

Multifeature Semantic Complementation Network for Marine Oil Spill Localization and Segmentation Based on SAR Images

Jianchao Fan¹, Member, IEEE, Shuai Zhang, Xinzhe Wang², and Jun Xing³

Abstract—Marine oil spill causes severe damage to the marine ecological environment. Synthetic aperture radar (SAR) is widely used in marine oil spill detection due to its all-day and all-weather advantages. However, long stripe shape oil spill areas make it challenging to extract the oil spills effectively. A multifeature semantic complementation network (MFSCNet) is proposed for oil spill localization and segmentation of SAR images in one framework to address these problems. The long strip shape interference of oil spill is reduced by extracting intensity and damping ratio characters from nonpolarimetric features and entropy, anisotropy, and mean scattering angle from polarization features to form a multifeature SAR image. Then, the backbone feature network and feature fusion module are used for feature extraction. The decoupled head and the proposed oil spill semantic segmentation head are used for localization and semantic segmentation tasks, respectively. Also, the semantic complementation module is used in the training phase. It combines the results of localization and semantic segmentation to obtain complementation boxes for interactive iterative updating of the model parameters to enhance the detection accuracy of localization boxes. The effectiveness of the proposed model is demonstrated based on a lot of Sentinel-1 oil spill data compared with other state-of-the-art methods.

Index Terms—Localization, multifeature, oil spill detection, SAR, semantic complementation, semantic segmentation.

I. INTRODUCTION

MARINE oil spill is a huge hazard and will cause severe damage to marine ecosystem once happen [1]. Frequent marine activities increase the risk of marine oil spills. It is

Manuscript received 16 January 2023; revised 3 March 2023; accepted 29 March 2023. Date of publication 3 April 2023; date of current version 20 April 2023. The work was supported in part by the National Natural Science Foundation of China under Grant 42076184, Grant 41876109, Grant 41706195, in part by the National Key Research and Development Program of China under Grant 2021YFC2801000, in part by the National High Resolution Special Research under Grant 41-Y30F07-9001-20/22, and in part by Dalian High Level Talent Innovation Support Plan (2021RD04). (Corresponding authors: Xinzhe Wang; Jun Xing.)

Jianchao Fan is with the School of Control Science and Engineering, Dalian University of Technology, Dalian 116023, China, and also with the Department of Ocean Remote Sensing, National Marine Environmental Monitoring Center, Dalian 116023, China (e-mail: fjchaonmeme@163.com).

Shuai Zhang, Xinzhe Wang, and Jun Xing are with the School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China (e-mail: 20052085400522@xy.dlpu.edu.cn; wxzgm@dlpu.edu.cn; xingjun@dlpu.edu.cn).

The code of this work will be available at <https://github.com/fjc1575/Marine-Oil-Spill/tree/main/MFSCNet> for the sake of reproducibility.

Digital Object Identifier 10.1109/JSTARS.2023.3264007

estimated that about two million tons of oil annually leak into the marine environment [2]. Such a large-scale oil spill not only causes enormous economic losses and environmental damage, but also damages the marine ecological environment of coastal countries. A severe oil spill even threatens human health. The occurrence of marine accidents, such as ship accidents and pipeline ruptures, has led to the emergence of oil spills, which makes the source and location of oil spill random [3]. The early remediation of oil spills is critical to the result. Therefore, timely and accurate monitoring of oil spills is required to facilitate oil spill clean-up operations and protect the marine ecological environment [4], [5].

Satellite remote sensing is widely used in marine oil spill detection due to its comprehensive coverage and high timeliness. In recent years, related researchers have conducted studies in satellite remote sensing [6], [7]. Synthetic aperture radar (SAR) has become a powerful tool for oil spill detection because of its all-day monitoring and unaffected capability by clouds and fog [8], [9], [10]. The presence of oil film on the sea surface suppresses the short gravity capillary waves and reduces the roughness of the oil spill area, which results in oil spill areas appearing dark in SAR images [11]. However, oil spill areas are usually a small part of the whole SAR image, and many natural phenomena also appear as dark spots, such as internal waves and low wind speed areas [1], [12]. It is essential to quickly locate the oil spill area from the entire scene image, realize semantic segmentation, and reduce the calculation amount of many background areas.

Oil spill detection can be divided into three steps: feature extraction, dark spot detection, and segmentation. Feature extraction plays a vital role in oil spill detection, which can enhance oil spill information and reduce the interference of the long strip shape oil spill. Many researchers have extracted oil spill features from SAR images. Topouzelis et al. [13] used a decision forest approach to evaluate a total of 25 features in geometric, physical, and textural and found that a combination of nine features was more effective for oil spill detection. Mdakane et al. [14] extracted geometric, physical, and texture totaling 29 features for oil spill detection. Fifteen significant features that can be used for oil spill detection were identified by using a multifeature selection method. These traditional feature extraction methods only consider the intensity information of SAR images and do not view the phase information. Polarimetric features contain

physical information about oil spills, overcoming the limitations of single polarimetric technology [15], [16]. Ma et al. [17] used both the amplitude and phase information to extract the polarimetric features of SAR images and experimentally demonstrated the capability of multilayer features for oil spill detection. Ren et al. [18] extracted polarimetric features from full-polarimetric SAR data for oil spill detection and experimentally verified the ability of the proposed features and methods to distinguish between crude oil, vegetable oil, and seawater. Song et al. [19] extracted polarimetric features and constructed nine-channel PolSAR data. The experimental results showed that the proposed method could effectively distinguish between oil spills and bio-oil films. However, most of the above methods obtain features from oil spill targets and satellite characteristics themselves, which will produce more redundancy and incompatibility of subsequent algorithms. Selecting salient features organically combined with subsequent positioning and semantic segmentation is critical to ensure the accuracy and spatial consistency of the long strip shape oil spill area.

In recent years, with the development of deep learning, it has been gradually applied to the field of computer vision [20], [21], [22], [23]. Relevant researchers have also used deep learning to remote sensing, such as ship detection and marine aquaculture extraction [24], [25], [26], [27]. Also, many researchers have applied deep learning to oil spill detection. Raeisi et al. [28] combined the cuckoo search algorithm and nonnegative matrix factorization of different Zernike moment features to distinguish oil spills and look-alikes in SAR images. Aghaei et al. [29] proposed OSDES-Net methods for oil spill detection based on efficient ShuffleNet blocks. Laurentiis et al. [30] used airborne uninhabited aerial vehicle SAR data to classify oil spills based on the deep learning framework and successfully classified mineral oil film, biogenic oil film, and clean sea, which verified the oil film classification potential of the convolution neural network. Bianchi et al. [31] used a fully convolutional neural network for oil spill segmentation and then used a classification network to predict 12 different categories, such as shape and texture. Zeng et al. [32] constructed a relatively deep-level DCNN for oil spill classification of SAR images based on the VGG16 network. Aghaei et al. [33] extracted oil spill features after dark spot detection of SAR images and used an improved level set method for classification identification. The classification results of these methods depend on the dark spot detection results, and the inaccuracy of the dark spot detection increases the computation time and false alarms. Yang et al. [34] used You Only Look Once version 4 (YOLOv4) method for oil spill detection in the region of interest, which was effectively validated on an extended dataset by removing tiny oil spill areas in the study area. Huang et al. [35] used an object detection model called Faster R-CNN to detect oil spill areas in large-scale remote sensing images and analyzed the interference of additional factors on oil spill detection to determine the optimal wind speed and incidence angle degrees for oil spill detection. Nieto-Hidalgo et al. [36] used three pairs of convolutional neural networks to form a two-stage oil spill detection model for side-looking airborne radar (SLAR) images, with the first network performing coarse detection and the second acquiring the exact pixels of

the class, showing experimentally that the proposed method outperforms previous methods used for this task. Guo et al. [37] used 4200 images extracted from five original SAR images of oil spill to validate the potential of the SegNet model for dark spot detection in the oil spill area. Wang et al. [38] performed oil spill detection based on the AlexNet model and achieved oil spill segmentation of SAR images through slice classification and morphological filtering. Ma et al. [17] used the amplitude and phase information of SAR data to extract polarimetric features and combined multilayer features with a deep learning network model to achieve oil spill segmentation of SAR images, achieving excellent results in terms of accuracy and inference time. Zhu et al. [39] proposed a contextual and boundary-supervised network (CBD-Net) for oil spill detection, which improved the extraction results of oil spill regions in SAR images with intensity inhomogeneity, high noise, and boundary-blurring by fusing multiscale features, spatial and channel squeeze excitation block and joint loss functions. However, the localization approaches only capture the location of the oil spill and cannot obtain more information, such as the oil spill area. Semantic segmentation methods are susceptible to false detection due to the interference of coherent speckle noise and other marine targets. Therefore, fast localization and semantic segmentation collaboration are more suitable for the actual oil spill detection process.

Related researchers have already researched on simultaneous localization and segmentation. He et al. [40] proposed a two-stage target detection algorithm that enables segmentation during localization. The first stage uses a region proposal network to obtain proposals. The second stage uses RoIAlign to resample the proposal to the same size and perform classification, localization, and segmentation. Then, the target instance segmentation is achieved. Bolya et al. [41] proposed a one-stage instance segmentation model. It implements instance segmentation mainly through two parallel subnetworks. One network is used to generate the class, localization, and mask coefficients of each anchor. The other network generates a set of prototype masks, which are then multiplied with the mask coefficients to obtain the segmentation result for each target. Also, Yekeen et al. [42] used Mask R-CNN to detect marine oil spills, and they realized the detection and segmentation of oil spills, look-alikes, ships, and land areas after pretreatment. Although these methods allow for the localization and segmentation of oil spills, some issues still need to be solved. Due to the long strip shape of the oil spill with non-Gaussian distribution, as shown in Fig. 1, the localization box cannot completely cover the whole oil spill area. Incomplete localization can seriously affect the accuracy of semantic segmentation. However, due to other nonoil spill dark spots in SAR images, as shown in Fig. 1(e), the wrong segmentation tends to occur when only semantic segmentation is used.

To solve the above problems, this article proposes a multifeature semantic complementation network (MFSCNet) for oil spill localization and segmentation of SAR images. MFSCNet uses a deep convolutional neural network to extract depth features from multifeature SAR images. The feature extraction results are used in the decoupled head and the oil spill semantic segmentation

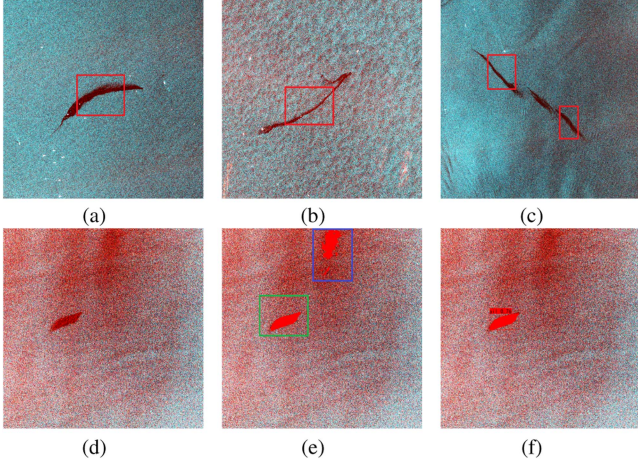


Fig. 1. Marine oil spill localization and segmentation based on SAR false color images with Pauli decomposition. (a)–(c) Localization imperfection when only localization is available, and the red box is the localization box. (d) A SAR image including an oil spill and low wind speed look-alike region with 3.2 m/s wind speed. (e) Oil spill segmentation result. The green box shows the correctly classified oil spill area, while the blue box shows the misclassification of the low wind speed look-alike area. (f) Result of the combination of localization and segmentation. The localization box restricts the segmentation area to avoid misclassification.

head for simultaneous implementation of oil spill localization and segmentation. The semantic complementation module obtains the complementation box by combining the localization and segmentation results in the training phase. Then, the model is updated interactively and iteratively to improve the localization results. The main contributions of this study are as follows:

- 1) MFSCNet is proposed for performing marine oil spill localization and semantic segmentation simultaneously, which can avoid wrong segmentation. It includes multi-feature extraction, PANet for feature fusion, a decoupled head, a segmentation head, and a semantic complementation module.
- 2) A multifeature extraction module is constructed to enhance the oil spill accuracy and spatial consistency and reduce the interference of long stripe shape oil spill area by combining intensity and damping ratio features in non-polarimetric features and entropy (H), anisotropy (A), and mean scattering angle ($Alpha$) in polarimetric features.
- 3) A semantic complementation module is designed in the training phase. It integrates the localization and segmentation results to generate the complementation box, which is iteratively updated to cover the entire oil spill area effectively.

The remainder of this article is organized as follows. Section II introduces the related work related to the research direction of this article. Section III introduces the proposed MFSCNet model in detail for marine oil spill in SAR image. Section IV shows the experimental results of the work. Section V summarizes the highlights of the article and future research.

II. RELATED WORK

Current object detection algorithms can be broadly classified into two categories, one is two-stage object detection algorithm

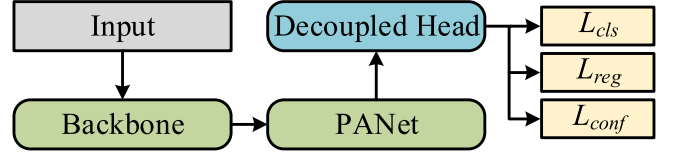


Fig. 2. Framework of YOLOX.

and the other is one-stage object detection algorithm [43]. Two-stage object detection algorithms have higher accuracy but are slower, such as the R-CNN series algorithm [22], [40], [44]. They obtain higher detection accuracy by acquiring region proposals and then performing detection. One-stage object detection algorithm treats detection as a regression problem. It uses a framework to directly implement classification or localization, such as the YOLO series algorithm and SSD [21], [41], [45], [46], etc. This type of algorithm is more advantageous in terms of speed. It is more important to identify the oil spill area for oil spill detection quickly, so the one-stage object detection algorithm is more suitable for oil spill detection. Further, one-stage object detection algorithms can be classified into anchor-based and anchor-free. Anchor-based generates prediction boxes based on a predetermined number of anchors with fixed scales and aspect ratios. In contrast, anchor-free generates prediction boxes directly based on points [47]. For regular targets such as people or cars, the anchor-based approach will be more advantageous in terms of accuracy because their sizes are relatively fixed. However, the dimensions are not suitable for oil spills. The anchor-based approach will limit the ability of the model. So, the anchor-free method is more suitable for oil spill detection.

With the development of the YOLO series, YOLOX [45] is proposed as a one-stage anchor-free object detection algorithm, which is more suitable for oil spill detection. As shown in Fig. 2, the YOLOX model can be divided into four parts from input to output: Backbone, PANet, decoupled head, and loss. For the input SAR image, the feature extraction is first performed using a backbone feature network named CSPDarknet53 [48]. Then, the PANet [49] module is used to fuse the feature extraction results of the last three layers. After fusion, the output is performed using the decoupled head for the three feature layers. Two parallel branches are used in the decoupled head to predict localization boxes, confidence, and classification, respectively. Finally, the SimOTA algorithm is used to perform label assignment to calculate the loss. The loss function is as follows:

$$L_{\text{loss}} = -\frac{1}{N} \sum_{i=1}^N (l_i \log(l'_i) + (1 - l_i) (1 - \log(l'_i))) \quad (1)$$

where the L_{loss} is the loss function of L_{cls} and L_{conf} , l_i is the predicted probability of pixel i , l'_i is the ground-truth label, N is the number.

$$L_{\text{reg}} = 1 - \left(\frac{l'_{\text{box}} \cap l_{\text{box}}}{l'_{\text{box}} \cup l_{\text{box}}} \right)^2 \quad (2)$$

where the L_{reg} is the loss function of the localization boxes, l_{box} is the predicted localization boxes, l'_{box} is the ground-truth boxes.

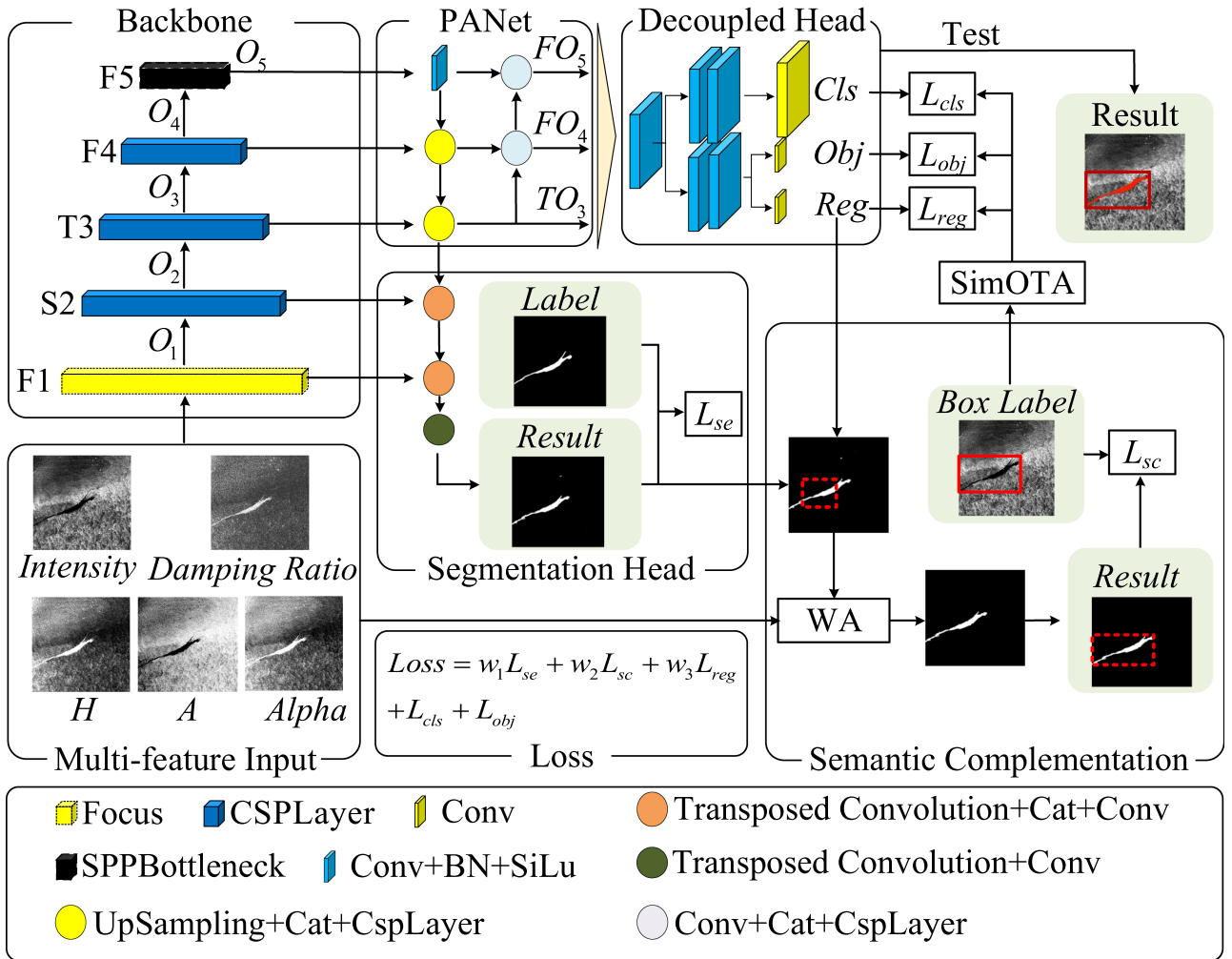


Fig. 3. Framework of MFSCNet. The framework is divided into five parts: multifeature input, backbone for feature extraction, PANet for feature fusion, detection head consisting of a decoupled head and a segmentation head for localization and segmentation, and semantic complementation module.

III. PROPOSED METHOD

The proposed MFSCNet can be divided into five parts, and the framework is shown in Fig. 3. The first part is the multifeature input, which extracts the nonpolarimetric features and polarimetric features of SAR images to form the multifeature SAR image for reducing the interference of long stripe shape oil spill area and enhancing the oil spill information. The second part is a backbone feature network composed of CSPDarknet53 for preliminary feature extraction. The third part is the PANet module, which fuses the semantic and spatial information of the last three feature layers of the backbone by a top-bottom and bottom-up approach. The fourth part is a detection head consisting of a decoupled head and a segmentation head that uses the feature extraction results for the localization and segmentation tasks. The fifth part is the semantic complementation module, which uses the watershed algorithm to obtain complementation boxes by combining the localization and segmentation results. Then, it is used to interactively and iteratively update the model to improve the accuracy.

A. SAR-Based Feature Extraction

For better oil spill detection, a multifeature input module is proposed. A good feature enhances the extracted target information and enables the deep learning model to learn the valuable features better and faster. The long stripe shape oil spill area in SAR images brings some interference to oil spill detection. It is difficult to extract oil spill information effectively by filtering methods, or intensity features only. Therefore, the intensity and damping ratio features in the nonpolarimetric features and the $H/A/Alpha$ feature in the polarimetric features are extracted for oil spill detection.

The intensity feature reflects the image brightness, which represents the backward scattering coefficient of the radar and reflects the roughness of the object. The damping ratio is the ratio of the backward scattering coefficient of the calm sea surface to the backward scattering coefficient of the oil film, which reflects some extent, the emulsification degree of the oil film in SAR images. A larger damping ratio value indicates a thicker oil film, while a smaller damping ratio value indicates a thinner oil film. The polarimetric features reflect the polarimetric scattering

characteristics of the object. H describes the randomness of the target scattering. A is a complementary parameter to H . $Alpha$ indicates the scattering type of the target.

1) *Nonpolarimetric Features*: Due to the interference of imaging weather, geographical location, marine targets, and other factors, most of the target information is in the low gray value area. In contrast, maritime targets, especially ships, are concentrated in the high gray value area. However, this information will interfere with oil spill detection. The intensity features obtained are decimalized to enhance the information in the low gray value region and compress the information in the high gray value region and the decibel is shown in the following:

$$x_{db} = 10 \log_{10} \sigma_0 \quad (3)$$

where σ_0 is the backscattering coefficient and x_{db} is the intensity feature after decibelization.

The damping ratio is the ratio of the backscattering coefficient of the calm sea surface to that of the oil film and some extent and it can be calculated according to the following:

$$x_{dr} = \frac{\sigma_{\text{water}}^0}{\sigma_{\text{oil}}^0} \quad (4)$$

where σ_{water}^0 is the backscattering coefficient of seawater, σ_{oil}^0 is the backscattering coefficient of oil, and x_{dr} is the extracted damping ratio feature.

2) *Polarimetric Features*: SAR can obtain different polarimetric information by controlling the polarimetric mode of transmission and reception. Different polarimetric modes have various sensitivities to other ground objects. Multipolarimetric SAR can get more abundant target information to analyze the scattering mechanism of targets and avoid the uncertainty of target information. For dual-polarimetric SAR data, Cloude–Pottier can be performed to obtain the target’s polarimetric information for better representation.

The C_2 polarimetric covariance matrix must first be extracted to extract the polarimetric information from the SAR image. After preprocessing operations such as multiview and radiometric calibration, the C_2 polarimetric covariance matrix is extracted with the following:

$$C_2 = \begin{bmatrix} \langle 2 |S_{VH}|^2 \rangle & \langle \sqrt{2} S_{VH} S_{VV}^* \rangle \\ \langle \sqrt{2} S_{VV} S_{VH}^* \rangle & \langle |S_{VV}|^2 \rangle \end{bmatrix} \quad (5)$$

with

$$S_{VV} = |S_{VV}| e^{j\phi_{VV}} \quad (6)$$

where $|S_{VV}|$ is the amplitude of VV polarimetric channel, S_{VV} is the complex information of VV polarimetric, $|S_{VH}|$ is the amplitude of VH polarimetric channel, S_{VH} is the complex information of VH polarimetric, ϕ_{VV} is the phase of polarimetric channel, $*$ is conjugate operator, $\langle \rangle$ is statistical mean, and j is imaginary unit.

The eigenvalue is calculated according to the polarimetric covariance matrix:

$$C_2 = \sum_{i=1}^2 \lambda_i e_i e_i^* = \lambda_1 e_1 e_1^* + \lambda_2 e_2 e_2^* \quad (7)$$

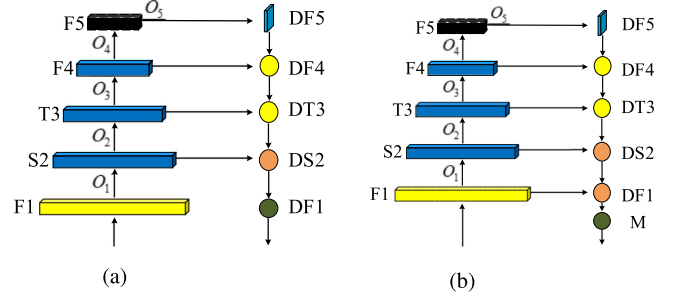


Fig. 4. Different semantic segmentation heads. (a) Rough segmentation head. (b) Fine segmentation head.

where λ_i is the eigenvalues in polarimetric covariance matrix and e_i is the feature vector.

Then the probability of the eigenvector λ_i corresponding to each eigenvalue e_i is calculated, p_i is as follows:

$$p_i = \frac{\lambda_i}{\lambda_1 + \lambda_2} \quad (8)$$

where the λ_1 and λ_2 are two eigenvalues of polarimetric covariance matrix, p_i is proportion probability.

Next, the $H/A/Alpha$ feature is calculated according to the pseudo probability and the eigenvalue obtained above. The $H/A/Alpha$ is shown in following:

$$H = - \sum_{i=1}^2 p_i \log_2 p_i \quad (9)$$

$$A = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \quad (10)$$

$$\alpha = \sum_{i=1}^2 p_i \alpha_i \quad (11)$$

with

$$a_i = \cos^{-1}(|e_{1i}|) \quad (12)$$

where H is the entropy, A is the anisotropy, α is the average polarimetric scattering angle, e_{1i} is the value of the first row, and the first column of the eigenvector.

After feature extraction, different results have different orders of magnitude. Two percent truncated linear stretching is used to unify all data into the same dimension [50].

B. Oil Spill Segmentation Head

An essential problem in oil spill detection of SAR images is the interference of coherent speckle noise. Generally, the shallow features in the depth learning model are more susceptible to the interference of coherent speckle noise. As shown in Fig. 4, two different segmentation heads are proposed for judging the influence of coherent speckle noise on segmentation.

The semantic segmentation module uses the output TO_3 from PANet and the outcomes O_1 and O_2 from Backbone. After TO_3 is resampled and spliced with the output of Backbone of the corresponding size, a convolutional block consisting of a combination of convolution, BN, and SiLu is then used for

feature extraction. Finally, the semantic segmentation result is output by convolution of 1×1 .

The input TO_3 is first resampled to the same size as O_2 using a transposed convolution, then stitched in the channel dimension, followed by feature extraction using two convolution blocks. The feature extraction result continues with a transposed convolution to resample it to the same size as O_1 , followed by feature extraction using the same two convolution blocks. To ensure that the output segmentation size is consistent with the input image, a transposed convolution is used to resample it to the same size as the input. After using one convolution block for feature extraction, a convolution of 1×1 is used to output a semantic segmentation result that is the same size as the original image.

In the training phase, the semantic segmentation loss function with the following:

$$L_{se} = L_{loss} + 1 - \frac{2 \sum_{i=1}^N y_i y'_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N y'_i} \quad (13)$$

where y_i is predicted probability of pixel i , y'_i is the ground-truth label, and N is the number.

In the prediction stage, the activation function is used to obtain the probability of the category to which each pixel on the output feature map belongs. Then the index of the maximum value corresponding to each topic is selected as the category to which it belongs, and the semantic segmentation result is output

$$y''_i = \frac{e^{y'_i}}{\sum_i e^{y'_i}} \quad (14)$$

where y''_i is the output after activation function of pixel i .

The final semantic segmentation result is obtained after the above operation and is denoted as y'_{ss} . In addition, the localization box will be combined with the semantic segmentation result, and the intersection region of the two will be taken as the final segmentation result with the following:

$$y''_{ss} = y'_{ss} \cap y'_{box} \quad (15)$$

where y'_{box} is the predicted localization box, y'_{ss} is the predicted semantic segmentation, y''_{ss} is the result of semantic segmentation after taking the intersection.

C. Semantic Complementation

After obtaining the localization and semantic segmentation results, k samples are accepted according to the label assignment strategy SimOTA. The semantic complementation is performed on these k samples using the watershed algorithm. The outer rectangle of the segmentation results after the complementation is used as the complementation box, and the loss is calculated with the actual value to update the model parameters.

First, get the current semantic segmentation result, which is defined as y'_{ss} . Next, the value of the Intersection over Union (IOU) between the k localization boxes and the actual boxes obtained by the label assignment strategy is calculated. When the result is less than 0.3, semantic complementation is performed. Otherwise, the development of the localization box remains unchanged, and no processing is performed. The intersection

is used as the label map for subsequent completion

$$s_{label} = \begin{cases} y'_{ss} \cap y'_{box}, IOU(y'_{box}, y_{box}) < 0.3 \\ \text{pass, otherwise} \end{cases} \quad (16)$$

where y_{box} is the ground-truth box, s_{label} is the intersection results, *pass* means no semantic complementation.

The intensity data x_{db} is then fed into the watershed algorithm along with the label map s_{label} to obtain the complementation result, which is only performed if there is an oil spill area in the label map. Otherwise, the complementation result is 0

$$s_{mask} = \begin{cases} wa(s_{label}, x_{db}), sum(s_{label}) > 0 \\ 0, \text{otherwise} \end{cases} \quad (17)$$

where $wa(\cdot)$ is the watershed algorithm and s_{mask} is the complementation result.

After obtaining the complementation result, get the outer rectangle of the result to obtain the final complementation box, which will be 0 when the complementation result does not contain the oil spill area

$$y''_{box} = \begin{cases} (h_1, v_1, h_2, v_2), sum(s_{mask}) > 0 \\ 0, \text{otherwise} \end{cases} \quad (18)$$

where y''_{box} is the semantic complementation box, h and v represent abscissa and ordinate, respectively, h_1 and v_1 represent the coordinates of the upper left corner of the complementation box, h_2 and v_2 represent the coordinates of the lower right corner of the complementation box.

The semantic complementation module is formally described in Algorithm 1, the loss function of the semantic complementation module is the same as that of L_{reg} .

Finally, the total loss function formula is shown below:

$$Loss = w_1 L_{se} + w_2 L_{sc} + w_3 L_{reg} + L_{cls} + L_{obj} \quad (19)$$

where $Loss$ is the total loss, L_{se} is the loss of semantic segmentation, L_{sc} is the loss of semantic complementation, L_{reg} is the loss of localization box, L_{cls} is the loss of classification, L_{cls} is the loss of confidence, w_1 is the weighting factor of L_{se} , w_2 is the weighting factor of L_{sc} , w_3 is the weighting factor of L_{reg} .

IV. EXPERIMENTS

A. Implementation Details

All experiments are compiled under Windows 10 with python 3.6, pytorch 1.7.1, and cuda 11.0, run with GeForce RTX 3080 GPU. The optimization algorithm uses Adam. The initial learning rate of the network is $1e-4$. SetpLR is used to adjust the learning rate of each epoch, and the adjustment multiple is 0.92. The value of w_1 and w_2 are 2, 0.1, respectively. The value of w_3 is 5, referring the result of literature [45]. The training epoch is set as 100. Moreover, the total training time is 5.6 h. The proposed model is trained without the semantic complementation module in the first 80 epochs, and it is utilized in the final 20 epochs.

Algorithm 1: Semantic Complementation.

Input: x_{db} : intensity data after processing
 k : k samples obtained by label allocation
 y'_{box} : location boxes obtained by prediction
 y'_{ss} : segmentation obtained by prediction
 y_{box} : ground-truth boxes
 s_{label} : intersection results
 s_{mask} : complementation results
 y''_{box} : semantic complementation boxes
 $IOU(\cdot)$: intersection over union
 $wa(\cdot)$: watershed algorithm

Output: Semantic complementation result.

for $epochs$ **do**
 for $i = 1$ to k **do**
 if $IOU(y'_{\text{box}}(i), y_{\text{box}}(i)) < 0.3$ **then**
 $s_{\text{label}}(i) = y'_{\text{box}}(i) \cap y'_{ss}$
 if $sum(s_{\text{label}}(i)) > 0$ **then**
 $s_{\text{mask}}(i) = wa(s_{\text{label}}(i), x_{db})$
 else
 $s_{\text{mask}}(i) = 0$
 end if
 if $sum(s_{\text{mask}}(i)) > 0$ **then**
 $y''_{\text{box}}(i) = (h_1, v_1, h_2, v_2)(i)$
 else
 $y''_{\text{box}}(i) = 0$
 end if
 else
 $y''_{\text{box}}(i) = y'_{\text{box}}(i)$
 end if
 end for
 Loss $L_{sc} = 1 - \left(\frac{y''_{\text{box}}(i) \cap y_{\text{box}}}{y''_{\text{box}}(i) \cup y_{\text{box}}}\right)^2$
 end for

B. Dataset

Existing oil spill datasets are difficult to obtain, and there are few publicly available datasets to evaluate the dependability of oil spill detection technologies. Therefore, 82 Sentinel-1 data scenes are obtained from the Alaska Satellite Facility (ASF) data distribution website (<https://search.asf.alaska.edu>) for the period 2014–2021 based on oil spill information provided by other researchers and the findings of long-term monitoring of the Chinese Bohai Sea [17], [39], [51], [52], [53], [54]. Sentinel-1 data in vertical emission vertical reception (VV) and vertical emission horizontal reception (VH) polarimetric modes in Interferometric Wide (IW) swath strip scan mode are used in this work to extract different features.

Thermal noise removal, radiometric calibration, filtering, terrain correction, and multilooking preprocessing are all conducted on the obtained data. The filtering employs a 7×7 refined Lee filter and the multilooking size is 1×5 . The oil spill region is then clipped out. Also, resampling and data augmentation are carried out. Finally, 1024×1024 pixels SAR images with 20×20 m resolution of 1050 views are obtained. The ratio of training and test sets is 7:3. This means that the number of training sets is 730, and the number of test sets is 320. Some sample images

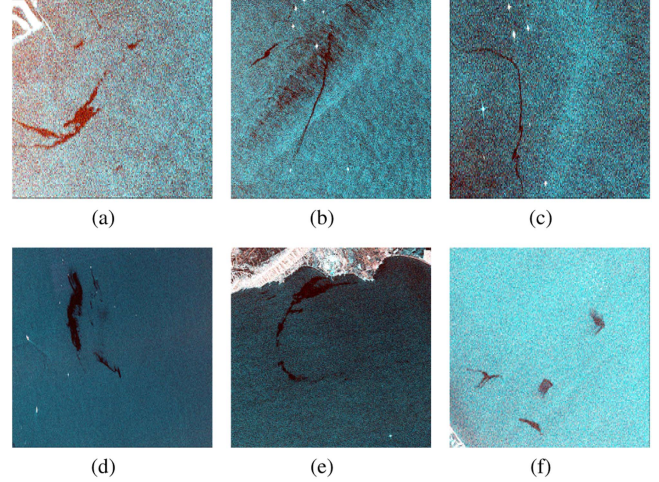


Fig. 5. Training dataset sample images.

are shown in Fig. 5. In the experiment, the ground-truth of oil spill SAR images are collected through relevant news reports, literature publications, and our daily oil spill monitoring. The software of labelme is also used for labeling ground-truth labels.

C. Evaluation Criteria

In order to objectively evaluate the performance of different models, different evaluation metrics are used to compare the results of localization and semantic segmentation. Average precision (AP) is used for localization. Overall accuracy (OA), mean intersection over Union ($MIOU$), $F1$ -score ($F1$), and $Kappa$ are used for semantic segmentation.

1) *The Evaluation Criteria of Localization:* Before AP calculation, $Precision_{\text{box}}$ and $Recall_{\text{box}}$ need to be obtained. These two values need to be obtained according to intersection over union (IOU) with the following:

$$IOU = \frac{P_B \cap T_B}{P_B \cup T_B} \quad (20)$$

where P_B is the prediction boxes and T_B is the ground-truth boxes.

Next, $Precision_{\text{box}}$ and $Recall_{\text{box}}$ are obtained for different confidence levels according to the confidence level of the prediction boxes, with the following:

$$Precision_{\text{box}} = \frac{TP}{TP + FP} \quad (21)$$

$$Recall_{\text{box}} = \frac{TP}{TP + FN} \quad (22)$$

where TP , TN , FP , and FN denote the number of true positives, true negatives, false positives, and false negative samples, respectively. $Precision_{\text{box}}$ and $Recall_{\text{box}}$ are the result of localization.

After receiving $Precision_{\text{box}}$ and $Recall_{\text{box}}$ at different confidence levels, the P-R curve is plotted, the area under the P-R

curve is calculated, and the result is the AP with the following:

$$AP = \int_0^1 p(r) dr \quad (23)$$

where $p(r)$ is the value of $\text{Precision}_{\text{box}}$ under different confidence levels.

2) *Evaluation Criteria of Semantic Segmentation*: The metrics OA , $F1$, $MIoU$, and $Kappa$ are used to compare the semantic segmentation results of different methods. There are calculated as follows:

$$OA = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (24)$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \quad (25)$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \quad (26)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (27)$$

$$MIoU = \frac{1}{2} \left(\frac{T_P}{T_P + F_P + F_N} + \frac{T_N}{T_N + F_N + F_P} \right) \quad (28)$$

$$Kappa = \frac{OA - P_e}{1 - P_e} \quad (29)$$

where

$$P_e = \frac{(T_N + F_P) \times (T_N + F_N) + (T_P + F_P) \times (T_P + F_N)}{N \times N} \quad (30)$$

D. Results and Analysis

1) *Comparison With Other Oil Spill Detection Methods*: To validate the performance of the proposed model adequately, it is compared with the present oil spill detection algorithm. The proposed method is compared with AlexNet¹ [38], SegNet² [37], Faster R-CNN³ [35], YOLOV4⁴ [34], YOLOX⁵, Mask R-CNN⁶ [42], and YOLACT⁷ [41]. Due to the differences between the various comparison methods, they are compared in terms of localization and semantic segmentation, respectively.

The experimental results are shown in Table I. The proposed MFSCNet achieves the highest accuracy in both localization and semantic segmentation. Unlike conventional object detection, the shape and size of the oil spill area are not uniform and exhibit a non-Gaussian normal distribution. This leads to anchor-based depth learning object detection algorithms such as Faster R-CNN, YOLOV4, Mask R-CNN, and YOLACT, which are difficult to find a suitable anchor. Thus, reducing the localization

¹[Online]. Available: <https://github.com/fjc1575/Marine-Oil-Spill/tree/main/AlexNet>

²[Online]. Available: <https://github.com/fjc1575/Marine-Aquaculture/tree/main/SegNet>

³[Online]. Available: https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html

⁴[Online]. Available: <https://github.com/bubliiiiing/yolov4-pytorch>

⁵[Online]. Available: <https://github.com/bubliiiiing/yolox-pytorch>

⁶[Online]. Available: https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html

⁷[Online]. Available: <https://github.com/bubliiiiing/yolact-pytorch>

TABLE I
COMPARISON OF MFSCNET WITH OTHER OIL SPILL DETECTION METHODS

Methods	$AP(\%)$	$OA(\%)$	$F1(\%)$	$MIoU(\%)$	$Kappa$
AlexNet [38]	-	88.22	55.85	66.46	0.55
SegNet [37]	-	98.12	55.08	69.46	0.54
Faster R-CNN [35]	54.76	-	-	-	-
YOLOv4 [34]	61.47	-	-	-	-
YOLOX [45]	79.05	-	-	-	-
Mask R-CNN [42]	55.58	99.2	57.29	74.6	0.57
YOLACT [41]	73.88	99.36	72.63	79.75	0.72
MFSCNet	86.24	99.41	79.10	83.20	0.79

TABLE II
COMPARISON OF THE COMPUTATIONAL COST OF MFSCNET TO OTHER OIL SPILL DETECTION METHODS

Methods	$Time(s)$	$GFLOPs$
AlexNet [38]	7.57	0.15
SegNet [37]	0.20	1310.72
Faster R-CNN [35]	0.17	168.41
YOLOv4 [34]	0.31	360.64
YOLOX [45]	0.30	66.98
Mask R-CNN [42]	0.24	206.10
YOLACT [41]	0.13	338.17
MFSCNet	0.31	141.60

accuracy. As an anchor-free algorithm, YOLOX does not rely on anchors, only on the model itself. It is more conducive to identify the targets with an irregular distribution and has better universality for oil spill detection. The proposed MFSCNet in this study has better detection results and achieves 86.24% AP , combined with the characteristics of YOLOX and SAR images.

This study also achieve the highest accuracy in semantic segmentation. The segmentation results are shown in Fig. 6.⁸ It can be seen that MFSCNet proposed in this article can extract oil spill regions of different sizes and shapes completely with high accuracy. The Alexnet method performs semantic segmentation by slider classification. Mask R-CNN resamples the feature maps to the same size by RoIAlign before performing semantic segmentation. This leads to poor prediction results in the edge region for both methods, which are prone to misclassification. In contrast, SegNet is semantic segmentation of the whole image, and the presence of targets such as coherent speckle noise and maritime targets affects the segmentation results. In contrast, YOLACT combines localization and semantic segmentation, and each localization box has a semantic segmentation result. The localization boxes limit the accuracy of semantic segmentation, leading to the truncation phenomenon shown in Fig. 6(f). This splits a complete oil spill target into two separate targets, reducing the accuracy of the split. Finally, the MFSCNet proposed in this article treats localization and semantic segmentation as two parallel tasks that restrict the semantic segmentation results through the localization box and takes the intersection region of both as the final segmentation result to extract the oil spill area effectively. Also, Table II shows the computational cost of different oil spill detection methods. Since the AlexNet method classifies for small images instead of segmentation, it takes a longer time but has lower Giga Floating Point of Operations

⁸Original SAR images in the experiments are presented as Pauli false color images.

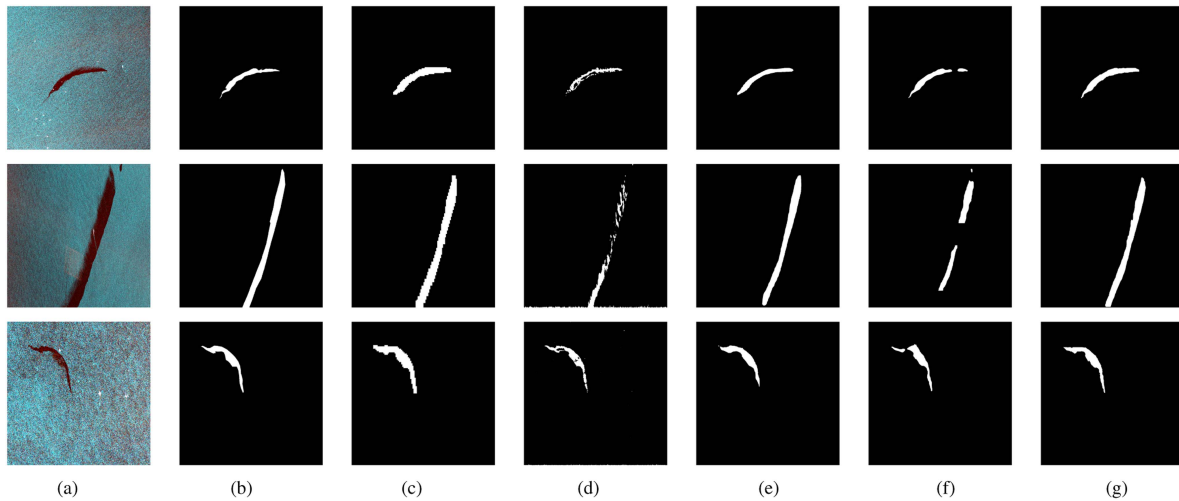


Fig. 6. Experimental results of semantic segmentation. (a) Original images. (b) Ground-truth. (c) AlexNet. (d) SegNet. (e) Mask R-CNN. (f) YOLACT. (g) MFSCNet.

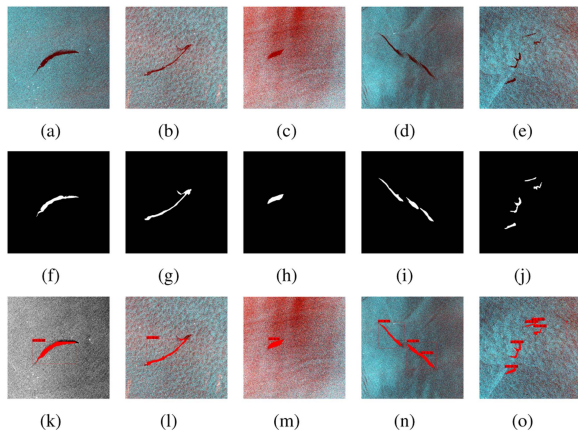


Fig. 7. Results of MFSCNet. (a)–(e) original images. (f)–(j) ground-truth maps. (k)–(o) predict results.

(GFLOPs) [45] values. The MFSCNet proposed in this study achieves a balance between time and GFLOPs. Fig. 7 shows the prediction results of the proposed method in this article. It can be seen that the SAR oil spill images of different sizes and shapes can be extracted well. Also, the image with other nonoil dark spots can also be well extracted.

2) *Comparison of Different Features*: This study uses the intensity and damping ratio features in the nonpolarimetric features and the $H/A/Alpha$ features in the polarimetric features. Oil spill detection is carried out by combining polarimetric with nonpolarimetric features. The detection results are shown in Table III, and the visualization results under different features are shown in Fig. 8. The combination of different features has higher accuracy in localization than single features. This indicates that the extracted multifeature module is effective. Although, the addition of the damping ratio feature enhances the information on the dark spot region and improves the detection accuracy. However, the damping ratio reflects the emulsification degree of the oil film. It will have different feature extraction results for

TABLE III
COMPARISON OF DIFFERENT FEATURES

<i>Intensity</i>	<i>Damping Ratio</i>	<i>H/A/Alpha</i>	<i>AP(%)</i>
✓			79.05
	✓		81.63
		✓	81.12
✓	✓		81.96
✓		✓	81.67
	✓	✓	81.95
✓	✓	✓	82.85

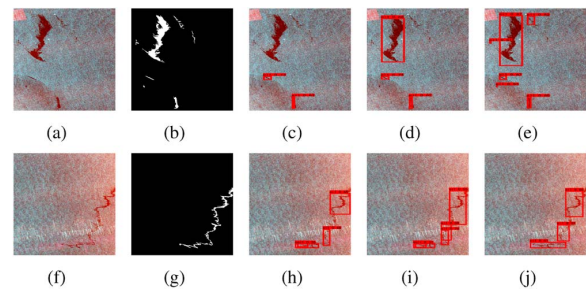


Fig. 8. Result of different features. (a) and (f) are original images, (b) and (g) are ground-truth maps, (c) and (h) are the result of intensity, (d) and (i) are the result of intensity and damping ratio, (e) and (j) are the result of intensity, damping ratio, and $H/A/Alpha$.

the same oil spill area, which is difficult to have consistency. Thus, the phenomenon of overlapping localization boxes will occur, shown in Fig. 8(i) and polarimetric features can effectively maintain the spatial consistency of the oil spill area by obtaining richer target information. Both the sparse oil spill area in Fig. 8(a) and the long strip of the oil spill area in Fig. 8(f) can be detected well. Finally, 82.85% is achieved in AP , which indicates that the multifeature module proposed in this study is practical and feasible to effectively enhance the spatial consistency of oil spills and improve the oil spill detection accuracy.

TABLE IV
COMPARISON OF DIFFERENT MODULES

<i>MF</i>	<i>SS</i>	<i>SC</i>	<i>AP</i> (%)	<i>OA</i> (%)	<i>F1</i> (%)	<i>MIOU</i> (%)	<i>Kappa</i>
✓			82.85	-	-	-	-
✓	✓		85.26	99.37	77	82.74	0.77
✓	✓	✓	86.24	99.41	79.10	83.20	0.79

TABLE V
COMPARISON DIFFERENT SEMANTIC SEGMENTATION HEADS

	<i>AP</i> (%)	<i>OA</i> (%)	<i>F1</i> (%)	<i>MIoU</i> (%)	<i>Kappa</i>
rough segmentation head	81.06	99.28	75.2	81.65	0.75
fine segmentation head	85.26	99.37	77	82.74	0.77

3) *Comparison of Different Modules*: Different modules will have different impacts on the results. The effects of the modules proposed in this study on the oil spill detection accuracy are shown in Table IV. With the addition of modules, the accuracy of oil spill detection is further improved. After using the multifeature module, the interference of speckle noise and offshore targets on oil spill detection is reduced, and the localization evaluation metric *AP* is increased from 82.85% to 85.26%, which shows that the proposed feature is feasible and can effectively improve the accuracy of oil spill detection. Adding a semantic segmentation module not only realizes the segmentation of the oil spill area, but also enhances the accuracy of the localization box. Finally, the semantic complementation module enhances the localization and segmentation branches through the organic combination of the localization and semantic segmentation branches. MFSCNet model obtains *AP*, *OA*, *F1*, *MIOU*, and *Kappa* are 86.24%, 99.41%, 79.10%, 83.20%, and 0.79.

Considering the effect of coherent speckle noise in SAR images, two semantic segmentation heads shown in Fig. 3 are proposed in this study. The evaluation results are shown in Table V, and the segmentation results are displayed in Fig. 9. The proposed fine segmentation head offers better detection performance. It can be seen from Fig. 9 that the two modules are disturbed by the coherent speckle noise, which produces different phenomena. The rough segmentation head has poorer prediction results for the edge region. The expansion phenomenon occurs in the edge region, and the background area is predicted as the oil spill area. The misclassification region in Fig. 9(h) and (i) appear because the shallow features are not considered. For the fine segmentation head, the edge region is better predicted, and it is easier to distinguish the boundary area of the oil spill and the background. However, it will be affected by the background noise and the phenomenon indicated by the red boxes shown in Fig. 9(k) appears. Still, it has less impact on the overall semantic segmentation results. So, this study uses the fine segmentation head for the semantic segmentation task.

4) *Comparison of Different Hyperparameters*: Different values of the weighting factors of the loss function will affect the results, and experiments are conducted to find the appropriate weighting factors.

The results of semantic segmentation loss with various weighting factors are compared, and the implementation results are shown in Table VI. Because the presence of the localization

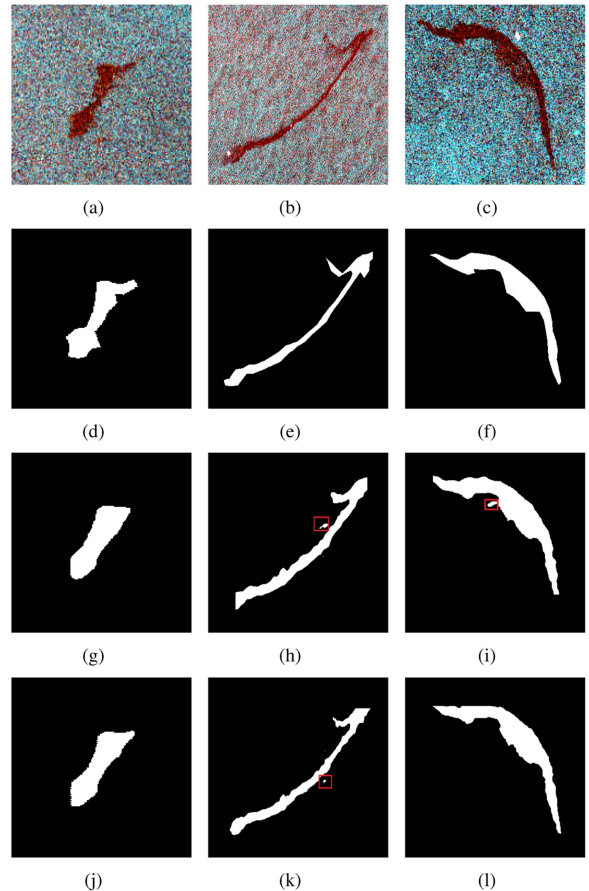


Fig. 9. Results of different semantic segmentation heads. (a)–(c) original image. (d)–(f) ground-truth map. (g)–(i) the results of rough segmentation head. (j)–(l) the results of fine segmentation head.

TABLE VI
COMPARISON OF DIFFERENT SEMANTIC SEGMENTATION LOSS WEIGHTING FACTORS

Weighting factor	<i>AP</i> (%)	<i>OA</i> (%)	<i>F1</i> (%)	<i>MIOU</i> (%)	<i>Kappa</i>
1	82.82	99.2	68.41	78.6	0.68
2	85.26	99.37	77	82.74	0.77
3	82.84	99.12	67.14	78.05	0.67
4	80.58	99.38	76.3	82.45	0.76
5	79.4	99.3	74.96	81.86	0.75

box limits the semantic segmentation results, the segmentation accuracy is affected by the localization accuracy. As the value of the weighting factors increases, both localization and semantic segmentation results show a trend of increasing and then decreasing. When the value is 2, the greatest accuracy for both localization and segmentation is achieved. Therefore, the weighting factor of the semantic segmentation loss function is set to 2.

The semantic complementation module uses the results of localization and semantic segmentation to produce a label map. Then, the watershed algorithm is used to obtain a complementation box for semantic complementation. After the loss calculation of the complementation box with the ground-truth box, updating the model parameters. However, not all localization boxes are subjected to semantic complementation, and the

TABLE VII
COMPARISON OF DIFFERENT THRESHOLDS IN SEMANTIC COMPLEMENTATION MODULE

Thresholds	AP(%)	OA(%)	FI(%)	MIoU(%)	Kappa
0.0	85.26	99.37	77.00	82.74	0.77
0.1	85.53	99.39	78.42	83.41	0.78
0.2	85.33	99.39	78.01	83.23	0.78
0.3	86.24	99.41	79.10	83.20	0.79
0.4	86.16	99.37	75.86	82.26	0.76
0.5	85.99	99.39	73.36	82.94	0.77
0.6	85.91	99.38	77.35	82.92	0.77
0.7	85.82	99.39	77.43	82.97	0.77
0.8	85.43	99.39	78.12	83.30	0.78
0.9	85.76	99.39	77.55	83.05	0.77
1.0	85.66	99.38	77.70	83.05	0.77

TABLE VIII
COMPARISON OF DIFFERENT SEMANTIC COMPLEMENTATION LOSS WEIGHTING FACTORS

Weighting factor	AP(%)	OA(%)	FI(%)	MIoU(%)	Kappa
0.1	85.50	99.38	77.73	83.08	0.77
0.5	85.53	99.39	78.06	83.25	0.78
1	86.24	99.41	79.10	83.20	0.79
2	86.02	99.38	77.49	83.05	0.77
3	85.81	99.38	77.40	82.97	0.77

intersection ratio between localization boxes and ground-truth boxes needs to be judged and different thresholds affect the number of localization boxes that perform complementation, so the results of varying threshold cases are compared. The results are shown in Table VII, and the highest accuracy is achieved when the threshold value is 0.3. Therefore, the threshold value for the semantic complementation module is determined to be 0.3. Although, the semantic complementation module is executed in the last period of the training phase. However, there is a specific error between the output of the training phase and the ground-truth, and the watershed algorithm cannot completely complement the oil spill region, which leads to the error between the final acquired complementation box and the ground-truth box as well. Therefore, it is crucial to choose a suitable loss function weighting factor. Then, considering the existence of errors, 0.1, 0.5, 1, 2, and 3 are selected for comparison. The results are shown in Table VIII, and the highest accuracy is achieved in localization and segmentation when the value is 1. As the value of the weighting factor increases, the accuracy gradually decreases. Still, all of them are higher than the accuracy when the module is not used. This indicates that the proposed semantic complementation module is practical and can improve localization and semantic segmentation accuracy. Therefore, the value of the loss function weighting factor of the semantic complementation module is determined to be 1.

V. CONCLUSION

This study proposes MFSCNet model to implement oil spill localization and semantic segmentation in the single neural network model. The multifeature input module eliminates the

long stripe interference of oil spills with the non-Gaussian distribution. A semantic complementation module is proposed to improve the accuracy of the localization box by combining the localization and semantic segmentation results. The best optimal parameters of the model are determined through hyperparameter experiments. MFSCNet is superior to other oil spill detection methods and more suitable for daily monitoring of oil spill events. In addition, how to distinguish oil spills from many kinds of look-alikes can be researched in the future.

APPENDIX

TABLE IX
THE STRUCTURE OF MFSCNET

Layer Name	Output Size	Structure
F1	512 × 512	3 × 3, 32
S2	256 × 256	$\begin{bmatrix} 3 \times 3, 64, \text{stride } 2 \\ 1 \times 1, 32 \\ 1 \times 1, 32 \\ 3 \times 3, 32 \\ 1 \times 1, 64 \end{bmatrix} \times 2$
T3	128 × 128	$\begin{bmatrix} 3 \times 3, 128, \text{stride } 2 \\ 1 \times 1, 64 \\ 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 128 \end{bmatrix} \times 3$
F4	64 × 64	$\begin{bmatrix} 3 \times 3, 256, \text{stride } 2 \\ 1 \times 1, 128 \\ 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
F5	32 × 32	$\begin{bmatrix} 3 \times 3, 512, \text{stride } 2 \\ 1 \times 1, 256 \\ 5 \times 5 \text{max pool, padding } 2 \\ 9 \times 9 \text{max pool, padding } 4 \\ 13 \times 13 \text{max pool, padding } 6 \\ 1 \times 1, 512 \\ 1 \times 1, 256 \\ 1 \times 1, 256 \\ 1 \times 1, 256 \end{bmatrix}$
DF5	32 × 32	$\begin{bmatrix} 1 \times 1, 512 \\ 1 \times 1, 256 \\ 1 \times 1, 256 \\ 1 \times 1, 256 \end{bmatrix}$
DF4	64 × 64	$\begin{bmatrix} \text{upsampling } 2 \\ 1 \times 1, 128 \\ 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 256 \\ 1 \times 1, 128 \end{bmatrix}$
DT3	128 × 128	$\begin{bmatrix} \text{upsampling } 2 \\ 1 \times 1, 64 \\ 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 128 \end{bmatrix}$
DS2	256 × 256	$\begin{bmatrix} \text{convTransposed2d } 2 \times 2, 64 \\ 3 \times 3, 128 \\ 3 \times 3, 64 \end{bmatrix}$
DF1	512 × 512	$\begin{bmatrix} \text{convTransposed2d } 2 \times 2, 32 \\ 3 \times 3, 64 \\ 3 \times 3, 32 \end{bmatrix}$
M	1024 × 1024	$\begin{bmatrix} \text{convTransposed2d } 2 \times 2, 32, \text{stride } 2 \\ 3 \times 3, 16 \\ 1 \times 1, \text{classes} \end{bmatrix}$

REFERENCES

- [1] Y. Dong, Y. Liu, C. Hu, I. R. MacDonald, and Y. Lu, "Chronic oiling in global oceans," *Science*, vol. 376, no. 6599, pp. 1300–1304, 2022.
- [2] E. Asihene et al., "Toward the detection of oil spills in newly formed sea ice using c-band multipolarization radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4302615, doi: 10.1109/TGRS.2021.3123908.

- [3] C. Quigley, C. Brekke, and T. Eltoft, "Comparison between dielectric inversion results from synthetic aperture radar co-and quad-polarimetric data via a polarimetric two-scale model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4200218, doi: [10.1109/TGRS.2020.3038366](https://doi.org/10.1109/TGRS.2020.3038366).
- [4] X. Kang, Z. Wang, P. Duan, and X. Wei, "The potential of hyperspectral image classification for oil spill mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538415, doi: [10.1109/TGRS.2022.3205966](https://doi.org/10.1109/TGRS.2022.3205966).
- [5] G. De Carolis, M. Adamo, and G. Pasquariello, "On the estimation of thickness of marine oil slicks from sun-glittered, near-infrared MERIS and MODIS imagery: The Lebanon oil spill case study," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 559–573, Jan. 2014.
- [6] H. Zhang et al., "Optical and SAR image dense registration using a robust deep optical flow framework," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1269–1294, 2023.
- [7] D. Xiang, Y. Xie, J. Cheng, Y. Xu, H. Zhang, and Y. Zheng, "Optical and SAR image registration based on feature decoupling network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5235913, doi: [10.1109/TGRS.2022.3211858](https://doi.org/10.1109/TGRS.2022.3211858).
- [8] A. Taravat, D. Latini, and F. Del Frate, "Fully automatic dark-spot detection from SAR imagery with the combination of nonadaptive weibull multiplicative model and pulse-coupled neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2427–2435, May 2014.
- [9] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, and I. Kompatsiaris, "Oil spill identification from satellite images using deep neural networks," *Remote Sens.*, vol. 11, no. 15, 2019, Art. no. 1762.
- [10] C. Brekke and A. H. Solberg, "Oil spill detection by satellite remote sensing," *Remote Sens. Environ.*, vol. 95, no. 1, pp. 1–13, 2005.
- [11] M. M. Espeseth, C. E. Jones, B. Holt, C. Brekke, and S. Skrunes, "Oil-spill-response-oriented information products derived from a rapid-repeat time series of SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3448–3461, 2020.
- [12] K. Li, J. Ouyang, H. Yu, Y. Xu, and J. Xu, "Overview of research on monitoring of marine oil spill," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 787, no. 1, 2021, Art. no. 012078.
- [13] K. Topouzelis and A. Pysillos, "Oil spill feature selection and classification using decision tree forest on SAR image data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 68, pp. 135–143, 2012.
- [14] L. W. Mdakane and W. Kleynhans, "Feature selection and classification of oil spill from vessels using Sentinel-1 wide-swath synthetic aperture radar data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4002505.
- [15] M. Migliaccio, A. Gambardella, and M. Tranfaglia, "SAR polarimetry to observe oil spills," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 506–511, Feb. 2007.
- [16] M. Migliaccio, F. Nunziata, and A. Buono, "SAR polarimetry for sea oil slick observation," *Int. J. Remote Sens.*, vol. 36, no. 12, pp. 3243–3273, 2015.
- [17] X. Ma, J. Xu, P. Wu, and P. Kong, "Oil spill detection based on deep convolutional neural networks using polarimetric scattering information from Sentinel-1 SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4204713, doi: [10.1109/TGRS.2021.3126175](https://doi.org/10.1109/TGRS.2021.3126175).
- [18] H. Ren, F. Li, B. Han, W. Hong, Y. Dong, and B. Wu, "Ocean oil spill classification with polarimetric SAR based on VGG16 multi-feature extraction," in *Proc. SAR Big Data Era*, 2021, pp. 1–13.
- [19] D. Song et al., "A novel marine oil spillage identification scheme based on convolution neural network feature extraction from fully polarimetric SAR imagery," *SAR Big Data Era*, vol. 8, pp. 59801–59820, 2020.
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.
- [21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
- [22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 1440–1448, 2015.
- [23] N. Davari, G. Akbarizadeh, and E. Mashhour, "Corona detection and power equipment classification based on Googlenet-alexnet: An accurate and intelligent defect detection model based on deep learning for power distribution lines," *IEEE Trans. Power Del.*, vol. 37, no. 4, pp. 2766–2774, Aug. 2022.
- [24] X. Wang, J. Zhou, and J. Fan, "IDUDL: Incremental double unsupervised deep learning model for marine aquaculture SAR images segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4209412, doi: [10.1109/TGRS.2022.3203071](https://doi.org/10.1109/TGRS.2022.3203071).
- [25] Z. Han, J. Xing, X. Wang, F. Xue, and J. Fan, "A robust LCSE-ResNet for marine man-made target classification based on optical remote sensing imagery," *Int. J. Artif. Intell. Tools*, vol. 31, no. 6, 2022, Art. no. 2240022.
- [26] J. Wang, J. Fan, and J. Wang, "MDOAU-Net: A lightweight and robust deep learning model for SAR image segmentation in aquaculture raft monitoring," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4504505.
- [27] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kaviani, "Ship classification in SAR images using a new hybrid CNN-MLP classifier," *J. Indian Soc. Remote Sens.*, vol. 47, pp. 551–562, 2019.
- [28] A. Raeisi, G. Akbarizadeh, and A. Mahmoudi, "Combined method of an efficient cuckoo search algorithm and nonnegative matrix factorization of different Zernike moment features for discrimination between oil spills and lookalikes in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4193–4205, Nov. 2018.
- [29] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "Osdes-Net: Oil spill detection based on efficient_shuffle network using synthetic aperture radar imagery," *Geocarto Int.*, vol. 37, no. 26, pp. 13539–13560, 2022.
- [30] L. De Laurentiis, C. E. Jones, B. Holt, G. Schiavon, and F. Del Frate, "Deep learning for mineral and biogenic oil slick classification with airborne synthetic aperture radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8455–8469, Oct. 2021.
- [31] F. M. Bianchi, M. M. Espeseth, and N. Borch, "Large-scale detection and categorization of oil spills from SAR images with deep learning," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2260.
- [32] K. Zeng and Y. Wang, "A deep convolutional neural network for oil spill detection from spaceborne SAR images," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 1015.
- [33] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "GreyWolfLSM: An accurate oil spill detection method based on level set method from synthetic aperture radar imagery," *Eur. J. Remote Sens.*, vol. 55, no. 1, pp. 181–198, 2022.
- [34] Y. J. Yang, S. Singha, and R. Mayerle, "Fully automated SAR based oil spill detection using YOLOv4," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 5303–5306.
- [35] X. Huang, B. Zhang, W. Perrie, Y. Lu, and C. Wang, "A novel deep learning method for marine oil spill detection from satellite synthetic aperture radar imagery," *Mar. Pollut. Bull.*, vol. 179, 2022, Art. no. 113666.
- [36] M. Nieto-Hidalgo, A. J. Gallego, P. Gil, and A. Pertusa, "Two-stage convolutional neural network for ship and spill detection using SLAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5217–5230, Sep. 2018.
- [37] H. Guo, G. Wei, and J. An, "Dark spot detection in SAR images of oil spill using segnet," *Appl. Sci.*, vol. 8, no. 12, 2018, Art. no. 2670.
- [38] X. Wang, J. Liu, S. Zhang, Q. Deng, and J. Fan, "Detection of oil spill using SAR imagery based on AlexNet model," *Comput. Intell. Neurosci.*, vol. 2021, 2021, Art. no. 4812979.
- [39] Q. Zhu et al., "Oil spill contextual and boundary-supervised detection network based on marine SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5213910.
- [40] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.
- [41] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLOACT: Real-time instance segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9157–9166.
- [42] S. Temitope Yekeen, A. Balogun, and K. B. Wan Yusof, "A novel deep learning instance segmentation model for automated marine oil spill detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 190–200, 2020.
- [43] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.
- [44] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
- [45] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding Yolo series in 2021," 2021, *arXiv:2107.08430*.
- [46] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7263–7271.
- [47] X. Sun, Y. Liu, Z. Yan, P. Wang, W. Diao, and K. Fu, "SRAF-Net: Shape robust anchor-free network for garbage dumps in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6154–6168, Jul. 2021.
- [48] A. Bochkovskiy, C. Wang, and H. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

- [49] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8759–8768.
- [50] S. Zhang, J. Xing, X. Wang, and J. Fan, "Improved YOLOX-S marine oil spill detection based on SAR images," in *Proc. 12th Int. Conf. Inf. Sci. Technol.*, 2022, pp. 184–187.
- [51] S. Naz, M. F. Iqbal, I. Mahmood, and M. Allam, "Marine oil spill detection using synthetic aperture radar over indian ocean," *Mar. Pollut. Bull.*, vol. 162, 2021, Art. no. 111921.
- [52] V. Corcione, A. Buono, F. Nunziata, and M. Migliaccio, "A sensitivity analysis on the spectral signatures of low-backscattering sea areas in Sentinel-1 SAR images," *Remote Sens.*, vol. 13, no. 6, 2021, Art. no. 1183.
- [53] A. Abimanyu, W. S. Pranowo, I. Faizal, N. K. Afandi, and N. P. Purba, "Reconstruction of oil spill trajectory in the java sea, Indonesia using SAR imagery," *Geogr., Environ., Sustainability*, vol. 14, no. 1, pp. 177–184, 2021.
- [54] I. A. El-Magd, M. Zakzouk, A. M. Abdulaziz, and E. M. Ali, "The potentiality of operational mapping of oil pollution in the mediterranean sea near the entrance of the suez canal using Sentinel-1 SAR data," *Remote Sens.*, vol. 12, no. 8, 2020, Art. no. 1352.



Jianchao Fan (Member, IEEE) received the B.S. and Ph.D. degrees in control science and engineering from the School of Control Science and Engineering, Dalian University of Technology, Dalian, China, in 2007 and 2012, respectively.

During 2013–2016, he was a Postdoctoral Fellow jointly cultivated by The Chinese University of Hong Kong and Dalian University of Technology, and during 2018–2019, he was a Visiting Scholar with the Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO,

USA. He has been a Professor with the Department of Ocean Remote Sensing, National Marine Environmental Monitoring Center, Dalian, China, since 2019. He has authored/coauthored three books and more than 90 scientific papers in international journals and conference proceedings, such as the IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition*, etc. His current research interests include marine SAR image processing, polarimetric SAR, deep learning neural networks, and heuristic swarm intelligent optimization.

Dr. Fan is a Program Chair, Workshop Chair, Publicity Chair, and Local Arrangement Chair of several international conferences, such as the International Symposium on Neural Networks, International Conference on Information Society and Technology, International Conference on Intelligent Control and Information Processing, etc.



Shuai Zhang received the B.S. degree in computer science and technology from Department of Computer Science and Technology, Xu Hai College of China University of Mining and Technology, Xuzhou, China, in 2020. He is currently working toward the M.S. degree in electronic and information with the School of Information Science and Engineering, Dalian Polytechnic University, Dalian, in 2023.

His main research interests include deep learning object detection and remote sensing image processing.



Xinzhe Wang received the Ph.D degree in control science and engineering from the School of Control Science and Engineering, Dalian University of Technology, Dalian, China, in 2012.

He is with the School of Information Science and Engineering, Dalian Polytechnic University, Dalian, China. His research interests include data driven modeling and deep learning neural networks.



Jun Xing received the B.Sc. and M.Sc. degrees in computer technology and application from Northeastern University, Boston, MA, USA, in 1996 and 1999, respectively. He received the Ph.D. degree in computer technology and application from Dalian University of Technology, Dalian, China, in 2008, respectively.

He is currently a Lecturer with the School of Information Science and Engineering, Dalian Polytechnic University, China. His research interests include research and application of large data algorithm.