood window
N	The noisy SAR ship image
C	The coordinate domain
T	The denoised image
x, y	The pixel point
$s(x, y)$	The similarity between x and y
$V(x)$	The image block centered on x
$V(\mathrm{y})$	The image block centered on y
$Z(\mathrm{x})$	The normalized coefficient
h	The smoothing parameter
H	Height of noising image
W	Width of noising image
(i, j)	A point to be filtered in the noise image
(m, n)	The center of the neighborhood window
$S_{m,n}$	Represents the pixel value at the search window (m, n)
C_1	Feature maps of 2 times down-sampling
C_2	Feature maps of 4 times down-sampling
C_3	Feature maps of 8 times down-sampling
C_4	Feature maps of 16 times down-sampling
C_5	Feature maps of 32 times down-sampling
$Conv(\cdot)$	The 3×3 convolutional layer
$U(\cdot)$	Up sampling
$F(\cdot)$	The fusion process
P	The feature of the first level fusion
L	The fused feature of the second layer
I	Final fused feature map
X	The input feature map
R	A regular grid for sampling over the features X
l_i	The locations in R
$\triangle l_i$	Kernel offset
G	Refined feature map
$(x_t,\ y_t)$	The center point of the ship in T
z	The times of down-sampling
$(x_t/z,\ y_t/z)$	The center point of the ship on the heat map
o_x	The offsets in the horizontal directions
o_y	The offsets in the vertical directions
M	The number of ships in the image T
α, β	The penalty factors
γ_{xy}	The value of (x, y) positions in the predicted heat map
w	The width of the object box
h	The height of the object box
L_{center}	The loss functions of the center prediction branch
L_{size}	The loss functions of the size prediction branch
L_{offset}	The loss functions of the offset prediction branch
μ	The average value
σ	The variance value

(continued on next page)

References

Anfinsen, S., Doulgeris, A., Eltoft, T., 2009. Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery. IEEE Trans. Geosci. Remote Sens. 47 (11), 3795–3809.

Bai, T., Pang, Y., Wang, J., Han, K., Luo, J., Wang, H., Lin, J., Wu, J., Zhang, H., 2020. An optimized faster R-CNN method based on drnet and RoI align for building detection in remote sensing images. Remote Sens. 12 (5), 762.

Bc, A., Prakash, H.N., 2020. Multimodal biometric recognition: Fusion of modified adaptive bilinear interpolation data samples of face and signature using local binary pattern features. Int. J. Eng. Adv. Technol. 9 (3), 3111–3120.

Bera, S., Biswas, P., 2021. Noise conscious training of non local neural network powered by self attentive spectral normalized Markovian patch GAN for low dose CT denoising. IEEE Trans. Med. Imaging 40 (12), 3663–3673.

Cai, Z., Vasconcelos, N., 2018. Cascade R-CNN: Delving into high quality object detection. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 6154–6162.

Chen, K., Ouyang, W., Loy, C.C., Lin, D., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., 2019. Hybrid task cascade for instance segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 4974–4983.

Dai, W., Mao, Y., Yuan, R., Liu, Y., Pu, X., 2020. A novel detector based on convolution neural networks for multiscale SAR ship detection in complex background. Sensors 20 (9), 2547.

Dai, J., Qi, H., Xiong, Y., Li., Y., Zhang, G., Hu, H., Wei, Y., 2017. Deformable convolutional networks. In: Proc. IEEE Int. Conf. Comput. Vis. pp. 764–773.

Deledalle, C., Denis, L., Tupin, F., Reigber, A., Jager, M., 2014. NL-SAR: a unified non-local framework for resolution-preserving (Pol)(In)SAR denoising. IEEE Trans. Geosci. Remote Sens. 53 (4), 2021–2038.

Elyouncha, A., Eriksson, L., Brostrm, G., Axell, L., Ulander, L., 2021. Joint retrieval of ocean surface wind and current vectors from satellite SAR data using a Bayesian inversion method. Remote Sens. Environ. 260, 112455.

Fortino, G., Zhou, M., Hassan, M., Pathan, M., Karnouskos, S., 2021. Pushing artificial intelligence to the edge: Emerging trends, issues and challenges. Eng. Appl. Artif. Intell. 103, 104298.

Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J., 2021. YOLOX: Exceeding YOLO series in 2021. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit.. http://dx.doi.org/10.48550/arXiv.2107.08430.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 580–587.

Glowacz, A., 2021a. Thermographic fault diagnosis of ventilation in BLDC motors. Sensors 21 (21), 7245.

Glowacz, A., 2021b. Ventilation diagnosis of angle grinder using thermal imaging. Sensors 21 (8), 2853.

Guo, H., Yang, X., Wang, N., Gao, X., 2021. A CenterNet++ model for ship detection in SAR images. Pattern Recognit. 112 (7), 107787.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–1916.

Hosseini, H., Hessar, F., Marvasti, F., 2015. Real-time impulse noise suppression from images using an efficient weighted-average filtering. IEEE Signal Process. Lett. 22 (8), 1050–1054.

Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X., 2019. Mask scoring RCNN. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 6409–6418.

Huang, G., Liu, Z., Laurens, V., Weinberger, K., 2017. Densely connected convolutional networks. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 4700–4708.

Huang, L., Liu, G., Wang, Y., Yuan, H., Chen, T., 2022. Fire detection in video surveillances using convolutional neural networks and wavelet transform. Eng. Appl. Artif. Intell. 110, 104737.

Katayama, T., Takahashi, T., Iga, K., 2000. Optical pattern recognition experiments of walsh spatial frequency domain filtering method. Japan. J. Appl. Phys. 39 (Part 1, No. 3B), 1576–1581.

Khatir, S., Boutchicha, D., Thanh, C., Tran-Ngoc, H., Nguyen, T., Wahab, M., 2020. Improved ANN technique combined with jaya algorithm for crack identification in plates using XIGA and experimental analysis. Theor. Appl. Fract. Mech. 107, 102554.

Klaine, L., Vozel, B., Chehdi, K., 2005. Varational method using the Kuan filtering approach for the restoration of blurred images with multiplicative noise. In: Proceedings of the Eighth International Symposium on Signal Processing & Its Applications. pp. 343–346.

Lee, J.S., 2009. Digital image enhancement and noise filtering by use of local statistics. IEEE Trans. Pattern Anal. Mach. Intell. 2 (2), 165–168.

Li, M., Hsu, W., Xie, X., Cong, J., Gao, W., 2020a. SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network. IEEE Trans. Med. Imaging 39 (7), 2289–2301.

Li, J., Qu, C., Shao, J., 2017. Ship detection in SAR images based on an improved faster R-CNN. In: Proc. Sar in Big Data Era: Models, Methods and Applications. pp. 1–6.

Li, R., Wang, X., Wang, J., Song, Y., Lei, L., 2020b. SAR target recognition based on efficient fully convolutional attention block CNN. IEEE Geosci. Remote Sens. Lett. 19, 1–5.

Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature pyramid networks for object detection. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 2117–2125.

Lin, T., Goyal, P., Girshick, R., He, K., Dollar, P., 2017b. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 99, 2999–3007.

Nguyen-Le, D., Tao, Q., Nguyen, V., Wahab, M., Nguyen-Xuan, H., 2020. A data-driven approach based on long short-term memory and hidden Markov model for crack propagation prediction. Eng. Fract. Mech. 235, 107085.

Oloumi, D., Winter, R., Kordzadeh, A., Boulanger, P., Rambabu, K., 2020. Microwave imaging of breast tumor using time-domain UWB circular-SAR technique. IEEE Trans. Med. Imaging 39 (4), 934–943.

Pan, Y., Meng, Y., Zhu, L., 2020. SAR image despeckling method based on improved frost filtering. Signal Image Video Process. 15 (1), 843–850.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 779–788.

Redmon, J., Farhadi, A., 2018. YOLOv3: An incremental improvement. http://dx.doi.org/10.48550/arXiv.1804.02767.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149.

Serrano, W., 2022. Deep reinforcement learning with the random neural network. Eng. Appl. Artif. Intell. 110, 104751.

Shi, Q., Tang, X., Yang, T., L, R., Zhang, L., 2021. Hyperspectral image denoising using a 3-D attention denoising network. IEEE Trans. Geosci. Remote Sens. 59 (12), 10348–10363.

Sinha, S., Santra, A., Mitra, S., 2020. Semi-automated impervious feature extraction using built-up indices developed from space-borne optical and SAR remotely sensed sensors. Adv. Space Res. 66 (6), 1372–1385.

Sun, W., Wang, R., 2018. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. IEEE Geosci. Remote Sens. Lett. 15 (3), 474–478.

Tan, M., Pang, R., Le, Q., 2020. EfficientDet: Scalable and efficient object detection. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 10781–10790.

Tang, S., Lin, J., 2008. Image denoising with combination of wavelet transform and median filtering. J. Harbin Inst. Technol. 40 (8), 1334–1336.

Tran-Nogc, H., Khatir, S., Ho-Khac, H., Roeck, G., Bui-Tien, T., Wahab, M., 2021. Efficient artificial neural networks based on a hybrid metaheuristic optimization algorithm for damage detection in laminated composite structures. Compos. Struct. 262, 113339.

Tran-Nogc, H., Khatir, S., Le, X., Roeck, G., Wahab, M., 2020. A novel machine-learning based on the global search techniques using vectorized data for damage detection in structures. Internat. J. Engrg. Sci. 157, 103376.

Wang, Q., Lu, Y., Zhang, X., Hahn, J., 2021a. Region of interest selection for functional features. Neurocomputing 422 (2), 235–244.

Wang, Q., Min, W., Han, Q., Liu, Q., Zha, C., Zhao, H., Wei, Z., 2021b. Inter-domain adaptation label for data augmentation in vehicle re-identification. IEEE Trans. Multimed. http://dx.doi.org/10.1109/TMM.2021.3104141.

Wang, S., Wang, H., Zhou, Y., Liu, J., Dai, X., Wahab, M., 2020. Automatic laser profile recognition and fast tracking for structured light measurement using deep learning and template matching. Measurement 169, 108362.

Wei, S., Su, H., Ming, J., Wang, C., Yan, M., Kumar, D., Shi, J., Zhang, X., 2020. Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet. Remote Sens. 12 (1), 167.

Yang, H., Liu, L., Min, W., Yang, X., Xiong, X., 2021. Driver yawning detection based on subtle facial action recognition. IEEE Trans. Multimed. 23, 572–583.

Zanchetta, A., Zecchetto, S., 2020. Wind direction retrieval from sentinel-1 SAR images using ResNet. Remote Sens. Environ. 253 (3), 112178.

Zhang, Z., Pan, X., He, K., Cheng, L., 2020. SAR image segmentation with parallel region merging. Multimedia Tools Appl. 80 (5), 1–21.

Zhang, T., Zhang, X., Liu, C., Shi, J., Wei, S., Ahmad, I., Zhan, X., Zhou, Y., Pan, S., Li, J., 2021. Balance learning for ship detection from synthetic aperture radar remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 182, 190–207.

Zhu, M., Hu, G., Li, S., Zhou, H., Wang, S., 2022. FSFADet: Arbitrary-oriented ship detection for SAR images based on feature separation and feature alignment. Neural Process. Lett. 1–11.

Zhu, L., Xie, Z., Liu, L., Tao, B., Tao, W., 2021. Iou-uniform R-CNN: Breaking through the limitations of RPN. Pattern Recognit. 112 (3), 107816.