DEPDet: A Cross-Spatial Multiscale Lightweight Network for Ship Detection of SAR Images in Complex Scenes

Jing Zhang, Fan Deng, Yonghua Wang, Jie Gong, Ziyang Liu, Wenjun Liu, Yinmei Zeng, and Zeqiang Chen

***Abstract*—Nowadays, the intricate nature of synthetic aperture radar (SAR) ship scenes, coupled with the presence of multiscale targets, poses a significant challenge in detection accuracy. Furthermore, to reduce the financial outlay on hardware, there is also a considerable challenge in lightweighting the model. In order to resolve the aforementioned concerns, we propose a cross-spatial multiscale lightweight network, designated as DEPDet. First, a new efficient multiscale detection backbone network DEMNet is redesigned. To improve the feature extraction capability of the network, a cross-spatial multiscale convolution (CSMSConv) is designed and a new CSMSConv module CSMSC2F is constructed. Meanwhile, we introduce an efficient multiscale attention module. DEMNet is capable of more effectively extracting information pertaining to multiscale ships. Moreover, to enhance the fusion of features at diverse scales, we design a new path aggregation feature pyramid network DEPAFPN, which combines deformable convolution and CSMSC2F. Finally, we introduce partial convolution to construct a lightweight detection head module PCHead, which can be employed to extract spatial features with greater efficiency. The publicly available SAR ship datasets, SAR Ship Detection Dataset and High-Resolution SAR Image Dataset, are employed for the purpose of conducting experiments. The mean average precision (mAP) obtained was 98.2% (+1.4%) and 91.6% (+1.6%), respectively. The F1 obtained 0.950 (+1.7%) and 0.871 (+1.4%), respectively. Concurrently, the Params decreased from 3.2M to 2.1M, a decrease of approximately 34%. The floating-point operations (FLOPs) decreased from 8.7G to 4.5G, a decrease of approximately 48%. The experimental results indicate that the network achieves an effective balance between detection accuracy and lightweight effect with good generalization and extensibility.**

***Index Terms*—Complex scenes, cross-spatial multiscale convolution (CSMSConv), lightweight network, multiscale ships, synthetic aperture radar (SAR).**

Received 5 August 2024; revised 7 September 2024; accepted 23 September 2024. Date of publication 26 September 2024; date of current version 15 October 2024. This work was supported in part by the Open Fund of National Engineering Research Center of Geographic Information System under Grant 2023KFJJ12 and in part by the Hubei Key Research and Development Program under Grant 2023DJC154. *(Corresponding author: Fan Deng.)*

Jing Zhang, Fan Deng, Ziyang Liu, Wenjun Liu, and Yinmei Zeng are with the School of Geosciences, Yangtze University, Wuhan 430100, China (e-mail: zhangjing.stu@yangtzeu.edu.cn; dengfan@yangtzeu.edu.cn; 2022710471@yangtzeu.edu.cn; 2022720559@yangtzeu.edu.cn; 2023720564@yangtzeu.edu.cn).

Yonghua Wang and Jie Gong are with the Wuhan Huaxin Lianchuang Technology Engineering, Company Ltd., Wuhan 430074, China (e-mail: wangyonghua@hxgis.com; gongjie@hxgis.com).

Zeqiang Chen is with the National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan 430074, China (e-mail: chenzeqiang@cug.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3469209

I. Introduction

SYNTHETIC aperture radar (SAR) is an active microwave imaging radar that provides all-day, all-weather observation capabilities, including the ability to penetrate clouds and ground, enabling it to capture ground feature information even in harsh weather and lighting conditions [1]. Consequently, it has been extensively employed in a multitude of fields, such as target reconstruction [2], target detection [3], and disaster and environmental monitoring [4]. In these applications, the use of SAR ship detection is of paramount scientific, technical, and operational significance in both military and civilian fields, including maritime transportation supervision and rescue, national defense and security, and fishing vessel monitoring [5], [6], [7], [8]. Nevertheless, the continuous improvement of SAR image resolution and the development of diversified ship sizes have posed greater challenges for the accurate detection of multiscale ships in complex scenes. At the same time, reducing the cost of hardware, increasing its availability, and deploying models on limited-resource devices have become the current pursuit [9]. This necessitates the development of innovative approaches to address these issues.

Lately, a variety of methods have been put forth with the aim of SAR ship detection. The aforementioned methods may be broadly classified into two distinct categories based on their feature extraction methods: traditional methods and deep learning methods.

Traditional SAR ship detection algorithms are mainly based on grayscale characteristics, whereas the constant false alarm rate (CFAR) [10] is the most commonly used algorithm based on grayscale features. By comparing the greyscale value of a single pixel with the discrimination threshold, it completes the detection of the target pixel. The discriminative threshold is determined based on the probabilistic properties of the backscatter, assuming a given false alarm rate. However, the uncertainty factor of background clutter is strong, especially in complex scenes where there are too many interference factors, resulting in low detection accuracy of the algorithm.

For the past few years, deep learning methods have found a wide range of applications in remote sensing, yielding promising results in tasks such as building change detection [11] and forest fire detection [12]. Concurrently, object detection algorithms based on deep learning performed exceptionally well in the computer science community, showcasing high detection accuracy,

advanced capabilities for extensibility, and the ability to learn features automatically, particularly in complex scenes [13], [14]. Consequently, these methods have been adapted for use in the detection of ships in SAR images [15].

SAR ship detection methods based on deep learning may be categorized according to their network architectures as either one-stage or two-stage. One-stage detection methods are characterized by the ability to accomplish object detection with a single feature extraction step, which results in faster processing in comparison to the two-stage. Representative algorithms in the category include the You Only Look Once (YOLO) [16] series and the Single Shot Multibox Detector (SSD) [17] series. Conversely, two-stage detection methods involve first locating candidate regions in the input image features and subsequently classifying these regions. This approach, exemplified by the earlier proposed region with convolutional neural network (R-CNN) [18] method and its derivatives such as Faster R-CNN [19] and Mask R-CNN [20], is more complex and computationally intensive, posing significant challenges for real-time detection and model deployment. Consequently, most researchers opt for one-stage detection algorithms for SAR ship detection.

However, despite their speed, one-stage detection algorithms still exhibit shortcomings in detection accuracy, particularly concerning small- and medium-sized ships in complex scenes [21]. Thus, existing SAR ship detection methods struggle to give attention to both detection accuracy and lightweight effect. SAR ship detection continues to confront challenges in achieving this balance, including the following.

1) Ship scenes are complex. In SAR images, the real-time captured ships are not just simple images of ships on the sea but more often in complex scenes such as rivers, ports, and islands, which affect the accurate detection of ships.
2) Multiscale ships. The scale of ships is diverse, in particular with a higher proportion of small- and medium-sized dense ships.
3) Models are not lightweight enough. The majority of ship detection methods exhibit high computational complexity and memory consumption, which presents significant challenges in their deployment on resource-limited hardware devices. This, in turn, increases the hardware cost and markedly reduces the usability and ease of use of the model.

In order to overcome all of these aforementioned challenges, DEPDet is proposed in this article, a cross-spatial multiscale lightweight network. The essential dedications made by the algorithm are presented as follows.

1) A new efficient multiscale detection backbone network DEMNet is redesigned. To tackle the multiscale problem, we design CSMSConv and construct a new CSMSConv module CSMSC2F, which better extracts and fuses the details and metaphorical features of multiscale ships, providing increased ship detection capability. In addition, we introduce efficient multiscale attention (EMA) mechanisms, which improve the performance of the model in terms of feature extraction, in particular with diminutive dense ships in complex scenes. It effectively suppresses the interference of complex backgrounds and highlights ship targets.
2) We design a new feature fusion module DEPAFPN that fuses deformable convolutions, enhancing the merging of feature information at different levels of detail.
3) A lightweight detection head module PCHead is constructed by introducing partial convolution (PConv), which can abstract spatial information more effectively through a concerted effort to reduce duplicated computations and storage operations.
4) As a basis for evaluating the effectiveness of DEPDet, comprehensive comparative and ablation experiments are conducted on two challenging public datasets, SAR Ship Detection Dataset (SSDD) and High-Resolution SAR Image Dataset (HRSID). Our model demonstrated satisfactory performance with regard to both detection accuracy and computational complexity in both datasets. To further evaluate the lightweight performance of the model, we deployed it on a resource-limited device for experiments. Generalizability experiments and extensibility experiments are also conducted, and our model can be applied to a large remote sensing scene for ship detection and can be extended to multispectral ship detection tasks. Our model has good ease of use, generalization, and extensibility.

The rest of this article is organized as follows. In Section II, we provide a detailed introduction to some related work. The overall architecture and details of DEPDet are presented in Section III. Experimental results are presented in Section IV. We present a discussion of our model and delineate future work directions in Section V. Section VI is devoted to the drawing of conclusions.

II. Related Work

A. Traditional SAR Ship Detection Methods

Most of the traditional methods are improved algorithms based on the CFAR, such as GO-CFAR [22], SO-CFAR [23], and CA-CFAR [24]. In addition, in order to enhance the detection accuracy, a number of CFAR detection methods have also emerged in combination with other methods [25], [26], [27].

Li et al. [25] developed an adaptive superpixel level CFAR detector for dense ship detection tasks in nearshore areas. This method proposes a nonlocal superpixel topology strategy that has the capacity of distinguishing between pure clutter superpixels and mixture superpixels, adaptively determining an adequate quantity of pure superpixels for taking stock of detection thresholds. Ai et al. [26] proposed AIS-RCFAR. With the assistance of AIS data, this method has great improvement in both uniform and nonuniform multitarget environments and has better detection speed, but for particularly complex scenarios, this article does not make corresponding experimental evaluation. Madjidi and Laroussi [27] proposed an improved algorithm based on the CFAR, a bilateral automatic review method based on an approximate maximum likelihood estimation. This method demonstrates superior performance compared to any other CFAR improvements in the field of simple maritime

scenes, but the algorithm is not applicable to scenes with complex backgrounds.

In general, traditional methods still rely on the setting of parameters. Even though many researchers are carrying out the research of adaptive parameters with the objective of enhancing the accuracy of detection, the majority of application scenarios are still limited to simple scenes of ships on the sea, and there is still a certain degree of difficulty to use it in complex scenes, especially in the remote sensing large scene maps. Therefore, this type of detection method has weak generalization ability and poor universality.

B. SAR Ship Detection Methods Based on Deep Learning

Due to the high-accuracy advantage of two-stage target algorithms, some researchers have begun to attempt to introduce two-stage algorithms into the field of SAR ship detection. ISAS-DNet [28] was proposed in 2021, which is a two-stage detector with two branches drawing inspiration from Mask R-CNN. For the multiscale problem of ships, it designs a global reasoning module and a mask-assisted ship detection module to interact with the information of the two branches, thereby enhancing the detection results. Tang et al. [29] proposed PPA-Net. The design strategy of the model is to enhance its ability to learn multiscale ship features by introducing an attention mechanism, spatial pooling, and feature balancing methods. Lv et al. [30] proposed SRDet, which is an anchor-free SAR ship detection method. This algorithm uses the two-stage method CenterNet2 [31] as a de facto standard. In addition, it develops a novel hybrid domain attention mechanism, which effectively suppresses the impact of complex land backgrounds.

Due to a simpler structure and stronger real-time performance of the one-stage detector, more researchers choose one-stage object detection algorithms as the baseline model. For example, Zhang et al. [32] proposed the introduction of LFO-Net, which originated from the SSD. The approach focuses on using a new bidirectional feature fusion module, comprising a semantic aggregation module and a feature reuse module, with the objective of enhancing the performance of multiscale object detection. This is achieved by optimizing the features of low and high feature layers. Zhao et al. [33] proposed D2ADet, improving by RetinaNet [34]. Specifically, it incorporates an enhanced feature pyramid network (FPN) that leverages its inherent strengths to more effectively and adaptively acquire superior ship features.

In response to the challenge of multiscale ship detection in complex scenes, Zhou et al. [35] presented HRLE-SARDet, which is a lightweight SAR target detection algorithm. It draws inspiration from the design concept of YOLOv5, a lightweight scattering feature extraction backbone and a lightweight hybrid representation learning enhancement module has been redesigned. It has proven effective in terms of both model accuracy and its lightweight design. In addition, Zhang et al. [36] proposed ESarDet. It adopts YOLOX [37] as the baseline model. The authors propose the CAA-Net with a larger ERF to facilitate the integration of contextual and semantic information pertaining to ships, and design a new convolutional block A2CSPlayer and a new spatial pyramid pooling network A2SPPF to heighten the fusion of feature images across distinct scales. ESarDet has achieved good results in accuracy and speed. In light of the recent advances in object detection algorithms, CSD-YOLO [38] based on improved YOLOv7 [39] was put forth for consideration. Its principal innovation is to propose SAS-FPN that combines spatial pyramid pooling and attention transfer to enable the model to focus on the most pertinent information regarding ships. Meanwhile, it integrates different scales ship features, thereby improving detection accuracy. It is experimentally verified that CSD-YOLO performs well in overall performance. In addition, Sun et al. [40] and Xiong et al. [41] designed a rotating ship detection model so that it can detect ships with arbitrary heading, which greatly enhances the accuracy of ship detection.

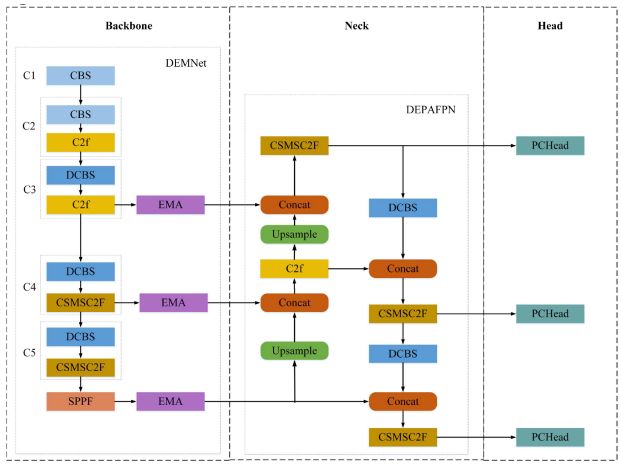

Fig. 1. Overall structure of DEPDet.

III. Method

A. Overall Network Structure

Addressing the complexity of SAR ship backgrounds and multiscale targets of SAR ships, while also considering the industrialized needs of easier deployment, we put forward a cross-spatial multiscale lightweight network named DEPDet. Fig. 1 shows its overall network structure. We choose the mature and efficient anchor-free real-time detection algorithm YOLOv8 as the reference model, and redesign the backbone, neck, and detection head. We design a new CSMSConv to replace part of the traditional convolution, so as to construct a new convolution module CSMSC2F. In backbone and neck of the model, the deformable convolution and the CSMSC2F are fused to achieve more efficient feature extraction and feature fusion. Furthermore, in addition to solving the problem of complex scenes, the backbone component incorporates the EMA mechanism to enhance the saliency of the ships. In the meantime, we also introduce the PConv to design a lighter and more efficient detection head.

YOLOv8 represents the most mature iteration of the YOLO algorithm and is currently being employed extensively in object detection [42], [43]. Our model is based on the framework structure of Yolov8, with certain components of Yolov8 retained

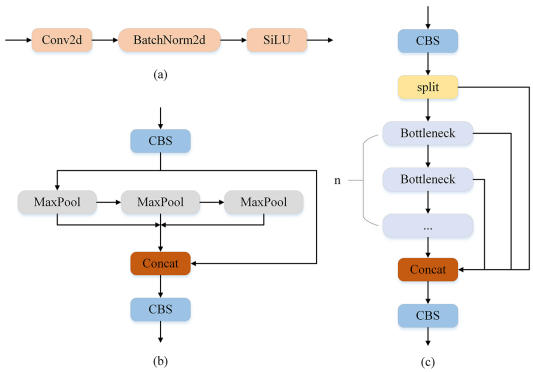

Fig. 2. Specific process structure diagrams. (a) CBS structure diagram. (b) SPPF structure diagram. (c) C2f structure diagram.

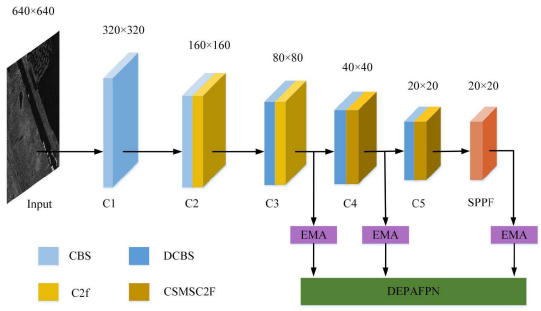

Fig. 3. Overall module structure of DEMNet.

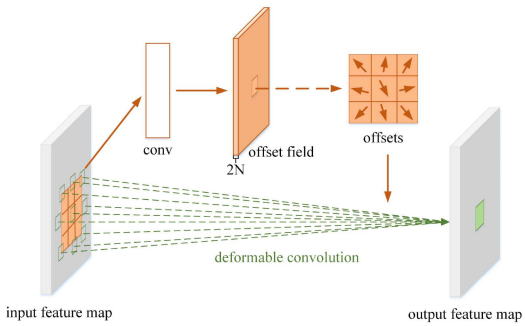

Fig. 4. Feature extraction of the deformable convolution.

in the backbone and neck network parts: the CBS, C2f, and SPPF. Fig. 2 shows the specific process structure of each module. The role of the CBS module is to extract the feature maps. The role of the C2f structure is to obtain richer information about the gradient flow. Finally, the role of the SPPF module is to fuse the different scales of the features. The role of the backbone is to extract features. The neck primarily performs feature fusion and employs the path aggregation network (PAN) [44] and the FPN [45] structures, which enhance the ability to express and localize semantic information across multiple scales. Head partially replaces the current mainstream decoupled head structure and separates the classification and detection heads, which solves the issue of the distinct focuses of classification and positioning.

B. DEMNet

The composition framework of DEMNet adopts the cross-stage partial network (CSPDarkNet53), mainly using information extracted from C3, C4, and C5 modules to create features and classify things. It may optimize the network's learning capacity while reducing memory costs. Nevertheless, multiscale SAR ship detection in complex scenes represents a significant challenge. Therefore, when we use simple traditional convolutional networks for feature extraction, it is easy to lose important information. To solve the problem of complex scenes, we introduced deformable convolution to enhance the edge feature extraction of ships. In addition, to address the issue of multiscale, we designed the CSMSConv and constructed a new convolution module CSMSC2F to enhance feature extraction at different scales by introducing an EMA attention mechanism to enhance feature fusion. Fig. 3 shows the overall module structure of DEMNet.

The core of DEMNet consists of six stages. First, C1 and C2 layers are located in the shallow region of the network, and we used traditional convolution to obtain high-resolution feature information. Then, we added deformable convolutions to C3, C4, and C5 layers instead of traditional convolutions to enhance ship edge features and obtain richer detailed information. In addition, due to the low resolution of C4 and C5 layers, traditional convolution can cause many small ships to lose some or all of their detailed features. Therefore, the CSMSConv was added to C4 and C5 layers to reconstruct the convolution module CSMSC2F. CSMSC2F facilitates the acquisition of feature information via disparate receiving fields, thereby enhancing the semantic information available to the network. Finally, with the aim of improving the efficacy of feature fusion, EMA attention mechanism was added to enhance feature fusion at different scales before transmitting the output feature information to the network neck DEPAFPN in C3, C4, and SPPF layers.

1) Deformable Convolution: Ships exhibit irregular shapes and sizes in SAR images, accompanied by intricate edge details. The ordinary convolution operation is typically fixed in size, which lacks the flexibility required for the shape perception field of ships, thereby diminishing the model's ability to recognize objects with significant deformation. To address this issue, we employ deformable convolution to enhance feature extraction, enabling adjustment of the target's shape based on its actual configuration for extracting feature images. In comparison to ordinary convolution, deformable convolution incorporates learnable offsets into the receptive field. These offsets can adjust the sampling position of the convolution kernel on the feature map, rendering the receptive field no longer fixed in size but closely aligned with the actual shape of the object. In Fig. 4, on the input feature map, the ordinary convolution operation corresponds to a convolutional sampling region of a square with the size of a convolutional kernel (orange box), while the deformable convolution corresponds to a convolutional sampling region of some points indicated by green boxes. This illustrates the difference between deformable convolution and ordinary convolution.

Fig. 4 provides a detailed illustration of the feature extraction process using deformable convolution. First, the input image is

subjected to processing in order to extract the feature maps, with the use of conventional convolution kernels. Then, in order to obtain the offset of deformations of the deformable convolution, we take the obtained feature maps as inputs and add an additional convolution layer to the feature maps. This convolution is not the same as the final convolution operation to be performed. The *N* in the figure refers to the number of channels within the input feature. In our proposed algorithm, we use a $3 * 3$ convolutional kernel, resulting in $N = 9$. The orange process in the figure represents the convolution process for offset learning, where the dimension of channels of the offset field is 2*N*, indicating that the convolutional kernel learns the offset in the x-direction and y-direction, respectively. Finally, during the training phase, convolution kernel employed in generating output features and those involved in generating offsets are synchronously learned, where the learning of offset is achieved through backpropagation using interpolation algorithms.

The deformable convolution formula is presented as

$$y\ (p_0) = \sum_{p_n \in R} w\,(p_n)\ \cdot x\,(p_0 + p_n + \Delta p_n) \tag{1}$$

where p_0 represents the input feature image pixel position, R represents the relational proximity of other pixel points in relation to the central pixel point, p_n represents the relative position in R with respect to p_0, Δp_n represents the offset, x symbolizes the input feature map, y symbolizes the output feature map, besides, and w symbolizes the weight of the sampling position.

However, since the offset Δp_n is usually a fraction, the image value of the input feature map x cannot be obtained directly, and the backpropagation cannot be performed if the offset is rounded directly, and it is necessary to perform a bilinear interpolation, the expression of which is denoted as

$$x\ (p) = \sum_{q} G\,(q, p)\ \cdot x\,(q) \tag{2}$$

where $G(q, p)$ is the bilinear interpolation kernel function, q symbolizes the total integral spatial positions within the map of features, and p represents the sum of p_0, p_n, and Δp_n.

2) CSMSC2F: To dispose of the challenge of SAR multiscale ships, we designed the CSMSConv and constructed a new convolution module CSMSC2F to augment feature extraction across various scales. The CSMSConv draws inspiration from grouped convolution and pointwise convolution. Nonetheless, it is not equivalent to grouped convolution or pointwise convolution. It cleverly utilizes interchannel information exchange to achieve the combination of detailed positional information from low-order feature maps and rich semantic information from high-order feature maps while reducing network parameters. This better avoids the situation of redundant information in feature maps and thus enhances the detection capabilities.

The function of C2f structure is mainly to enrich the gradient flow information, but the existence of traditional convolution and residual structures incurs substantial computational costs for the model, and the capacity of extracting multiscale ship features, especially small- and medium-sized ship features, is limited. Consequently, we integrate CSMSConv into the C2f structure and reconstruct a new CSMSConv module CSMSC2F.

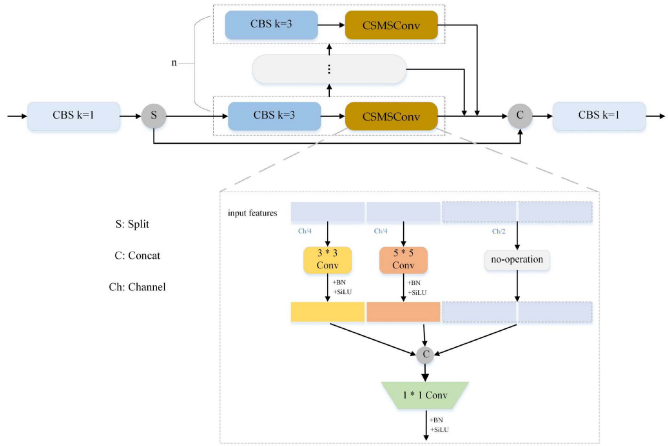

Fig. 5. Specific implementation process of CSMSConv and CSMSC2F.

CSMSC2F achieves the enrichment of the interaction of multiscale gradient flow information while mitigating excessive redundancy in channel information.

Fig. 5 illustrates the specific implementation process of CSMSC2F and CSMSConv. The feature extraction of CSMSC2F mainly relies on multibranch gradient flow. During the iterative process of the gradient flow information, CSMSConv is designed to facilitate the extraction and fusion of multiscale ship features. The design idea of CSMSConv is as follows.

1) Initially, the input feature map is partitioned into four subchannels, each of which is defined by the channel dimension. Half of them do not perform any operations, while the other half use convolution kernels of different sizes to obtain perceptual fields at different scales. Given that diminutive ships occupy the majority of the ship detection task, we have opted for two types of convolution kernels: $3 * 3$ and $5 * 5$. One-fourth of the subchannels employ a $3 * 3$ convolution operation to extract features, while the remaining one-fourth utilize a $5 * 5$ convolution operation to extract features. These two subchannels facilitate the acquisition of target information at varying sizes and scales.
2) Subsequently, after BN and SiLU operations, they are restored to the original dimensional order.
3) Ultimately, the attributes of all subchannels are integrated and merged, and the number of channels is reinstated through the application of $1 * 1$ convolution, BN, and SiLU operations, thereby generating a multiscale feature map with a size equivalent to that of the original feature map.

3) Efficient Multiscale Attention: The complexity of scenes and the presence of multiscale ships present a significant challenge to the accuracy. In order to overcome the aforementioned obstacles, we introduce an attention mechanism allowing for the dynamic reallocation of the model's weight to each region within the image before transmitting the feature information to the neck network in C3, C4, and SPPF output feature layers

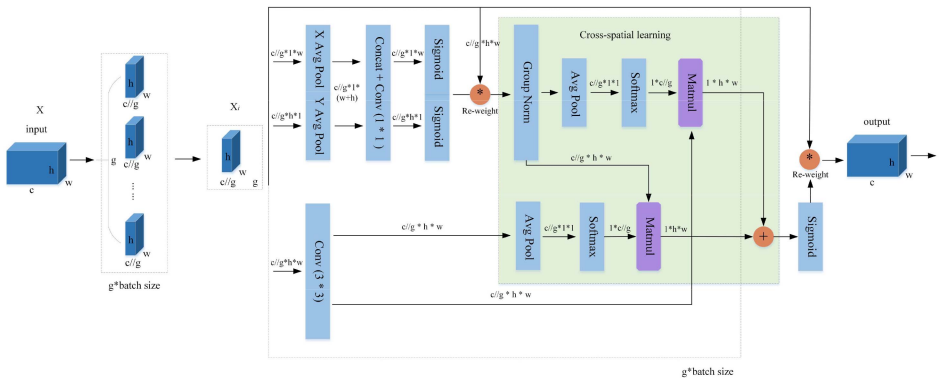

Fig. 6. Overall module structure of EMA.

of the backbone part. This enhances the attention of the model to the target area and improves the capability of the detection model to extract multiscale ships in complex scenes.

Presently, numerous scholars have investigated attention mechanisms. For example, the SE [46] attention mechanism learns the weight of each channel through global average pooling and fully connected layer, a procedure that can significantly enhance the representation and generalization capacity, but it lacks the understanding of spatial information. Conversely, the CA [47] attention mechanism is concerned with the relationship between pixels within the feature map. It obtains the spatial remote interaction relationship by calculating the coordinate information and the relationship between the feature values but the significance of the entire spatial location interaction is ignored.

EMA [48] mechanism employs parallel substructures to reduce network depth and furnish more suitable attention information for sophisticated features without reducing channel dimensions. EMA not only adjusts the channel weights of parallel subnetworks by encoding global information but also fuses the output of the above by establishing cross-dimensional interaction. Fig. 6 shows the overall module structure of EMA. The g represents the divided groups, X Avg Pool stands for 1-D horizontal global pooling, and Y Avg Pool stands for 1-D vertical global pooling in the figure.

First, for the input feature $X \epsilon R^{c*h*w}$, it is categorized to g subfeatures based on the number of channels to learn different semantics, where the input feature $X = [X_0, X_i, \ldots, X_{g-1}]$, $X_i \epsilon R^{c//g*h*w}$. Let us assume that g is much smaller than c without losing generality, and assume that the learned weights will be used to result in an improvement in the focus area of the subfeatures.

Then, EMA employs two parallel paths on the $1 * 1$ branch and one parallel path on the $3 * 3$ branch to extract the figured weights of filtered feature maps. In $1 * 1$ branch, X and Y Avg Pool are employed for the purpose of encoding spatial channel information in two distinct directions; besides, two encoded features are connected. These will keep it from dimensionality reduction on the $1 * 1$ branch. Furthermore, the output is decomposed into two vectors and subsequently mapped using two sigmoid functions; besides, finally a multiplicative aggregation of the channel attention feature maps to facilitate distinct cross-channel interactions between two parallel paths. In $3 * 3$ branch, one $3 * 3$ convolution is employed for the purpose of capturing multiscale feature representations.

Finally, a cross-spatial learning method is employed in order to obtain a more extensive and detailed feature aggregation. The green part in Fig. 6 represents the process of cross-spatial learning. It achieves cross-spatial channel information aggregation of feature maps with different spatial dimension directions. First, Avg Pool function is applied to the output of $1 * 1$ branch with the intention of encoding the global spatial information. Output results are transformed into the corresponding dimensional shape of $R^{1*c//g}$ by fitting a linear transformation with the nonlinear function softmax, while the spatial feature information of the $3 * 3$ branch is passed in at the same time with the dimensionality of $R^{c//g*h*w}$. The results of the aforementioned two branches are subsequently matrix-dot multiplied, thus yielding the initial set of spatial attention maps. In the same way, Avg Pool function is applied to the output of $3 * 3$ branch with the intention of encoding the global spatial information. The output undergoes a transformation to assume a corresponding dimensional shape of $R^{1*c//g}$ through softmax fitting linear transformation, and the spatial feature information of $1 * 1$ branch is passed in with a dimension of $R^{c//g*h*w}$. The results of the aforementioned two branches are subsequently matrix-dot multiplied, thus yielding the second set of spatial attention maps. In this way, the two sets of spatial attention maps exhibit a uniform scale, with dimensions of R^{1*h*w}. Subsequently, the outputs for each group are calculated as a function of the corresponding weight values, which are then fitted using sigmoid. Different scales of spatial information interaction are achieved through multiplication aggregation.

C. DEPAFPN

DEPDet adopts the PAN-FPN method framework in the neck part, which transfers the deep and shallow information to each other, achieves the unification of parameters, and improves the effective extraction of target deep semantic details and superficial positional details by the model. In the domain of the SAR ship detection task, effective fusion of disparate

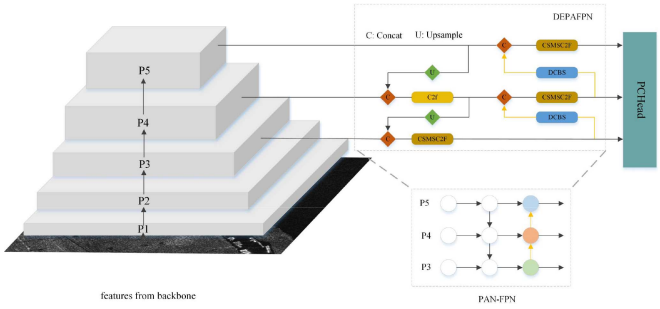

Fig. 7. Overall architecture of DEPAFPN.

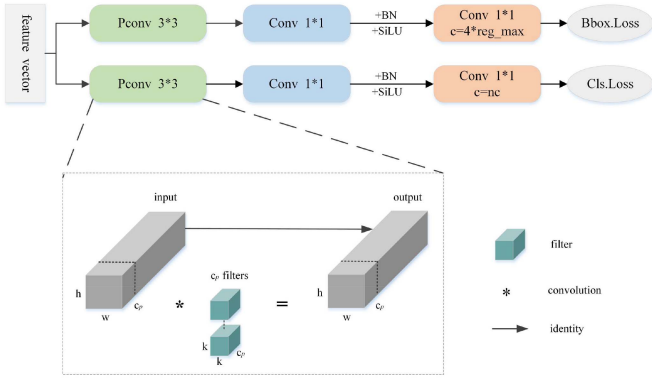

Fig. 8. Overall architecture of PCHead.

scale features extracted from the backbone represents a pivotal challenge. Therefore, we improve the structure of part of the neck and design a method framework DEPAFPN that integrates the deformable convolution and the CSMSC2F. Fig. 7 shows the overall architecture. DEPAFPN effectively merges feature maps derived from disparate scales.

DEPAFPN mainly consists of modules such as DCBS, C2F, and CSMSC2F, Upsample. It directly concatenates P3, P4, and P5 shallow information output from C3, C4, and SPPF layers of the backbone with the extracted deep information so that the model output results contain shallow position information. Since the upsample of the neck is an upsampling method, it inserts more elements between the individual image pixels, thereby enlarging the image and facilitating the acquisition of greater image information. In addition, the DCBS and the CSMSC2F modules expand the receptive field, which can provide powerful feature extraction from tiny ships scattered across multiple channels.

D. PCHead

PCHead uses three detection heads with different scales to detect targets from feature maps at different levels. Each detection head is tasked with the prediction of a range of bounding boxes, along with the respective class probability and target confidence associated with each box. However, with an increase in the number of layers in the convolutional neural network, semantic information of feature maps is gradually extracted and aggregated, resulting in deep feature maps often containing much similar information. In addition, due to the weight sharing mechanism of convolutional layers, different positions of the deep feature maps share convolutional kernel parameters, resulting in redundant feature map information, which in turn leads to the complexity of model computation. For the reason that reduces the complexity of the model, we construct a new lightweight detection head PCHead using the PConv, which is the convolution used in FasterNet [49]. Its working principle is shown in Fig. 8. By reducing redundant calculations and simultaneously storing access, spatial features can be extracted more effectively. It achieves higher operating speed than alternative neural networks while maintaining the desired levels of target task accuracy.

PConv employs conventional convolution operations on some subsets of input channels, whereas the residual channels undergo no modification. The computational complexity and memory access situation of PConv are as follows:

$$h \times w \times k^2 \times c_p^2 \tag{3}$$

$$h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \tag{4}$$

where the width and height of the feature map, denoted by h and w, respectively, k represents the size of the convolution kernel, besides, c_p represents the number of channels utilized for conventional convolution. As a matter of fact, there is mostly $r = c_p / c = 1/4$, resulting in a computational cost for PConv that is only $1/16$ of that for conventional convolution. It has been established that the memory access required of a PConv operation is only $1/4$ that of a conventional convolutional process, allowing for greater efficiency in data processing. Besides, it can be observed that the rest of $c - c_p$ channels are not implicated in the aforementioned calculation; thus, there is no necessity for any kind of memory access.

From this, it can be seen that introducing the PConv operator into the detection head can considerably alleviate computational complexity and memory access, thereby making the detection head network lightweight and accelerating the inference speed of the model.

TABLE I
CONFIGURATION OF THE EXPERIMENTAL ENVIRONMENT

Configuration	Parameter
CPU	Intel(R) Xeon(R) Gold 6330
RAM	80GB
GPU	NVIDIA GeForce RTX 3090 24GB
Operating system	Ubuntu20.04
Development tools	PyTorch2.0.0; Python3.8; CUDA11.8

IV. EXPERIMENTS AND RESULTS

A. Experimental Environment and Details

Table I illustrates the configuration of the experimental environment. All comparative and ablation experiments are executed on the same platform. In all experiments, the same parameter settings were employed for training a total of 200 epochs on the datasets, with a batch size of 32 and an initial learning rate of 0.01. In addition, the stochastic gradient descent optimizer was

TABLE II
DETAILS OF THE SSDD AND THE HRSID

Dataset	Date	Source	Resolution	Image Size	Images/Ships
SSDD	1 December 2017	Sentinel-1 RadarSat-2 TerraSAR	1 m–15 m	190–668	1160/2456
HRSID	29 June 2020	Sentinel-1 TerraSAR	0.5 m, 1 m, 3 m	800 × 800	5604/16,951

employed in conjunction with the weight decay of 0.0005 as well as the optimizer momentum of 0.937.

B. Datasets

So as to validate the efficacy and effectiveness of DEPDet, two datasets were selected for the purpose of experimentation: the SSDD [50] and the HRSID [51]. In experiment, the datasets are partitioned into three distinct subsets: training, validation, and testing. These subsets are allocated a ratio of 7:1:2, respectively.

1) SSDD: The SSDD, released in 2017, was the first publicly available SAR ship detection dataset. It has become the standard against which researchers assess and train their algorithms [50]. The SSDD contains a total of 1160 images and 2456 ships. In addition, the resolution range of the SSDD is from 1 to 15 m, with polarizations of HH, HV, VV, and VH, and it is derived from the combined data of three satellites: RadarSat-2, TerraSAR-X, and Sentinel-1. The SSDD covers locations such as Yantai in China and Vishakhapatnam in India. It contains diverse ships, including multiscale SAR ships captured from different scenes by different sensors with different resolutions under different polarizations. This includes small ships with complex scenes, as well as the dense distribution of ships near ports.

2) HRSID: The HRSID, a high-resolution SAR image dataset, is utilized for ship detection. It is a publicly available dataset released by Wei et al. [51] in January 2020. It has a richer representation of ship features, consisting of 99 Sentinel-1 images, 36 TerraSAR-X images, and 1 TanDEM-X image. The HRSID comprises 5604 images and 16 951 ships. The resolution of the HRSID is 0.5 m, 1 m, and 3 m. This dataset covers cities such as Houston in the United States, São Paulo in Brazil, the Aswan Dam in Egypt, and Shanghai in China. It includes multiple complex scenes such as nearshore, ports, and islands, as well as a richer variety of multiscale target ships.

Table II lists the details of the SSDD and the HRSID, including date, source, resolution, image size, image number, and ship number.

In addition, we calculate the pixel sizes of all ships in two datasets and display diverse ship sizes in the form of scatter plots. Fig. 9 illustrates that both datasets consist of multiscale ships with significantly large spans, predominantly comprising diminutive ships. This makes the multiscale ships detection capability of our trained DEPDet model possible.

C. Evaluation Metrics

The precision (P), recall (R), mean average precision (mAP), and F1 score (F1), which are commonly used in the field of target detection, are selected as the evaluation metrics of the model

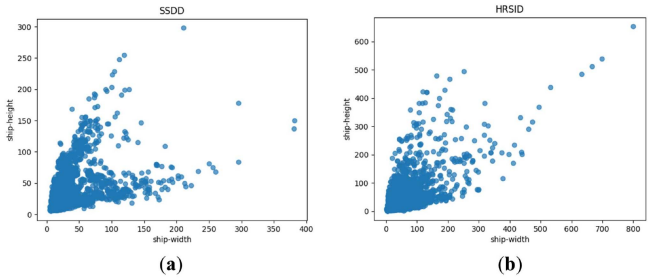

Fig. 9. Diverse ship sizes distribution. (a) SSDD. (b) HRSID.

detection performance. In addition, this article also quotes the parameters (Params) and the floating-point operations (FLOPs) to evaluate the complexity of the model. They are calculated as follows:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$\mathrm{AP} = \int_0^1 P\left(R\right) dR \tag{7}$$

$$\mathrm{mAP} = \frac{1}{N} \sum_{k\,=\,1}^{N} \mathrm{AP}\left(k\right) \tag{8}$$

$$\mathrm{F1} = 2 \times \frac{\mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}}. \tag{9}$$

TP stands for the number of instances in which the predicted and actual results are true, indicating that the detected ship is a true ship; FP stands for the number of instances in which the predicted result is true but the actual result is false, indicating that the detected ship is a fake ship; FN stands for the number of instances in which the predicted result is false but the actual result is true, that is, the real ship is not detected. Thus, P stands for the proportion of genuine ship targets among all detected ship target samples, while R stands for the percentage of accurately identified ship samples among real ship targets.

Generally, in order to provide a comprehensive evaluation of the model's detection accuracy, a plot of the P–R curve is required. The area between the curve and the coordinate axis is the average precision (AP). AP metric assesses the efficacy of the model across distinct categories. The mAP is taken as the average precision of the AP across all categories, providing an overall assessment of the model's performance across all categories.

In addition, due to the interdependent relationship between P and R, we introduce F1 to characterize the model's ability to have both good P and R.

$$\mathrm{Params} = \left(K_h K_w C_{\mathrm{in}}\right)\ C_{\mathrm{out}} \tag{10}$$

$$\mathrm{FLOPs} = 2\mathrm{HW}\left(C_{\mathrm{in}} K^2 + 1\right) C_{\mathrm{out}}. \tag{11}$$

The lightweight of the model is one of the important goals we aim to achieve. Params metric is employed to quantify the spatial complexity of the model, which mainly depends on the

TABLE III
COMPARISON WITH CLASSICAL OBJECT DETECTION MODELS ON SSDD

Method	P (%)	R (%)	mAP (%)	F1	Params (M)	FLOPs (G)
Faster R-CNN	90.9	87.9	91.2	0.894	41.1	134.4
SSD	90.1	87.6	92.7	0.888	24.3	87.7
YOLOv5	91.9	90.4	96.1	0.911	7.2	16.7
YOLOv8n	95.2	91.5	96.8	0.933	3.2	8.7
DEPDet (Ours)	**97.9**	**92.3**	**98.2**	**0.950**	**2.1**	**4.5**

The bold values indicate the best results.

dimensions of the convolution kernel and the number of channels in the input and output feature maps. In addition, FLOPs metric is employed to ascertain the time complexity associated with a given model, which mainly depends on the dimensions of the input feature map, the number of channels for the input and output feature maps, the dimensions of the convolution, and the number of convolution layers.

D. Experimental Results

1) Comparison Experiments on SSDD: SSDD represents the de facto standard for the SAR ship detection task, which is the most prevalent and authoritative in the field of SAR ship detection [15]. In order to ascertain the efficacy of DEPDet, we initially conduct comparative experiments with some excellent classical object detection algorithms. Table III presents the results of the comparative experimental study. We chose the classical two-stage algorithm Faster R-CNN, and the classical one-stage algorithms SSD, YOLOv5, and the baseline model YOLOv8n for comparison to ascertain the efficacy of DEPDet.

DEPDet demonstrated the most optimal performance on mAP, F1, Params, and FLOPs, with respective values of 98.2%, 0.950, 2.1M, and 4.5G. Compared with Faster R-CNN, DEPDet improves mAP and F1 by 7% and 5.6%, respectively, while our Params and FLOPs almost decrease by about 95% and 97%. Compared with SSD, DEPDet improves mAP and F1 by 5.5% and 6.2%, respectively, while Params and FLOPs decrease by about 91% and 95%, respectively. Compared with YOLOv5, mAP and F1 increase by 2.1% and 3.9% respectively, while Params and FLOPs decrease by about 71% and 73%, respectively. Compared with YOLOv8n, mAP and F1 increase by 1.4% and 1.7%, respectively, while Params and FLOPs decrease by about 34% and 48%, respectively. The results of SSDD indicate that DEPDet is lighter than ordinary object detection models, and the detection accuracy and precision are also the best. In addition, we visualize the test results of DEPDet on SSDD, as shown in Fig. 10.

Fig. 10 visually illustrates the detection performance of our model. Even though the resolution of the SSDD dataset is low and there are many background clutter interference factors, our model can still accurately detect the position of ships.

So as to further verify the efficacy of DEPDet, we selected a number of models for comparison that had been developed for the purpose of detecting ships using SAR technology over the past few years. Due to the confidentiality of most SAR ship detection models and the fact that their source code has not been disclosed, we directly conduct a simple comparison based on the data they provided. We sort the models according to the year they were proposed, and the comparative experimental results are given in Table IV.

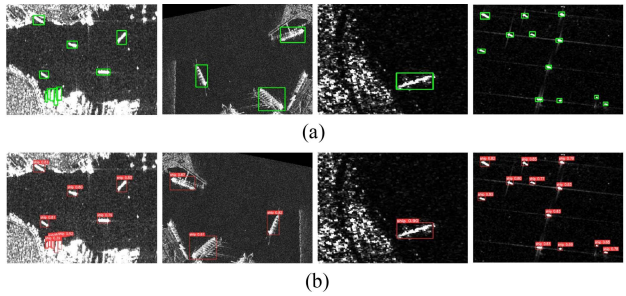

Fig. 10. Visualization test results of DEPDet on SSDD. (a) Ground truth. (b) Our results.

TABLE IV
COMPARISON RESULTS WITH OTHER SAR SHIP DETECTION MODELS ON SSDD

Method	mAP (%)	F1	Params (M)	FLOPs (G)
LFO-Net [32]	80.12	-	-	-
FBR-Net [52]	94.10	0.934	32.5	-
ISASDNet [28]	96.80	-	-	-
D2ADet [33]	96.41	-	-	-
undefined name [53]	94.70	0.913	33.6	236.8
LPEDet [54]	97.40	-	5.7	18.4
ATSD [55]	96.80	0.958	61.5	7.3
MHASD [56]	96.80	0.940	5.5	13.7
PPA-Net [29]	95.19	0.930	-	-
ESarDet [36]	97.96	**0.960**	6.2	7.5
DEPDet (Ours)	**98.20**	0.950	**2.1**	**4.5**

The bold values indicate the best results.

In previous years, most researchers chose a relatively single evaluation index, only concerned about the detection accuracy, besides, did not pay much attention to the issue of real-time detection. Over the past two years, the development of SAR ship detection and the emergence of new challenges have prompted a growing number of researchers to adopt a more comprehensive approach to evaluating the performance of models, employing multiple indicators. The experimental results indicate that our model still demonstrates the most optimal overall performance. MHASD has the smallest params except for our model. Compared with MHASD, we increase mAP and F1 by 1.4% and 1%, respectively, while ensuring the lowest Params and FLOPs. ATSD has the smallest FLOPs besides our model. Compared with ATSD, our mAP has increased by 1.4%, and Params is only about 1/30 of ATSD's Params. Therefore, our model has better lightweight performance. Although ESarDet has the highest F1, our model's mAP is 0.24% higher than ESarDet, while our Params decrease by approximately 66% and FLOPs decrease by approximately 40% compared with ESarDet. In a comprehensive comparison, our model has better performance. In addition, compared with PPA-Net, our model is 3.01% and 2% higher in mAP and F1, respectively. In summary, our model performs excellently in both detection accuracy and Lightweighting of the model.

2) Comparison Experiments on HRSID: So as to better ascertain the efficacy of our model for multiscale ship detection

TABLE V
COMPARISON WITH CLASSICAL OBJECT DETECTION MODELS ON HRSID

Method	P (%)	R (%)	mAP (%)	F1	Params (M)	FLOPs (G)
Faster R-CNN	91.2	78.6	87.8	0.844	41.1	134.4
SSD	89.5	74.7	84.3	0.814	24.3	87.7
YOLOv5	91.7	80.5	89.4	0.857	7.2	16.7
YOLOv8n	91.2	80.8	90.0	0.857	3.2	8.7
DEPDet (Ours)	**92.4**	**82.4**	**91.6**	**0.871**	**2.1**	**4.5**

The bold values indicate the best results.

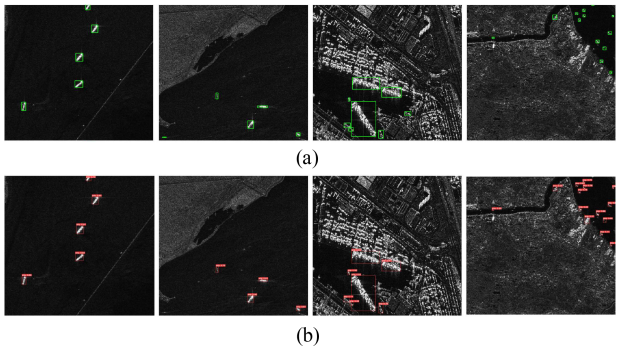

Fig. 11. Visualization test results of DEPDet on HRSID. (a) Ground truth. (b) Our results.

in complex scenes, we chose the HRSID dataset with larger data volume, more complex scenes, and containing abundance remote sensing large-scale scene images for further validation. The experimental results are given in Table V. Similar to the SSDD dataset, we compare DEPDet with Faster R-CNN, SSD, YOLOv5, and baseline model YOLOv8n.

DEPDet still has excellent detection performance on HRSID, with mAP and F1 reaching 91.6% and 0.871, respectively, while ensuring the lightweight. Compared with these models, DEPDet has the lowest Params and FLOPs, which are 2.1M and 4.5G, respectively. In addition, compared with Faster R-CNN, DEPDet improves mAP and F1 by 3.8% and 2.7%, respectively. Compared with SSD, DEPDet improves mAP and F1 by 7.3% and 5.7%, respectively. Compared with YOLOv5, mAP and F1 increased by 2.2% and 1.4%, respectively. Compared with YOLOv8n, mAP and F1 increased by 1.6% and 1.4%, respectively. The results of HRSID indicate that DEPDet is validated with more reliable results for the multiscale ship detection tasks in complex scenes. In addition, we visualize the test results of DEPDet on HRSID, as shown in Fig. 11.

In Fig. 11, HRSID has a higher resolution, but at the same time, the background factors are also more complex, adding many complex scenes such as islands and ports. In addition, the ship targets of HRSID are significantly multiscale, especially for small and medium-sized ships. In comparison to SSDD, HRSID is perceived to be more challenging and difficult to detect, but at the same time, HRSID is also closer to real remote sensing images. Our model still has high detection accuracy in HRSID.

Similarly, to provide additional corroboration for the efficacy of DEPDet, a comparison is made between DEPDet and the latest SAR ship detection model on HRSID. Table VI presents the comparative experimental results.

TABLE VI
COMPARISON RESULTS WITH OTHER SAR SHIP DETECTION MODELS ON HRSID

Method	mAP (%)	F1	Params (M)	FLOPs (G)
undefined name [53]	87.80	0.871	33.6	236.8
LPEDet [54]	89.70	-	5.7	18.4
ATSD [55]	88.19	**0.883**	61.5	7.3
YOLO-SD [57]	83.70	-	59.6	-
SRDet [30]	90.60	-	35.1	-
CSD-YOLO [38]	86.10	-	-	-
DEPDet (Ours)	**91.60**	0.871	**2.1**	**4.5**

The bold values indicate the best results.

Our model has the highest mAP, as well as the lowest Params and FLOPs. Although ATSD is 1.2% higher than our model's F1, our Params have decreased by about 97% compared to ATSD's Params. As a result, our model better meets the lightweight requirements of the model. LPEDet is the smallest Params besides our model. Compared with LPEDet, we increase mAP by 1.9% while ensuring the lowest Params and FLOPs. Therefore, our model has better detection performance. Compared with the improved algorithms YOLO-SD and CSD-YOLO in the YOLO series, our model has increased mAP by 7.9% and 5.5%, respectively. Experimental results indicate that DEPDet balances detection performance and lightweight requirements and can satisfy the detection task of multiscale ships in complex scenes.

3) Visualization Results of Comparative Experiments: So as to ascertain the efficacy of DEPDet in different backgrounds, we visualized the results of comparative experiments. First, four SAR images of large dense ships with simple maritime scenes as well as with fewer disturbing elements were selected, and detection results of distinct models are provided in Fig. 12. Then, we selected six SAR images of multiscale ships with complex backgrounds, such as rivers, ports, and islands. Fig. 13 provides detection results of distinct models. Among them, the first column of each figure is the ground truth of the scene.

In Fig. 12(a), when detecting ships of nondense areas in simple scenes, only Faster R-CNN and SSD have false detections. Originally, there was only one ship but it is detected as two. Other models have excellent detection performance; in short, there are no missed detections in this case. In Fig. 12(b), when ships are denser and multiscale with different sizes, both Faster R-CNN and SSD exhibit false detections and missed detections, and Faster R-CNN is severely missed in the area in the upper left corner of the SAR image. YOLOv5 and YOLOv8n also exhibit slight false detections, with only detection results of DEPDet approach are observed to be in close approximation to the ground truth. However, in Fig. 12(c), when conducting ship detection of dense area in simple scenes, Faster R-CNN shows multitudinous missed detections, and the other three models also have a few missed detections and false detections. Only our model approach is observed to be in close approximation to the ground truth. In addition, in Fig. 12(d), when the background contains a minor degree of clutter interference and ships are densely packed, the difficulty of ship detection increases. Faster R-CNN and SSD have a significant degree of false detection, and they incorrectly detect the background clutter in the upper left corner of a ship.

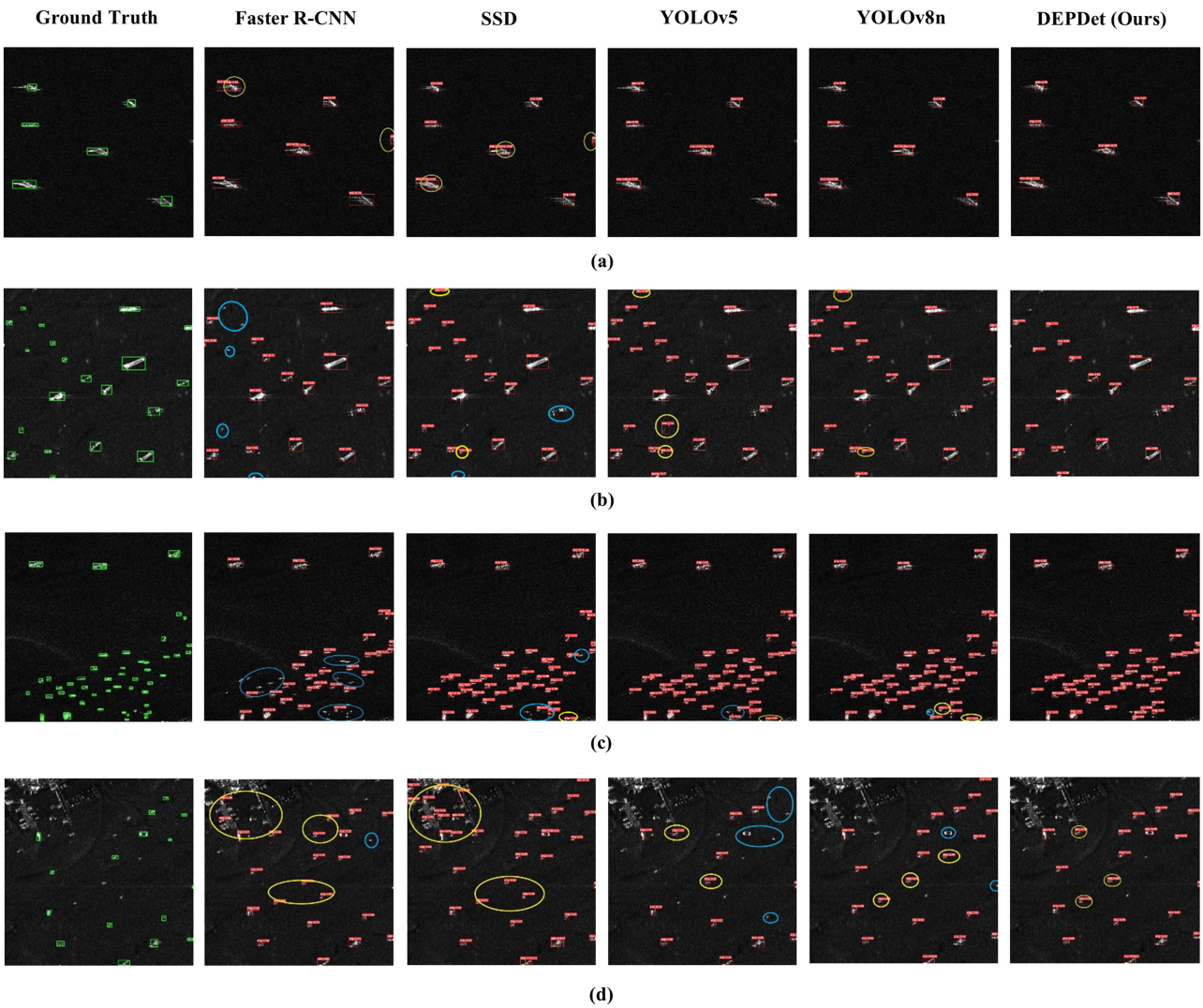

Fig. 12. Comparative experimental results in simple maritime scenes. (a) Scene one. (b) Scene two. (c) Scene three. (d) Scene four. The yellow circle indicates the false detection area, while the blue circle indicates the missed detection area.

Although YOLOv5 distinguishes the clutter, it has multitudinous missed detections, and extremely small ships in the upper right corner have not been detected. YOLOv8n has also experienced many missed detections. Only our model DEPDet did not show any missed detections, with only a very small number of false detections of ships. Compared with other models, our model detects closer the ground truth. In conclusion, our model has excellent detection performance in simple scenes, whether it is for multiscale ships or small dense ships.

Fig. 13 provides detection results of distinct models in six different complex scenes. In Fig. 13(a), the size of ships in the river is inconsistent, leading to the problem of multiscale ships. The detection results of Faster R-CNN and SSD demonstrate that both models have significant instances of missed detections and false positives. Moreover, YOLOv5 missed two ships and did not detect them. YOLOv8n identified a background clutter as a ship. Only our results approach is observed to be in close approximation to the ground truth. In Fig. 13(b), there are multitudinous dense diminutive ships in the river. SSD only detected a very small number of ships. Faster R-CNN and YOLOv5 also exhibit a high degree of false and missed detections. YOLOv8n has a better detection effect but there are still cases of missed detections. Only our model has no missed detections, and by comparison, our model is closer to the ground truth. In Fig. 13(c), the port area, which is subject to the additional challenge of accommodating ships of varying dimensions, was selected as the site for this study. Due to stronger interference factors at the port, the detection performance of Faster R-CNN and SSD is very poor. YOLOv5 and YOLOv8n also have a large number of false detections. Our model only has one false detection area in the bottom right corner, so it is closer to the ground truth. In Fig. 13(d), we chose a port scene containing containers, which has a more complex background. All four models have missed detections, and only our model can accurately detect ships. In Fig. 13(e), we selected the image near islands containing small ships. Faster R-CNN and SSD have a large number of missed detections, while YOLOv5 and YOLOv8n have superior performance in terms of detection. However, there are still some

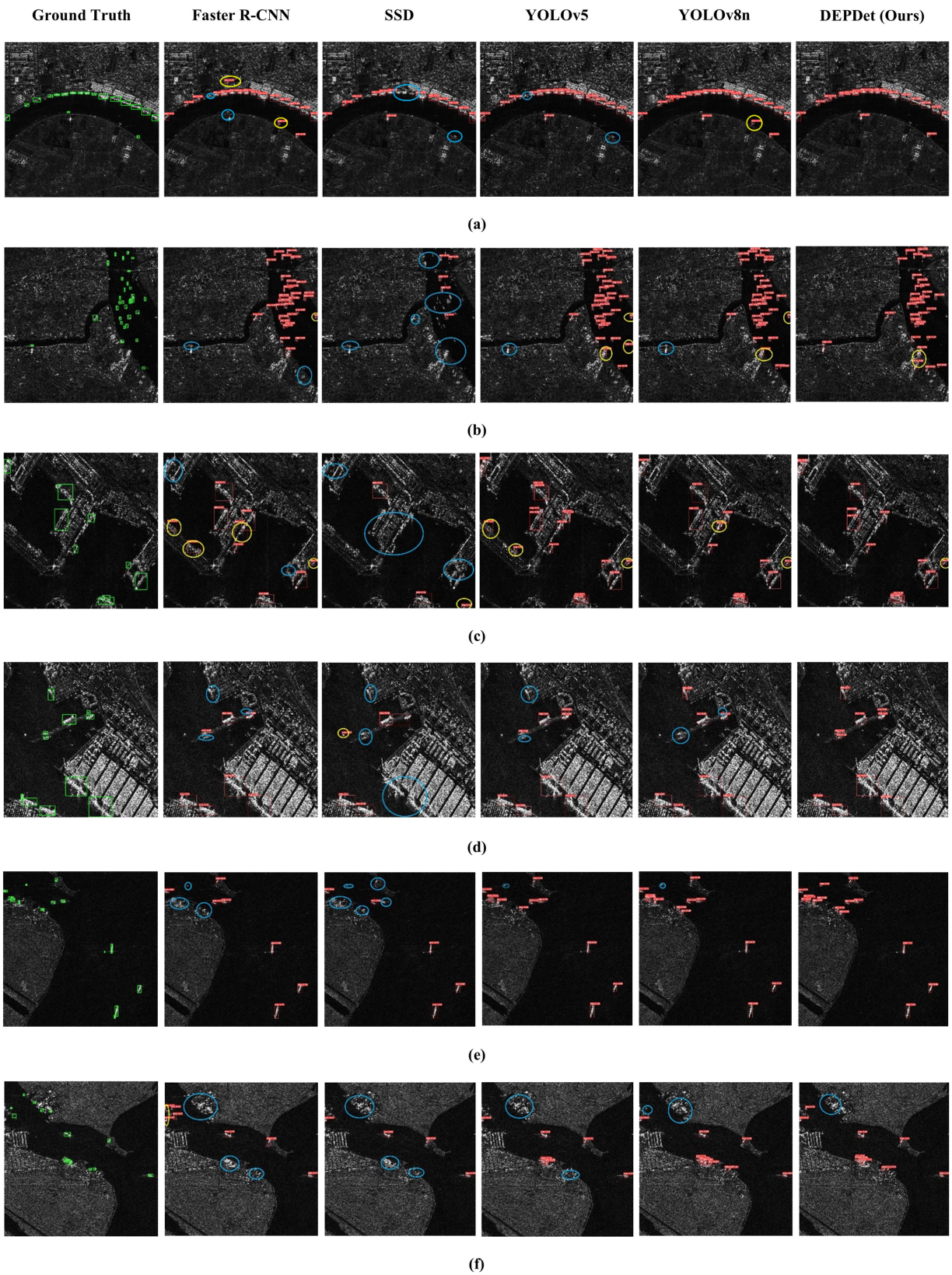

Fig. 13. Comparative experimental results in complex scenes. (a) Scene one. (b) Scene two. (c) Scene three. (d) Scene four. (e) Scene five. (f) Scene six. The yellow circle indicates the false detection area, while the blue circle indicates the missed detection area.

TABLE VII
ABLATION EXPERIMENT

DEMNet	DEPAFPN	PCHead	SSDD				HRSID				Params (M)	FLOPs (G)
			P (%)	R (%)	mAP (%)	F1	P (%)	R (%)	mAP (%)	F1		
			95.2	91.5	96.8	0.933	91.2	80.8	90.0	0.857	3.2	8.7
√			95.1	**95.3**	97.5	**0.951**	90.8	**83.8**	90.8	**0.871**	2.8	7.6
	√		94.8	93.9	97.1	0.943	90.9	80.9	90.2	0.856	2.8	7.8
		√	92.8	92.4	97.0	0.926	91.8	80.9	90.1	0.860	2.3	5.5
√	√	√	**97.9**	92.3	**98.2**	0.950	**92.4**	82.4	**91.6**	**0.871**	**2.1**	**4.5**

The bold values indicate the best results.

missed detections of individual ships. In contrast to them, our model has detected all ships. In Fig. 13(f), we chose a more complex image near the island. Our model has the lowest missed detection rate, with only small ships docked on the island not detected. Compared with other models, our model approach is observed to be in close approximation to the ground truth. In summary, our model effectively solves the problem of multiscale ship detection in complex scenes.

4) Experiment on Resource-Limited Devices: In order to mitigate the financial burden associated with hardware costs, it has become feasible to deploy models on resource-limited computing devices by lightweighting models. To evaluate the lightweight effect of our model, we select a laptop with an Intel Core i7-10750H CPU and an NVIDIA GeForce GTX 1650 4G GPU for the deployment of our model in an experimental setting. The SSDD dataset is selected as the validation test, and the results demonstrate that our model exhibits satisfactory detection performance on a laptop with limited computational resources, where the P is 93.7%, R is 93.3%, mAP is 96.2%, and F1 is 0.935.

E. Ablation Experiment

In order to ascertain the efficacy of DEPDet, we conducted two sets of ablation comparative experiments using SSDD and HRSID datasets. On each dataset, we conducted five ablation experiments for comparison. First, the first experiment used YOLOv8 with no improvement as the baseline, serving as a foundation for subsequent experiments. The objective of the second experiment is to evaluate the effectiveness of feature extraction by replacing the backbone network with our proposed DEMNet. In the third experiment, in order to ascertain the efficacy of the proposed feature fusion, the neck network was replaced with our DEPAFPN module. In the fourth experiment, we replaced the detection head with our newly designed lightweight detection head PCHead to verify the effectiveness of lightweighting. Finally, in the fifth experiment, we superimposed the second, third, and fourth experiments to validate the effectiveness of the proposed DEPDet. The results of ablation experiments are presented in Table VII.

Our proposed model DEPDet achieved 98.2% and 91.6% mAP on the SSDD and HRSID datasets, respectively, and the F1 values reached 0.950 and 0.871. Compared with the baseline model, our model has improved in accuracy, with mAP improving by 1.4% and 1.6%, and F1 improving by 1.7% and 1.4%, respectively. Meanwhile, our model achieves a more lightweight effect. Among them, the Params of the model decreased from 3.2M to 2.1M, a decrease of approximately 34%, and the FLOPs of the model decreased from 8.7G to 4.5G, a decrease of approximately 48%.

In the second experiment, the DEMNet backbone network greatly demonstrated a significant enhancement in the model's feature extraction capabilities. In comparison to the baseline model, it demonstrated an improvement of mAP by 0.7% and F1 by 1.8% on SSDD, with mAP by 0.8% and F1 by 1.4% on HRSID. In the third experiment, so as to better fuse extracted features, the neck network used our designed DEPAFPN. Compared with the baseline model, it increased mAP by 0.3% on SSDD and 0.2% on HRSID. These two experiments have verified the effectiveness of our improvement strategy, and while the accuracy has improved, the number of model parameters has not increased. However, the model is not lightweight enough at this point, so in the fourth experiment, we used our designed lightweight detection head PCHead. Compared with the baseline model, although the detection accuracy on SSDD and HRSID was not significantly improved, it reduced the Params of the model by about 28%, and the FLOPs of the model by about 37%, achieving a good lightweight effect.

Furthermore, so as to gain a more intuitive understanding of the performance of our model, we introduced LayerCAM [58] heatmaps to visualize the ablation experiment. Fig. 14 shows the fitting effect of different modules on the ship.

From the results of the heat map, compared with the baseline model, DEPDet fits the ship features better and takes more care of the overall features of the ship. In Fig. 14(a), DEMNet enhances the model's learning ability and focuses more on the edge features of ships, achieving multiscale feature extraction. In Fig. 14(b), DEPAFPN helps the model achieve better feature fusion and reduces the influence of interference factors. In Fig. 14(c), the final output result of PCHead is the output result of DEPDet. The experimental results show that the three modules proposed in this article can effectively improve the performance of multiscale SAR ship detection in complex scenes.

F. Generalization Experiment

Ultimately, one of China's Gaofen-3 complex large-scene remote sensing images from the AIR_SARShip_1.0 [59] dataset was selected to assess the generalization performance of the proposed model. The remote sensing image, which has the dimensions of approximately 3000×3000 pixels, depicts a multitude of concentrated maritime traffic, as well as a variety of ships navigating within the confines of inland ports. As illustrated in Fig. 15, our model is capable of extracting the ship positions with a high degree of effectiveness, which can be applied to

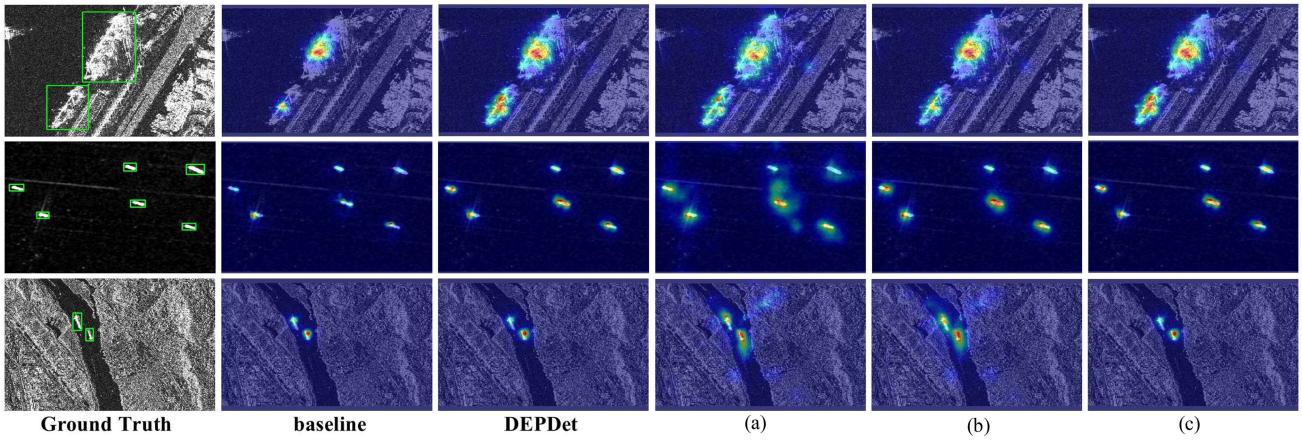

Fig. 14. Visual results of ablation experiment. The first column is the ground truth. The second column is the output of the baseline model. The third column is the output of DEPDet. The last three columns represent the process of feature extraction, feature fusion, and feature output of DEPDet. (a) Output of the backbone network layers. (b) Output of the backbone network layers and the neck network layers. (c) Output of the backbone network layer, the neck network layers, and the detection head layers.

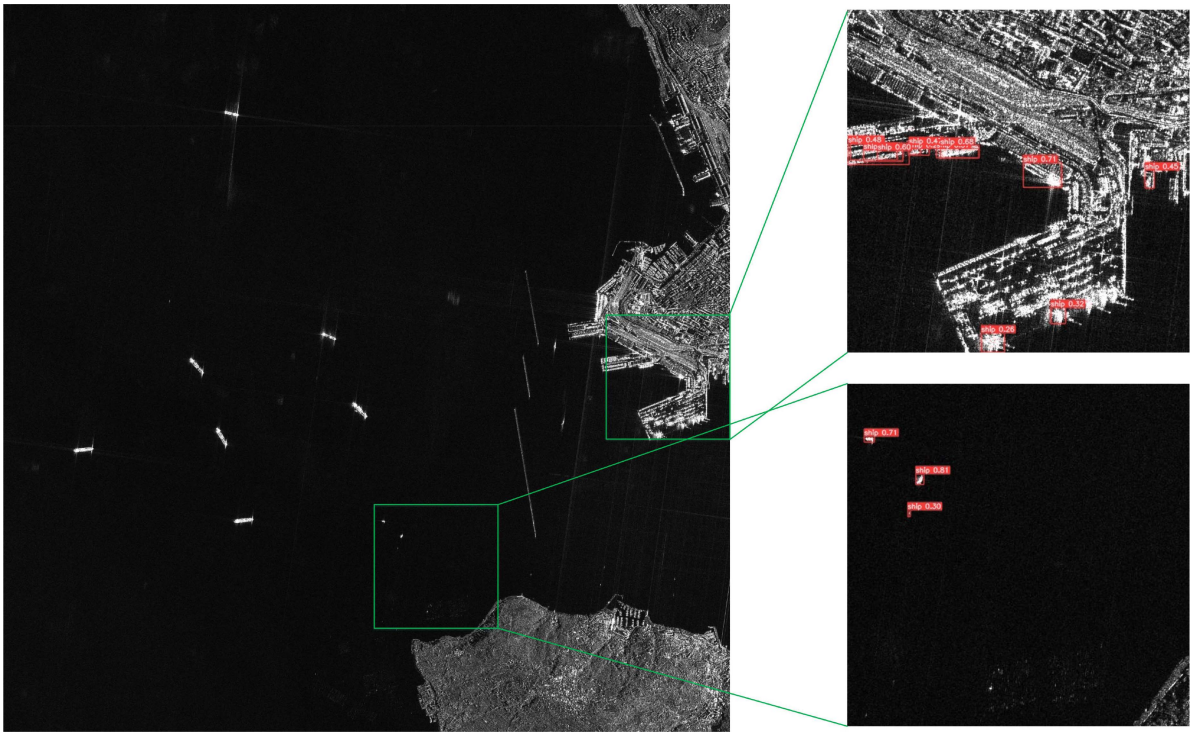

Fig. 15. Visual result of the Gaofen-3 complex large-scene remote sensing image.

real remote sensing scenarios with satisfactory generalization performance.

V. Discussion

So as to address the challenges posed by complex scenes, multiscale ships, and difficult model deployment in SAR images, we propose DEPDet and design a backbone network DEMNet, a neck network DEPAFPN, and a lightweight detection head PCHead. In order to ascertain the efficacy of the proposed DEPDet, a series of comparative experiments were conducted on the SSDD and HRSID datasets. Empirical evidence indicates that our model obtains 98.2% and 91.6% mAP on SSDD and HRSID, respectively, and the F1 reaches 0.950 and 0.871. At the same time, the Params and FLOPs of our model are only 2.1M and 4.5G, respectively. Compared with some classical target detection models and SAR ship detection models, our proposed model demonstrates effective detection of multiscale ships in complex scenes while the model is lightweighted, with

Fig. 16. Multispectral image ship detection effect.

outstanding performance with regard to both detection accuracy and lightweight effect.

To validate the reliability of our proposed modules, we conducted a series of ablation experiments and employed heat maps to illustrate the impact of each module. DEMNet has been sufficiently enhanced with CSMSConv and EMA, thereby facilitating the more efficient extraction of edge features and multiscale features. DEPAFPN is more effective at fusing multiscale features and suppressing the interference of background factors. PCHead is capable of achieving efficient detection accuracy while maintaining a lightweight effect. Finally, when applied to complex large-scale scene remote sensing images, the model is still able to effectively extract ship positions with good generalization performance.

The model's extensibility can enhance its potential for application. In ship detection tasks, in addition to SAR remote sensing images, multispectral remote sensing images are frequently employed for this purpose. In order to evaluate the extensibility of our proposed model, the optical remote sensing image dataset NWPU VHR-10 [60] was selected for experimental testing. The dataset comprises a total of 800 VHR optical remote sensing images.

Following experimentation, the values of P, R, mAP, and F1 are 88.1%, 76.6%, 88.1%, and 0.819, respectively. Fig. 16 provides a visual representation of the multispectral image ship detection effect. Although our model can be extended to multispectral image ship detection tasks, the presence of cloud cover leads to a reduction in detection accuracy. In addition, multispectral images also lack the capacity of providing comprehensive ship detection. It is for these reasons that ship detection applications are more commonly used in SAR images, which allows for comprehensive monitoring of ships and thus facilitates real-time detection.

Although DEPDet has good performance, it still has some limitations. Due to the fact that our detection results are not directional boxes, we cannot detect the direction of ships. When the real ships on the ground are densely overlapped, our model may detect multiple ships as one, resulting in incorrect detection. Furthermore, as our methodology employs a fixed dataset and lacks an associated embedded device, it is unable to provide an accurate assessment of the detection speed.

In future work, we will further improve these bottlenecks that affect model performance and attempt to apply directional boxes for ship detection. At the same time, we will investigate further research into the potential of lightweight methodologies, including pruning and distillation. In addition, we will introduce pertinent embedded devices with the objective of evaluating the detection speed of the model using FPS metrics, thereby facilitating the deployment of our detection model on hardware platforms and enabling real-time detection.

VI. Conclusion

This article introduces DEPDet, which is a cross-spatial multiscale lightweight network for SAR ship detection. First, so as to better enhance the efficacy of the detection of multiscale ships, we design the CSMSConv, and introduce deformable convolution to enrich the edge information of ships. Then, it is combined with the EMA to design the DEMNet, which enhances the multiscale information and contextual information of the ship and effectively eliminates the background clutter, thereby enabling the operator to focus on the target ship. Furthermore, so as to facilitate the integration of disparate scale features, we design a new path aggregation FPN DEPAFPN, which combines deformable convolution and CSMSConv module CSMSC2F, which effectively avoids information redundancy. Second, we introduce PConv to construct a lightweight detection head module PCHead, which can abstract spatial information more effectively through a concerted effort to reduce duplicated computations and storage operations. Extensive comparative experiments are carried out on SSDD and HRSID, and the results demonstrated that our model has excellent performance. In addition, we successfully deployed the model on low-configuration device, and our model still maintained good detection accuracy. Furthermore, a sequence of ablation and heatmap experiments prove the efficacy of DEMNet, DEPAFPN, and PCHead, and they can prove beneficial in terms of the efficacy of DEPDet. Ultimately, the model has been tested on ship detection in remote sensing of large-scale scene images and multispectral images. This provides evidence that the model has good generalization and extensibility.

In conclusion, DEPDet achieves a balance between detection accuracy and lightweight effect in the SAR ship detection task. The experimental results demonstrate that the model is capable of achieving satisfactory results in the multiscale SAR ship detection task in complex scenes, and at the same time, while achieving the lightweight requirements of the SAR ship detection model, as well as good generalization and extensibility. Subsequently, our research team will continue to refine our model with the aim of implementing it on various hardware platforms and achieving real-time ship detection. The code can be obtained by https://github.com/Marine0011/DEPDet.

References

[1] J. Li, C. Xu, H. Su, L. Gao, and T. Wang, "Deep learning for SAR ship detection: Past, present and future," *Remote Sens.*, vol. 14, no. 11, Jun. 2022, Art. no. 2712.

[2] S. Wang, J. Guo, Y. Zhang, and Y. Wu, "Multi-baseline SAR 3D reconstruction of vehicle from very sparse aspects: A generative adversarial network based approach," *ISPRS J. Photogrammetry Remote Sens.*, vol. 197, pp. 36–55, 2023.

[3] Z. Zhou et al., "FSODS: A lightweight metalearning method for few-shot object detection on SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5232217, doi: 10.1109/TGRS.2022.3192996.

[4] Y. Han et al., "Water distribution based on SAR and optical data to improve hazard mapping," *Environ. Res.*, vol. 235, Oct. 2023, Art. no. 116694.

[5] K. Ding et al., "Towards real-time detection of ships and wakes with lightweight deep learning model in Gaofen-3 SAR images," *Remote Sens. Environ.*, vol. 284, Jan. 2023, Art. no. 113345.

[6] T. Zhang et al., "Balance learning for ship detection from synthetic aperture radar remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 182, pp. 190–207, Dec. 2021.

[7] P. W. Vachon, C. Kabatoff, and R. Quinn, "Operational ship detection in Canada using RADARSAT," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 998–1001.

[8] A. Kurekin et al., "Use of Sentinel-l and Sentinel-2 for monitoring illegal fishing off Ghana," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 6875–6878.

[9] X. Ma, K. Ji, B. Xiong, L. Zhang, S. Feng, and G. Kuang, "Light-YOLOv4: An edge-device oriented target detection method for remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10808–10820, 2021, doi: 10.1109/JSTARS.2021.3120009.

[10] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[11] H. Jiang, X. Hu, K. Li, J. Zhang, J. Gong, and M. Zhang, "PGA-SiamNet: Pyramid feature-based attention-guided Siamese network for remote sensing orthoimagery building change detection," *Remote Sens.*, vol. 12, no. 3, Jan. 2020, Art. no. 484.

[12] A. Namburu, P. Selvaraj, S. Mohan, S. Ragavanantham, and E. T. Eldin, "Forest fire identification in UAV imagery using X-MobileNet," *Electronics*, vol. 12, no. 3, Jan. 2023, Art. no. 733.

[13] Z. Deng, H. Sun, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3652–3664, Aug. 2017, doi: 10.1109/JSTARS.2017.2694890.

[14] Z. Hong et al., "Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6083–6101, 2021, doi: 10.1109/JSTARS.2021.3087555.

[15] J. Li, J. Chen, P. Cheng, Z. Yu, L. Yu, and C. Chi, "A survey on deep-learning-based real-time SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3218–3247, 2023, doi: 10.1109/JSTARS.2023.3244616.

[16] 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, doi: 10.1109/CVPR.2016.91.

[17] W. Liu et al., "SSD: Single shot multibox detector," in *Computer Vision – ECCV*. Berlin, Germany: Springer, 2016, pp. 21–37.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[21] X. Ren, Y. Bai, G. Liu, and P. Zhang, "YOLO-lite: An efficient lightweight network for SAR ship detection," *Remote Sens.*, vol. 15, no. 15, Jul. 2023, Art. no. 3771.

[22] J. Akhtar and K. E. Olsen, "GO-CFAR trained neural network target detectors," in *Proc. IEEE Radar Conf.*, 2019, pp. 1–5.

[23] J. Akhtar, "Training of neural network target detectors mentored by SO-CFAR," in *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 1522–1526.

[24] Y. Ge, Z.-Z. Huang, Q. Liu, and Q. Lan, "Fast implementation of CA-CFAR algorithm based on FFT," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Automat. Control Conf.*, 2021, pp. 2494–2498.

[25] M.-D. Li, X.-C. Cui, and S.-W. Chen, "Adaptive superpixel-level CFAR detector for SAR inshore dense ship detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4010405.

[26] J. Ai, Z. Pei, B. Yao, Z. Wang, and M. Xing, "AIS data aided Rayleigh CFAR ship detection algorithm of multiple-target environment in SAR images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 2, pp. 1266–1282, Apr. 2022.

[27] H. Madjidi and T. Laroussi, "Approximate MLE based automatic bilateral censoring CFAR ship detection for complex scenes of log-normal sea clutter in SAR imagery," *Digit. Signal Process.*, vol. 136, May 2023, Art. no. 103972.

[28] Z. Wu, B. Hou, B. Ren, Z. Ren, S. Wang, and L. Jiao, "A deep detection network based on interaction of instance segmentation and object detection for SAR images," *Remote Sens.*, vol. 13, no. 13, Jan. 2021, Art. no. 2582.

[29] G. Tang et al., "PPA-Net: Pyramid pooling attention network for multi-scale ship detection in SAR images," *Remote Sens.*, vol. 15, no. 11, Jan. 2023, Art. no. 2855.

[30] J. Lv et al., "An anchor-free detection algorithm for SAR ship targets with deep saliency representation," *Remote Sens.*, vol. 15, no. 1, Jan. 2023, Art. no. 103.

[31] X. Zhou, V. Koltun, and P. Krähenbühl, "Probabilistic two-stage detection," 2021, *arXiv:2103.07461*.

[32] X. Zhang et al., "A lightweight feature optimizing network for ship detection in SAR image," *IEEE Access*, vol. 7, pp. 141662–141678, 2019.

[33] S. Zhao, K. Qiao, W. Yu, R. Yang, and J. Lv, "D ^{2} ADet: Dual-activation domain RetinaNet based on a deformable feature pyramid network for ship detection in SAR images," *Remote Sens. Lett.*, vol. 13, no. 11, pp. 1153–1163, Nov. 2022.

[34] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.

[35] Z. Zhou et al., "HRLE-SARDet: A lightweight SAR target detection algorithm based on hybrid representation learning enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5203922, doi: 10.1109/TGRS.2023.3251694.

[36] Y. Zhang, C. Chen, R. Hu, and Y. Yu, "ESarDet: An efficient SAR ship detection method based on context information and large effective receptive field," *Remote Sens.*, vol. 15, no. 12, Jun. 2023, Art. no. 3018.

[37] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[38] Z. Chen, C. Liu, V. Filaretov, and D. Yukhimets, "Multi-scale ship detection algorithm based on YOLOv7 for complex scene SAR images," *Remote Sens.*, vol. 15, no. 8, Apr. 2023, Art. no. 2071.

[39] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7464–7475.

[40] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, Oct. 2021, Art. no. 4209, doi: 10.3390/rs13214209.

[41] B. Xiong, Z. Sun, J. Wang, X. Leng, and K. Ji, "A lightweight model for ship detection and recognition in complex-scene SAR images," *Remote Sens.*, vol. 14, no. 23, Nov. 2022, Art. no. 6053, doi: 10.3390/rs14236053.

[42] L. Zhang, G. Ding, C. Li, and D. Li, "DCF-Yolov8: An improved algorithm for aggregating low-level features to detect agricultural pests and diseases," *Agronomy*, vol. 13, no. 8, Jul. 2023, Art. no. 2012.

[43] G. Yang, J. Wang, Z. Nie, H. Yang, and S. Yu, "A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention," *Agronomy*, vol. 13, no. 7, Jul. 2023, Art. no. 1824.

[44] 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.

[45] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944.

[46] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," presented at the Proc. of the IEEE Conf. Comput. Vis. and Pattern Recognit., 2018, pp. 7132–7141.

[47] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13708–13717.

[48] D. Ouyang et al., "Efficient multi-scale attention module with cross-spatial learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2023, pp. 1–5.

[49] J. Chen et al., "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 12021–12031.

[50] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era, Models, Methods Appl.*, 2017, pp. 1–6.

[51] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.

[52] J. Fu, X. Sun, Z. Wang, and K. Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021, doi: 10.1109/TGRS.2020.3005151.
[53] H. Shi, B. Chai, Y. Wang, and L. Chen, "A local-sparse-information-aggregation transformer with explicit contour guidance for SAR ship detection," *Remote Sens.*, vol. 14, no. 20, Oct. 2022, Art. no. 5247.
[54] Y. Feng et al., "A lightweight position-enhanced anchor-free algorithm for SAR ship detection," *Remote Sens.*, vol. 14, no. 8, Apr. 2022, Art. no. 1908.
[55] C. Yao, P. Xie, L. Zhang, and Y. Fang, "ATSD: Anchor-free two-stage ship detection based on feature enhancement in SAR images," *Remote Sens.*, vol. 14, no. 23, Jan. 2022, Art. no. 6058.
[56] N. Yu, H. Ren, T. Deng, and X. Fan, "A lightweight radar ship detection framework with hybrid attentions," *Remote Sens.*, vol. 15, no. 11, Jan. 2023, Art. no. 2743.
[57] S. Wang et al., "YOLO-SD: Small ship detection in SAR images by multi-scale convolution and feature transformer module," *Remote Sens.*, vol. 14, no. 20, Oct. 2022, Art. no. 5268.
[58] P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, and Y. Wei, "LayerCAM: Exploring hierarchical class activation maps for localization," *IEEE Trans. Image Process.*, vol. 30, pp. 5875–5888, 2021.
[59] X. Sun, Z. Wang, Y. Sun, W. Diao, Y. Zhang, and K. Fu, "AIR-SARShip-1.0: High-resolution SAR ship detection dataset," *J. Radars*, vol. 8, no. 6, pp. 852–862, 2019.
[60] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS J. Photogrammetry Remote Sens.*, vol. 98, pp. 119–132, Dec. 2014.

Jing Zhang received the B.E. degree in data science and big data technology from Shandong Jiaotong University, Jinan, China, in 2022. She is currently working toward the M.S. degree in surveying and mapping engineering with the School of Geosciences, Yangtze University, Wuhan, China.

Her research interests include object detection and remote sensing image processing.

Fan Deng received the B.S. degree in geographical information systems from the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China, in 2006, and the M.S. and Ph.D. degrees in physical geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, in 2010 and 2013, respectively.

He is an Assistant Professor with Yangtze University, Hubei, China. His research interests lie in remote sensing monitoring of ecological environment, including vegetation remote sensing, thermal environment remote sensing, and ecohydrological remote sensing.

Yonghua Wang received the B.E. degree in surveying and mapping engineering from Wuhan University, Wuhan, China, in 2001.

He is currently with Wuhan Huaxin Lianchuang Technology Engineering, Co., Wuhan. His research interests include meteorological big data development and remote sensing research and development and application.

Jie Gong received the Ph.D. degree in geographical information systems from the China University of Geosciences, Wuhan, China, in 2024.

He is a Postgraduate Tutor with a renowned university and a Senior Engineer. He currently serves as the Chairman, General Manager, and Party Branch Secretary with Wuhan Huaxin Lianchuang Technology Engineering, Co., Wuhan.

His research interests encompass meteorological big data development and remote sensing research and development and application, with a particular focus on the key areas of meteorological observation informatization, geological hazards, hydrology, and ecological and environmental protection informatization.

Ziyang Liu is currently working toward the M.S. degree in cartography and geographic information engineering with the School of Geosciences, Yangtze University, Wuhan, China, under the guidance of Dr. F. Deng.

His main research interests include remote sensing image processing and applications.

Wenjun Liu received the B.S. degree in geographic information science in 2022 from Yangtze University, Wuhan, China, where she is currently working toward the M.S. degree in surveying and mapping engineering.

Her research interests include thermal environment remote sensing and intelligent extraction of remote sensing information.

Yinmei Zeng received the B.S. degree in geographic information science from the Guilin University of Technology, Guilin, China, in 2023. She is currently working toward the M.S. degree in surveying and mapping engineering with the School of Geosciences, Yangtze University, Wuhan, China.

Her research interests include intelligent extraction of remote sensing information and ecohydrological remote sensing.

Zeqiang Chen received the Ph.D. degree in geographical information systems from Wuhan University, Wuhan, China, in 2012.

He is a Professor with the National Engineering Research Center of Geographic Information System, China University of Geosciences at Wuhan, Wuhan. His research interests include sensor web, spatiotemporal Big Data, GeoAI, and their applications in smart watersheds.