Adaptive Multiscale Reversible Column Network for SAR Ship Detection

Tianxiang Wang and Zhangfan Zeng

Abstract—Ship detection via synthetic aperture radar (SAR) is widely used in maritime safety and maritime traffic control, etc. Recently, deep learning methods are employed to improve the performance of SAR ship detection to a large extent. However, the presence of clutter in SAR images and the large-scale difference of ships result in the diminished detection performance in complex environments. As such, in this article, a novel adaptive multiscale reversible column network is proposed. First, the idea of disentangled feature learning is applied to construct reversible column networks with a C2f module to alleviate the problem of large-scale differences and the loss of ship information. Second, the multiplexed adaptive spatial pyramid pooling module is proposed to alleviate the impact of complex background clutter through multiplexed pooling operations and adaptive fusion. Finally, a novel feature pyramid network with an adaptive downsampling module is designed to reduce the information loss caused by downsampling while enhancing the multiscale ship detection capability. The effectiveness of the proposed method is validated on two public datasets: 1) SAR ship detection dataset and 2) high-resolution SAR images dataset. The experimental results show that the proposed method is able to achieve better results than current state-of-the-art methods for SAR ship detection in complex environments and large-scale differences of ships.

Index Terms—Adaptive downsampling, feature extraction, reversible column network, ship detection, synthetic aperture radar (SAR).

I. INTRODUCTION

THE development of synthetic aperture radar (SAR) technology has significantly advanced the field of remote-sensing target detection, particularly demonstrating outstanding potential in maritime monitoring [1], [2], [3]. Unlike optical sensors, which are suboptimal in stringent conditions due to sensitivity to weather and lighting conditions, SAR exceeds in various weather conditions and offers higher data rates and larger processing capacities [4]. Furthermore, SAR benefits from the unique ability to penetrate challenging environments for imaging and to continuously acquire geographical information even in complex weather conditions. These advantages make SAR valuable for applications in various fields, such as target detection and recognition [5], [6], geomorphology and terrain mapping [7], and segmentation and classification [8], [9], [10].

Maritime ship detection is a crucial task in the fields of maritime traffic control, maritime resource exploitation, and maritime environmental protection [11], [12], [13]. Recent studies [1], [2], [3] have shown that SAR also plays an important role in ship detection due to its unique imaging mechanism. However, SAR data can be subject to various degradation factors, noise effects, and variabilities during the imaging process. Unlike spectral variability [14], SAR images mainly suffer from topology and adverse changes in environments, leading to variations in the microwave signals of targets. Topology in SAR imaging typically refers to the geometric arrangement of land features, including their spatial relationships and characteristics. More specifically, the variation in the incidence angle affects the interaction between the radar signal and the terrain. In areas with complex topography, radar signals may cast shadows or result in layover effects. This can lead to distorted or ambiguous representations in the SAR image. Speckle is a common artifact in SAR imaging, resulting from the interference of radar signals with multiple scattering centers on the ground. Consequently, SAR ship detection faces significant challenges in complex coastal environments and harsh marine conditions. For instance, in coastal environments, SAR microwave signals can encounter reflection, scattering, and refraction from ships, coastal constructions, and inshore constructions, which can result in false detection. In offshore scenarios, complex marine conditions such as sea surface turbulence and ship wakes introduce intricate background clutter in SAR images. Moreover, the large-scale difference in SAR images, often influenced by different observation geometries, poses challenges for SAR ship detection.

Recent methods can be broadly categorized into traditional methods and deep-learning-based methods [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36]. Most of the traditional methods are designed from the perspective of signal processing. Among them, constant false alarm rate (CFAR)-based methods are widely used in SAR ship detection, as explored in related research papers [15], [16], [17], [18], [19]. The CFAR method is a statistically based signal processing technique that adaptively estimates the statistical properties of the background clutter and determines an appropriate threshold to suppress the background clutter and detect the target signal. However, in complex environments with strong clutter, multiple targets, and inhomogeneous noise, the CFAR algorithm may not be able to

accurately estimate the statistical properties of the background, resulting in missed or false alarms.

In recent years, deep learning technology has been widely applied to SAR ship detection [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36]. Compared to the CFAR method, deep learning methods exhibit stronger automatic feature learning ability and robustness, avoiding the tediousness and uncertainty associated with manual design processes. However, deep learning methods in the fields of object detection, such as Faster R-CNN [37], FoveaBox [38], etc., and remote-sensing object detection of optical sensors [39], [40], when directly applied to SAR ship detection, are susceptible to the influence of clutter, significant scattering of target information, and large-scale differences, leading to a degradation in detection performance. As such, numerous improvement methods applicable to SAR ship detection have been proposed. Specifically, to enhance the performance of multiscale ship detection in SAR images, feature pyramid network (FPN)-based methods have been largely proposed [20]. Jiao et al. [21] proposed an end-to-end dense join to extract features at different scales. Cui et al. [22] reported a novel DAPN-based multiscale extraction method with a CBAM attention mechanism module. Zhang et al. [23] presented a novel quad FPN for integrating feature maps in SAR ship detection. Li et al. [24] proposed a multidimensional domain deep learning network with complementary features in spatial and frequency domains. Zhou et al. [25] used a novel Res2-based idea embedded into the backbone network to further extract multiscale ship information. Based on the bidirectional convolutional structure, Yu et al. [26] proposed the two-way convolutional network and utilized multiscale feature mapping to process SAR images. To mitigate the influence of complex background clutter, Zhao et al. [27] designed a dual-feature fusion attention mechanism that combines shallow features and denoised features, effectively reducing clutter from complex backgrounds. Zhang et al. [28] employed a one-stage detector with a frequency attention module, which can process frequency-domain information adaptively and suppress the sea cluster in the SAR images. Wang et al. [29] presented a multifeature fusion network, which reduced the interference of complex backgrounds by extracting frequency-domain features and combining spatial, high-frequency, and low-frequency texture information. Bai et al. [30] introduced the feature enhancement pyramid and the shallow feature reconstruction module to address scattered spots and noises. Furthermore, the significance of sidelobes is amplified by the weak scattering from small ships and the clutter of background clutter. To alleviate the impact of sidelobes, Zhou et al. [31] proposed feature extraction with sidelobe awareness (FESA), which addresses the sidelobe effect by incorporating maximum pooling and minimum pooling operations along the X and Y axes, respectively. Zhou et al. [32] integrate the respective advantages of a convolutional neural network (CNN) and self-attention to enhance the extraction of scattering information from small targets. Compared to anchor-based methods, researchers have observed advantages in anchor-free methods such as flexibility, simplified designs, adaptability to complex scenes, and lower computational complexity. Jiang et al. [33] constructed an anchor-free detector, which is designed with a foreground enhancement module to reduce the impact of complex backgrounds. Hu et al. [34] constructed an anchor-free balanced attention network, which introduced dynamic convolution to construct a local attention module to obtain local information and designed a nonlocal attention module to extract nonlocal features. To address effectively sparse labeled samples and imbalanced categories, Gao et al. [35] proposed an attention-dense-CycleGAN method for ship automatic target recognition, specifically designed for the optical to SAR transfer learning task. The transformer is also employed for SAR ship detection tasks due to its capability to capture long-term dependencies. A vision transformer architecture named CR-TransSar was proposed by Xia et al. [36], combining transformer and CNN to enhance context learning. However, the lack of utilization of local information and multilevel feature representation in transformer-based methods makes multiscale ship detection more challenging, especially for small-scale ships.

Despite progress in SAR ship detection with various methods, the challenges posed by environmental clutter and scale differences need to be further addressed. Specifically, due to the presence of environmental clutter, ship features are similar to surrounding noise features, which results in the masking or obscuring of semantically weaker ship features. Moreover, the prevalence of small-scale ships in SAR images leads to deterioration in the performance of multiscale ship detection. The previous methods, which entirely adopt the strategy based on FPN, alleviate the challenge of significant scale differences among ships to some extent. However, owing to the structural and available pixel constraints of small-scale ships and the unique imaging mechanism of SAR, feature information for small-scale ships will be lost or erroneously identified as background during the feature extraction process. It can be attributed to the fact that deep learning methods widely used in SAR ship detection usually follow the information bottleneck (IB) principle [41], [42], where the information will be compressed or even discarded. Therefore, from the perspective of feature learning, the application of the IB will be inappropriate for SAR image ship detection. The disentangled feature learning [43], [44] is widely used in the area of computer vision, especially for target detection. This method involves the embedding of task-relevant semantic and positional information into separate decoupling dimensions while maintaining the same amount of information as the input. Revcol is the most representative network based on disentangled feature learning, proposed by Cai et al. [45]. In Revcol, a multiplexed subcolumn network is utilized for information extraction, where the information from different layers in different subcolumns is embedded into the next subcolumn using a reversible transformation. Thanks to the reversible transformation, information partially lost in one subcolumn network can be acquired in another subcolumn network. It is proved that the disentangled feature learning is effective for multiscale ship detection, particularly for small-scale ships.

Based on the analysis provided, in order to enhance the performance of multiscale ship detection, especially for small-scale ships, as well as to address the impact of complex background clutters in SAR images, an adaptive multiscale reversible column
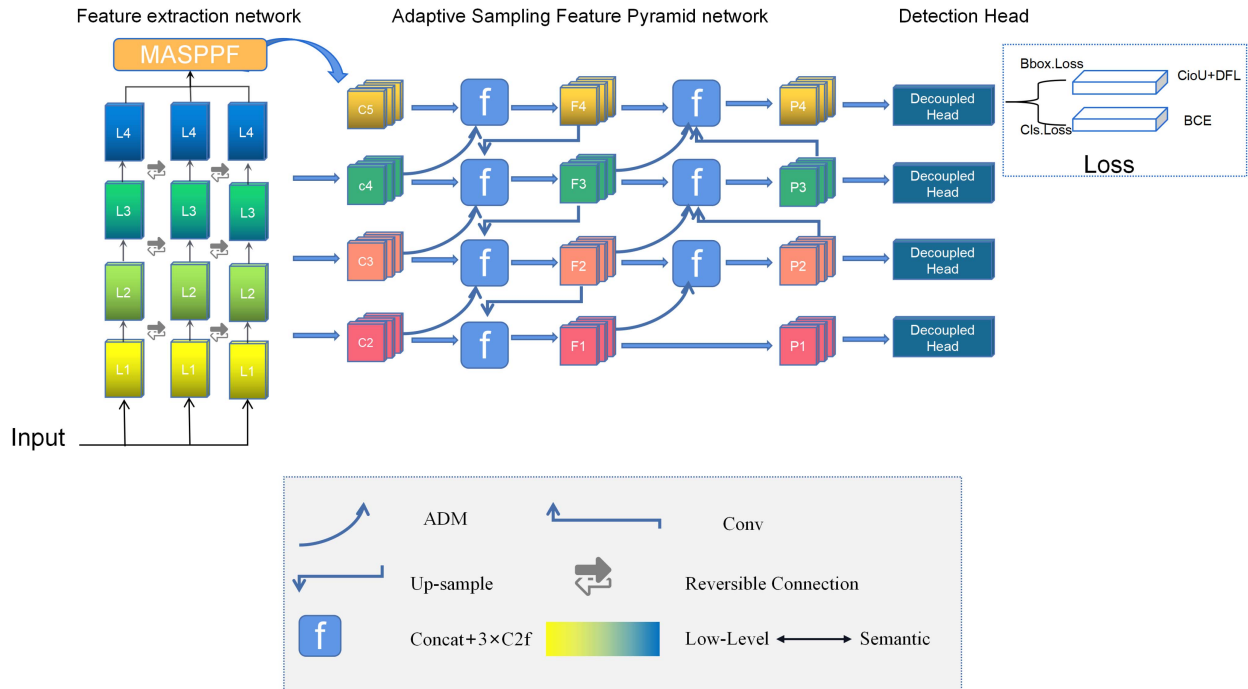
Fig. 1. General architecture of the AMRCNet network. It consists of three main parts: 1) The feature extraction part with Revcol-C2f as the backbone and MASPPF as the feature enhancement layer; 2) the neck combined by ADM and RepGFPN; and 3) the parallel detection head.

network for SAR ship detection (AMRCNet) is proposed in this article. The main contributions of this article are as follows.

1) A novel ship detection network, AMRCNet, is proposed, which demonstrates superior multiscale ship detection performance in complex environments and mitigates the negative effects of complex background clutters, improving the detection accuracy of ships in SAR images.

2) To address the degraded SAR multiscale ship detection caused by the loss of partial information due to the IB principle, Revcol is applied to reconstruct the backbone network of YOLOv8s [46], resulting in the reversible column networks with a C2f module (Revcol-C2f) backbone network for feature extraction.

3) To mitigate complex background clutter in SAR ship detection, this article proposes a novel multiplexed adaptive spatial pyramid pooling layer (MASPPF). In MASPPF, multiplexed large kernel pooling operations and adaptive fusion are utilized to improve the target perception capabilities of the network and mitigate the influence of background clutter in environments of all scales.

4) To further enhance the detection capabilities of multiscale ships, especially small ships, in SAR images, this article constructs an adaptive sampling FPN (SAFPN). Within SAFPN, an adaptive downsampling module (ADM) is devised to address issues related to acquiring irrelevant information and information loss during feature pyramid downsampling (DS). The ADM facilitates the extraction of semantic information and precise spatial localization. Furthermore, a bidirectional fusion of shallow and deep

features at different scales is implemented to merge feature information.

In order to validate the effectiveness of our method, a large number of ablation experiments and comparative experiments have been implemented on the SAR ship detection dataset (SSDD) [47] and high-resolution SAR images dataset (HRSID) [48].

The rest of the article is structured as follows: Section II describes the proposed method. Section III validates the proposed method. Section IV offers a comprehensive discussion on the experimental results. Finally, Section V concludes this article.

II. PROPOSED METHOD

The general architecture of the proposed method is shown in Fig. 1 and consists of three parts: 1) the backbone, 2) the neck, and 3) the head. The implementation details of each module in the proposed method are described. First, the general architecture of the network is described. Following that, the Revcol-C2f feature extraction network, the MASPPF module, and the ASFPN structure will be detailed, respectively.

A. General Architecture of AMRCNet

AMRCNet is the proposed adaptive multiscale reversible column network, where YOLOv8s is used as a baseline. The overall structure of the AMRCNet is illustrated in Fig. 1. AMRCNet consists of three parts: 1) the backbone, 2) the neck, and 3) the detection head. The Revcol-C2f backbone is proposed to alleviate information loss and acquire comprehensive multiscale
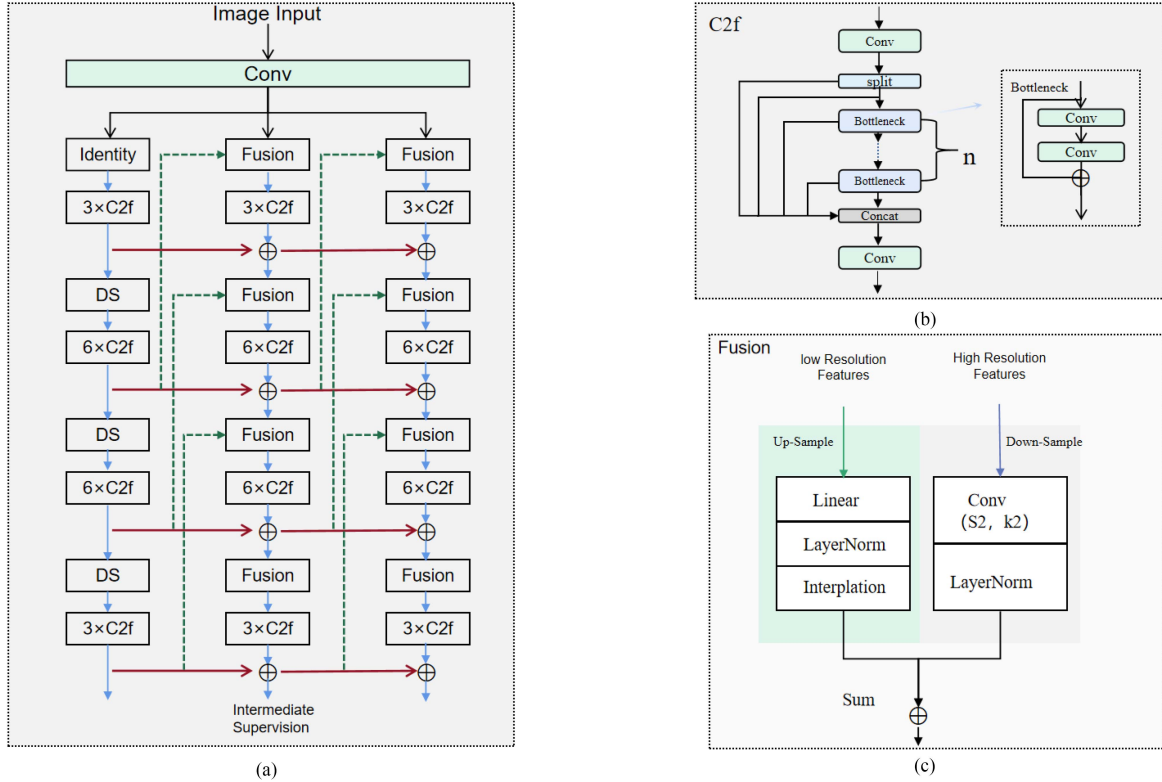
Fig. 2. (a) Revcol-C2f module (b) C2f module. (c) Fusion module. Red arrows represent reversible transformations, blue arrows denote feature extraction within the current subnetwork, and green arrows signify cross-subnetwork multiscale information embedding.

ship information in a complex environment, where Revcol is leveraged to reconstruct the YOLOv8 backbone network. The MASPPF module is introduced to augment the receptive field and mitigate the influence of background clutter after Revcol-C2f. The ASFPN neck is reconstructed by adding an ADM to obtain more adaptive multiscale ship information, which is based on the FPN and PAN [49] network structures. The detection head utilizes the decoupled head of YOLOv8, which processes classification and regression separately. In particular, the Revcol-C2f backbone is composed of four feature extraction layers denoted as Ln (where n ranges from 1 to 4), playing a pivotal role in extracting discriminative features. Thanks to the hierarchical feature extraction process, multiscale feature maps are obtained for each layer, where C2, C3, and C4 represent the outputs of the first three stages in the Revcol-C2f backbone. C5 is the feature maps obtained by MASPPF enhancement of the final output of Revcol-C2f. During the FPN fusion, intermediate feature maps, denoted as F1, F2, F3, and F4, are generated. These feature maps result from the fusion of intermediate levels within the feature pyramid, enabling the integration of information across different scales. Finally, the fused multilevel feature maps are generated, denoted as P1, P2, P3, and P4, which is beneficial to detect multiscale ships.

B. Features Extraction Backbone Revcol-C2f

In conventional SAR ship detection methods, the IB principle was widely applied. However, this approach can cause small and medium ship information to be lost or recognized as background in complex environments. On the contrary, disentangled feature learning may be more favorable for preserving this information. In this article, to obtain rich multiscale ship information in SAR images and mitigate the information loss, Revcol is utilized to reconstruct the YOLOv8s backbone, named Revcol-C2f backbone, as shown in Fig. 2(a). The multilevel feature extraction structure of YOLOv8s, specifically the C2f structure attributed to the CSPs structure, facilitates an improved capture of target information across various scales of ships. The use of reversible transformations is helpful in retaining large amounts of extracted information while mitigating information conflicts. In particular, the Revcol-C2f backbone combines embeddings from different layers and employs disentanglement mechanisms to address the potential information loss encountered when using a single-column network to extract features from deeper layers. In addition, Revcol-C2f promotes the fusion of deeper semantic information with shallow features, resulting in a more comprehensive and informative representation of the input data. Reversible transformation can be explained by the following formula:

$$\text{Forward}: x_t = F_t(x_{t-1}, x_{t-m+1}) + \gamma x_{t-m} \qquad (1)$$

$$\text{Inverse}: x_{t-m} = \gamma^{-1}\left[x_t - F_t(x_{t-1}, x_{t-m+1})\right] \qquad (2)$$

where t represents each subnetwork as the tth layer ($n \geq 1$), m represents the number of subnetworks ($m \geq 2$), F_t represents the $n \times$C2f module at the tth layer, γ denotes a fixed parameter set

to 0.5 in this article, and x_t represents the feature map of the tth layer in the network.

The detailed implementation process of Revcol-C2f is as follows:

1) To achieve consistent tensor dimensions with the down-sampled dimensions in YOLOv8s, a convolutional layer with a 4×4 kernel, a stride of 4, and no padding is employed.
2) According to the implementation of Revcol, the downsampled feature maps are subjected to an extraction network consisting of three subnetworks with reversible transformations.
3) The C2f module is introduced to extract the feature of the current subnetwork and the previous subnetwork, and the structure of the C2f module is shown in Fig. 2(b).
4) The first subnetwork is equivalent to the original feature extraction network in YOLOv8. In the second and third subnetworks, a fusion module is employed to combine the features from the previous subnetworks with the features of the current subnetwork, where features are extracted by C2f. This fusion process is illustrated in Fig. 2(c).
5) The information from the same layer in the previous subnetwork is embedded through reversible transformations.
6) To address the problem of information collapse, the intermediate supervision method proposed by RevCol is incorporated into the second subnetwork. Through this approach, the network receives feedback and gradients not only at the final output but also at multiple intermediate layers.

Feature fusion is accomplished by the fusion module through the upsampling of low-resolution features and the DS of high-resolution features, followed by a sum operation. DS refers to the process of reducing the spatial resolution of features using regular convolutional operations.

C. Multiplexed Adaptive Spatial Pyramid Pooling Layer

Due to the unique imaging mechanism of SAR, the effective feature information of ships can be affected by the clutter dispersed around the ship, leading to excessive semantic differences, especially for small-scale ships. According to attention mechanisms such as FESA [31], coordinate attention [50], and EMA [51], it can be demonstrated that pooling operations can effectively extract more relevant feature information in a local region. Moreover, the receptive field can be greatly enhanced by successive multiple identical kernel pooling operations in SPPF [52]. It is believed that utilizing different receptive fields to obtain a wider range of feature information can help mitigate the effects of scattered clutter, noise, and complex environments on the detection of ships in SAR images. However, the use of several fixed receptive fields in SPPF is prone to introducing background clutter. This approach potentially leads to the mixture of background noise features with ship features, resulting in blurred ship details. Therefore, multiple pooling layers with different kernels are added to obtain various scales of the receptive field. Subsequently, the receptive fields are enhanced by using multiplexed parallel pooling. Ultimately, the adaptive fusion of different

Fig. 3. Structure of MASPPF.

receptive fields is conducted to obtain more suitable features. The overall structure is referred to as the MASPPF, illustrated in Fig. 3. In particular, pooling operations with kernel sizes 5 and 7 are utilized for local feature extraction. The receptive fields of size 9×9 and 13×13 are acquired by the right branch after successive 5×5 pooling while the left branch acquires receptive fields of size 11×11 and 17×17, respectively, after successive 7×7 pooling. The process for the implementation of the adaptive fusion consists of three steps.

First, a learnable parameter w_n ($1 \leq n \leq 6$) is assigned for each feature map from the pooling layers. Second, the scoring scores are adaptively obtained by these learnable parameters. Finally, the scores are passed through the ReLU function to obtain the final adaptive scores, as shown in

$$\varepsilon_k = \text{Relu}\left(\frac{w_k}{\mu + \sum_{n=1}^{6} w_n}\right) \quad (3)$$

where w_n represents the learnable parameters, ε_k the adaptive score for the kth layer of the feature map, and k specifies the layer. μ represent a constant value of 0.0001, which is added to prevent a score of 1. Relu represents the activation function Relu.

The feature map of MASPPF for each layer can be described as follows:

$$X_1 = \text{CBS}(x) \quad (4)$$
$$X_2 = \text{MaxPool2d}_5(X_1) \quad (5)$$
$$X_3 = \text{MaxPool2d}_5(X_2) \quad (6)$$
$$X_4 = \text{MaxPool2d}_5(X_3) \quad (7)$$
$$X_5 = \text{MaxPool2d}_7(X_2) \quad (8)$$
$$X_6 = \text{MaxPool2d}_7(X_5) \quad (9)$$
$$X_{\text{out}} = \text{CBS}\left(\text{Cat}\left(\varepsilon_1 X_1, \varepsilon_2 X_2, \varepsilon_3 X_3, \varepsilon_4 X_4, \varepsilon_5 X_5, \varepsilon_6 X_6\right)\right) \quad (10)$$

where X is the input feature map, and the CBS consists of a full convolutional layer, BatchNorm2d, and the SiLU activation function. $\text{MaxPool2d}_5(.)$ and $\text{MaxPool2d}_7(.)$ represent the maximum pooling layers with kernel size 5, stride 1, padding 2 and kernel size 7, stride 1, padding 3, respectively. X_1 represents the output of CBS(.), and X_2, X_3, X_4, X_5, and X_6 represent the
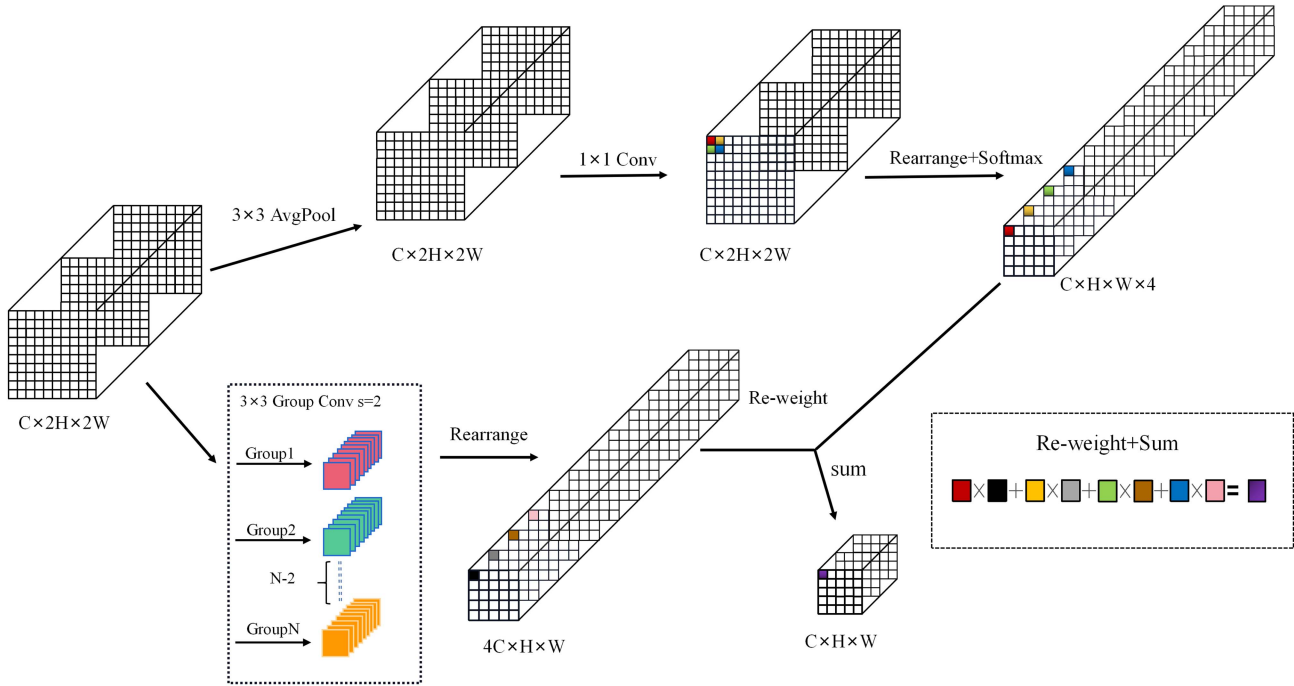
Fig. 4. Structure of ADM.

output of different pooling layers, respectively. X_{out} represents the final output of the MASPPF module.

D. Adaptive Sampling FPN

In SAR images, the scale difference of ship poses difficulty for existing feature fusion networks to effectively fuse multiscale ship features. In most existing FPNs, upsampling and DS operations are employed to modify the resolution of feature maps at different levels. However, as the number of upsampling and DS operations increases, the risk of losing valuable feature information rises, thereby diminishing the effectiveness of information fusion. Furthermore, the deficient correlation among features in individual layers subsequent to the fusion of shallow and deep feature maps through concatenation followed by convolutional processing hampers the detection performance of ships in SAR images. In order to alleviate these problems, an adaptive sampling feature pyramid network (ASFPN) is introduced, as shown in Fig. 1, to help the detection of multiscale ships in SAR images. In ASFPN, an ADM is introduced to alleviate information loss during the DS process. Furthermore, to address the introduction of irrelevant information resulting from upsampling, multiple bidirectional fusions are employed to effectively integrate the abundant shallow acquired from ADM and deep feature information from upsampling.

The implementation process of ADM is shown in Fig. 4 and the specific implementation process is as follows.

The ADM is divided into two components: 1) adaptive feature score acquisition and 2) sampling information fusion. In the first component, a kernel with a size of 3, stride of 1, and padding of 1 is applied to perform average pooling, which

helps alleviate the influence of the background. Subsequently, a fully convolutional layer is applied to capture the relationships between different channels and spatial locations. The generated feature maps are then divided into nonoverlapping regions of size 2×2. Furthermore, each region is transformed to increase its channel dimension by a factor of 4. Finally, applying the Softmax function to these transformed regions yields scores for each feature within the 2×2 kernel region, indicating their respective importance or relevance. In the second component, group convolution is first utilized (with the number of groups equal to the input channels divided by 16) to perform DS. The downsampled feature map is then resized to match the feature map obtained in the first component. Next, information weights are re-established by multiplying the obtained scores with the corresponding feature values from the downsampled feature map. Finally, the most relevant information is aggregated by summing the scores within each 2×2 window. By following the aforementioned steps, information loss during DS can be effectively reduced.

The feature map for each layer of the ADM can be described as follows:

$$P_1 = \text{Rearrange}\left(\text{Conv}_{1\times1}\left(\text{AvgPool}_{3\times3}(X_{\text{in}})\right)\right) \quad (11)$$

$$P_2 = \text{Rearrange}\left(\text{GConv}_{3\times3,2}(X_{\text{in}})\right) \quad (12)$$

$$X_{\text{out}} = \text{Sum}\left(\text{Reweight}\left(\text{Softmax}(P_1), P_2\right)\right) \quad (13)$$

where $\text{AvgPool}_{3\times3}$ represents the average pooling operation with kernel size 3, stride 1, padding 1, and $\text{Conv}_{1\times1}$ represents the full convolutional layer. $\text{GConv}_{3\times3,2}$ represents a group convolution operation with a kernel size of 3 and a stride of 2. The number of groups in the group convolution is equal to $C_{\text{in}}/16$,

TABLE I
OVERVIEW OF MULTISCALE SSDD DATASET INFORMATION

sensor	RadarSat-2,TerraSAR-X,Sentinel-1
Number of SAR image samples	1160
Average image size	500×500
Training and testing ratio	8:2
polarization	HH, VV, VH, HV
resolution	1 m — 15 m
location	Yantai, China, Visakhapatnam, India
Sea state	Good, Poor
scene	Complex onshore and offshore scenarios
Number of ships	2587
Minimum ship (width × height)	5×4
Largest ship (width × height)	180×308

where C_{in} represents the number of input feature map channels. Rearrange represents the special feature map adjustment operation, Softmax is the Softmax function, Reweight represents the multiplication to recapture the score weights, and Sum sums the obtained scores.

The feature map of each layer of ASFPN can be described as follows:

$$F_4 = f\left(\mathrm{Cat}\left(\mathrm{ADM}(C_4), \mathrm{MASPPF}(C_5)\right)\right) \quad (14)$$

$$F_3 = f\left(\mathrm{Cat}\left(\mathrm{ADM}(C_3), \mathrm{Ups}(F_4), C_4\right)\right) \quad (15)$$

$$F_2 = f\left(\mathrm{Cat}\left(\mathrm{ADM}(C_2), \mathrm{Ups}(F_3), C_3\right)\right) \quad (16)$$

$$P_1 = F_1 = f\left(\mathrm{Cat}\left(\mathrm{UpS}(F_2), C_2\right)\right) \quad (17)$$

$$P_2 = f\left(\mathrm{Cat}\left(\mathrm{Conv}_{3\times3,2}(P_1), F_2\right)\right) \quad (18)$$

$$P_3 = f\left(\mathrm{Cat}\left(\mathrm{ADM}(F_2), \mathrm{Conv}_{3\times3,2}(P_2), F_3\right)\right) \quad (19)$$

$$P_4 = f\left(\mathrm{Cat}\left(\mathrm{ADM}(F_3), \mathrm{Conv}_{3\times3,2}(P_3), F_4\right)\right) \quad (20)$$

where C_2, C_3, C_4, and C_5 represent the feature maps extracted from different layers of the backbone network, F_1, F_2, F_3, and F_4 represent intermediate features obtained from top-down and bottom-up paths, UpS represents the up-sampling operation, $\mathrm{Conv}_{3\times3,2}$ represents the full DS operation of a convolutional layer with convolutional kernel 3×3 and a step of 2, and ADM represents the adaptive sampling module. $f(.)$ represents 3 C2f feature fusion operations.

III. EXPERIMENT

The proposed method is validated on two public datasets: 1) the SSDD [47] and 2) HRSID [48]. In this section, the two datasets, along with the experimental setup, are briefly described in the first place. Then, evaluation metrics are presented. Finally, ablation experiments and comparison experiments are conducted.

A. Dataset Description

The two public datasets, SSDD and HRSID, include the data of nearshore and ocean-going ships from complex environments.

The overview of the SSDD dataset is presented in Table I. As shown in the table, the multiscale SSDD dataset provides 1160 SAR image samples from Radarsat-2, TerraSAR-X, and Sentinel-1 satellites. The average size of the images is about

(a)

(b)
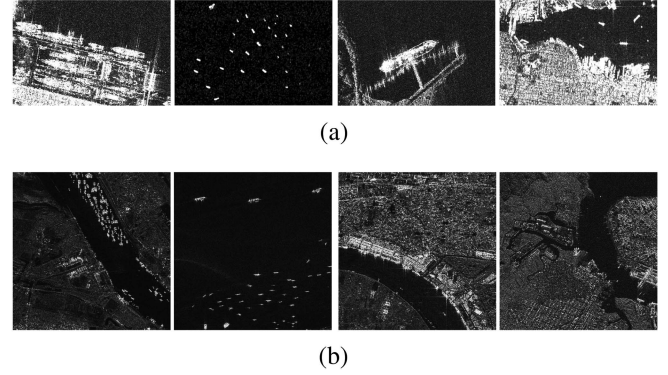
Fig. 5.   Sample images from two experimental datasets. (a) SSDD. (b) HRSID.

TABLE II
OVERVIEW OF COMPLEX HRSID DATASET INFORMATION

sensor	Sentinel-1, TerraSAR-X, TanDEM
Number of SAR image samples	5,604
Average image size	800×800
Training and testing ratio	13:7
polarization	HH, VV, VH, HV
resolution	0.5 m, 1 m, 3 m
location	Houston, São Paulo, Barcelona, etc.
Sea state	Good, Poor
scene	Complex onshore and offshore scenarios
Number of ships	16951
Minimum ship (width × height)	3×1
Largest ship (width × height)	800×653

500×500 pixels. Polarizations are HH, VV, VH, and HV. The resolution ranges from about 1 to 15 m. The environments of the ships in the dataset range from good to poor sea state, and from complex docking scenarios to simple offshore scenarios, as shown in Fig. 5(a). According to the statistics, the SSDD dataset comprises 2587 ships, with the smallest ship measuring 5 pixels in width and 4 pixels in height, occupying only 20 pixels. In contrast, the largest ship has dimensions of 180×308 pixels, occupying 55 440 pixels. It is clear that the largest size ship is 2772 times larger than the smallest ship, which indicates that the ship scales in the SSDD dataset vary greatly. Therefore, SSDD is very suitable for measuring the multiscale SAR ship detection performance of the network.

The HRSID dataset was published by Wei et al. [48] and the overview of information is shown in Table II. The HRSID dataset provides 5604 SAR image samples from Sentinel-1, TerraSAR-X, and TanDEM. These images have an average size of 800×800 pixels, polarization modes of HH, HV, and VV, and resolutions of 0.5 m, 1 m, and 3 m, and locations including Houston, São Paulo, Barcelona, Chittagong, Bangladesh, and other important international shipping lanes. The environments of the ships in the dataset range from good to poor sea state, and from complex docking scenarios to simple offshore scenarios, as revealed in Fig. 5(b). Compared with the multiscale SSDD dataset, the HRSID dataset contains a wider variety of complex scenarios, which makes the dataset suitable for measuring the model's performance in detecting SAR ships in complex scenarios.

TABLE III
MEANINGS OF COCO METRICS

Metric	Meaning
AP	IoU=0.50:0.05:0.95
AP50	IoU=0.50
AP75	IoU=0.75
APs	Area $< 32^2$
APm	$32^2 <$ Area $< 96^2$
APl	$96^2 <$ Area

TABLE IV
IMPACT OF INDIVIDUAL MODULES ON HRSID DATASET

Method	AP	AP50	AP75	APs	APm	APl
Baseline	64.2	88.9	73.3	52.9	79.5	30.9
Revcol-C2f	67.2	90.5	77.2	54.9	81.1	55.2
ASFPN	67.8	91.0	79.0	59.9	80.6	27.4
MASPPF	66.5	90.3	75.8	54.7	80.1	50.5

TABLE V
IMPACT OF INDIVIDUAL MODULES ON SSDD DATASET

Method	AP	AP50	AP75	APs	APm	APl
Baseline	68.8	97.4	81.5	67.6	73.9	65.9
Revcol-C2f	70.2	97.5	83.5	69.0	75.5	72.5
ASFPN	71.0	97.2	85.2	70.2	76.4	67.6
MASPPF	70.4	97.9	84.2	69.0	76.6	71.1

B. Hyperparameters and Environment Settings

The images in the SSDD and HRSID datasets are resized to 640×640 during input. All the experiments are implemented using PyTorch with hyperparameters fine-tuned based on the YOLO series. The network is trained using the stochastic gradient descent method with a learning rate of 0.01. The momentum is set to 0.937, and the weight decay is set to 0.0005. The model is trained for 300 epochs.

The software environment includes PyTorch 1.11.0, Python 3.8 (Ubuntu 20.04), and CUDA 11.3. The hardware environment consists of an Intel(R) Xeon(R) Gold 6330 CPU and an RTX 3090 with 24 GB memory. To enhance the diversity of the training dataset, mosaic augmentation is utilized during the training phase. In addition, the proposed model is trained from scratch rather than using pretrained models.

C. Evaluation Metrics

The object detection evaluation metrics in the COCO [53] dataset are utilized and detailed in Table III. It is summarized as follows.

AP: The primary metric for measuring the accuracy of algorithmic detection results. It is the average precision (AP) across different IoU thresholds ranging from 0.5 to 0.95. A higher AP indicates more accurate detection results.

AP50: The AP at an IoU threshold of 0.5. In many applications, 0.5 is a commonly used IoU threshold.

AP75: The AP at an IoU threshold of 0.75. In more stringent applications, a higher IoU threshold can better measure algorithm performance.

APs: The AP for small objects (area smaller than 32^2 pixels). Since small objects are usually more challenging to detect, this metric helps evaluate algorithm performance in such cases.

APm: The AP for medium objects (area between 32^2 and 96^2 pixels). This metric helps to evaluate algorithm performance in detecting medium objects.

APl: The AP for large objects (area larger than 96^2 pixels). This metric helps to evaluate algorithm performance in detecting large objects.

Precision and recall are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{21}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{22}$$

where TP (true positives), FP (false positives), and FN (false negatives) refer to the number of correct detections, false alarms, and missed targets, respectively.

AP is defined as

$$\text{AP} = \int_0^1 P(R)\, dR \tag{23}$$

where $P(R)$ is denoted as the curve of precision and recall, which can be defined as the area AP values contained in the $P(R)$ curve and axes will be calculated separately in this article depending on the IoU threshold.

D. Evaluations of the Proposed Method

In this section, the YOLOv8s is used as the baseline network, and ablation experiments are implemented to verify the effectiveness of the proposed modules. Comparison experiments are then conducted with the representative classical deep learning networks to verify the superiority of the proposed network.

1) Ablation Study: In AMRCNet, three methods (Revcol-C2f, ASFPN, and MASPPF) are proposed to enhance ship detection performance in SAR images. To validate the performance of AMRCNet, the experimental results of the designed modules are compared with the baseline. For the sake of objectivity and fairness, identical hyperparameters are used in the ablation experiment. The ablation experiment is divided into two parts. In the first part, the modules in the baseline network are replaced by the designed modules. In the second part, the designed modules incrementally are added to the baseline network.

Part 1. (a) Analysis of the effectiveness of Revcol-C2f: Revcol-C2f utilizes the hierarchical CSP structure in C2f to acquire rich multiscale ship information in various gradients. Subsequently, the ship information supplemented by Revcol is embedded with reversible transformation. Experiments are conducted to confirm the effectiveness of the proposed Revcol-C2f, and the results are shown in Tables IV and V. It is noteworthy that on the HRSID dataset, AP, AP50, AP75, APs, APm, and APl are improved by 3%, 1.6%, 3.9%, 2%, 1.6%, and 24.3%, respectively. On the
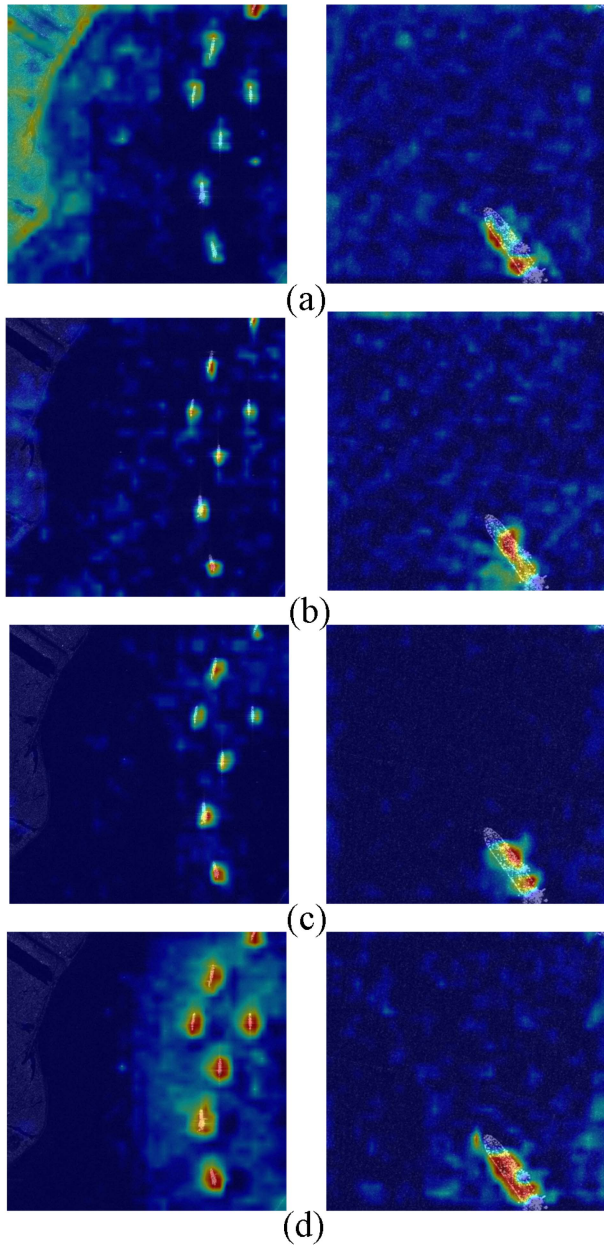
Method	AP	AP50	AP75	APs	APm	APl	Prams
SPPF	64.2	88.9	73.3	52.9	79.5	30.9	0.656 m
SPPCSPC	64.3	87.9	73.6	54.7	81.0	23.0	7.085 m
SPPFCSPC	65.2	89.9	74.2	52.9	79.5	48.4	7.085 m
MASPPF	66.5	90.3	75.8	54.7	80.1	50.5	0.919 m

shown in Figs. 7(c), 8(c), and 9(c). It is observed that the missed detection of ships has decreased.

(b) Analysis of the effectiveness of MASPPF: To validate the effectiveness of MASPPF, ablation experiments are conducted for analysis, and the results are presented in Tables IV and V. It can be observed that various metrics of the network exhibit effective improvements. AP75, being a more stringent object detection metric, reveals a notable increase in both SSDD and HRSID datasets, with improvements of 2.7% and 2.5%, respectively, compared to the baseline. This demonstrates MASPPFs ability to suppress the impact of clutter around ships. This is attributed to MASPPF employing multiplexed large-kernel pooling operations to capture comprehensive information about the ship hull within a region. Moreover, adaptive fusion is able to obtain receptive fields suitable for ships of different scales. This ultimately contributes to mitigating the impact of clutter. To verify the superior performance of MASPPF over similar methods, comparative experiments were conducted, introducing the most effective spatial pyramid methods, including SPPF [52], SPPCSPC [54], and SPPFCSPC [55]. Features were extracted by the YOLOv8s backbone network, and all experiments were conducted on the HRSID dataset. The results of the comparative experiments are shown in Table VI. It is evident that MASPPF achieves better results with minimal increase in parameters. In particular, MASPPF achieved a 1.6% AP75 improvement compared to the suboptimal SPPFCSPC. Furthermore, Grad-CAM [56] is used to display the heat map results of different methods, as revealed in Fig. 6. It is clearly shown that MASPPF is more sensitive to the specific location of ships and is able to obtain more valuable information, thereby mitigating the impact of clutters.

(c) Analysis of the effectiveness of ASFPN: ASFPN is an improved FPN, which is used to further enhance the performance of multiscale ships detection. Ablation experiments were conducted to demonstrate the effectiveness and superiority of ASFPN, and the results are presented in Tables V and VI. Compared to the FPN-PAN used in the baseline, there is a significant improvement in the HRSID dataset, with the evaluation metrics of AP, AP50, AP75, and APs increasing by 3.6%, 2.1%, 5.7%, and 7%, respectively. Likewise, for the SSDD dataset, the AP, AP50, AP75, APs, and APm increase by 2.2%, 3.7%, 2.6%, 2.5%, and 1.7%, respectively. The visual results of ASFPN are shown in Figs. 7(d) and 8(d), where the accuracy of detecting the small ships can be clearly observed. The previous experimental and visual results adequately demonstrate the effectiveness of ASFPN in multiscale ship detection. Within ASFPN, the ADM plays a crucial role in alleviating the acquisition of irrelevant information and mitigating information loss, leading to the

Fig. 6. Presentation of heat map for different spatial pyramid methods on HRSID dataset. (a) SPPF. (b) SPPCSPC. (c) SPPFCSPC. (d) MASPPF.

SSDD dataset, the gains in the AP, AP75, APs, APm, and APl are reaching 1.4%, 2%, 1.4%, 1.6%, and 6.6%, respectively. From the gains brought by APs, APm, and APl, it can be confirmed that Revcol-C2f captures information about ships at different scales. The improvements are mainly due to the fact that Revcol-C2f, which includes multiple subnetworks, enables the extraction of features at different hierarchical levels. Meanwhile, the improvements in AP, AP50, and AP75 can be attributed to the application of reversible transformation with a disentangled feature learning concept. This mechanism effectively alleviates the challenge of ship information loss, thereby contributing to the overall improvement in performance. To illustrate the experimental effect, the visualization results of Revcol-C2f are
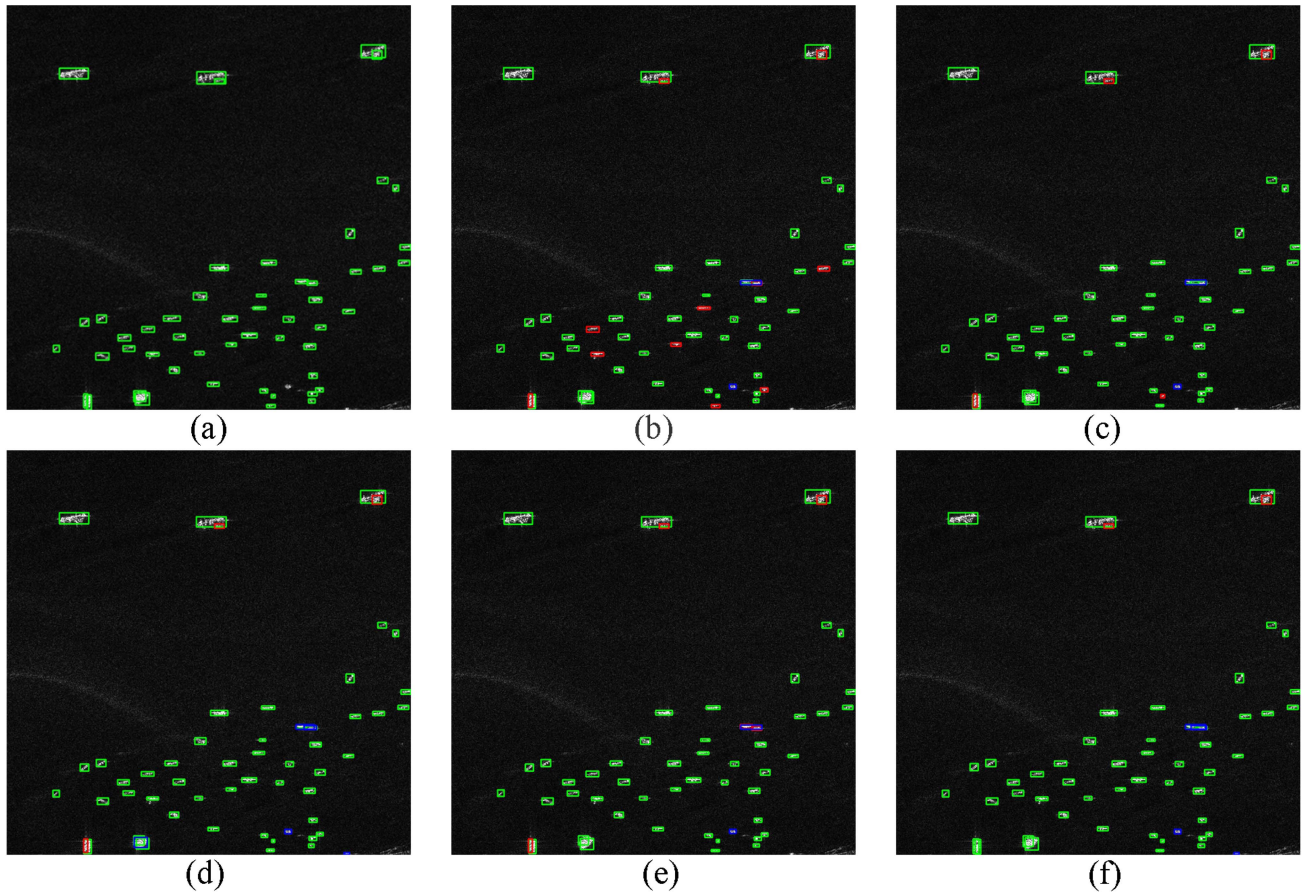
Fig. 7. Detection results for different methods in complex marine environments on the HRSID dataset. Green, blue, and red boxes represent TP, FP, and FN, respectively. (a), (b), (c), (d), (e), and (f) represent real boxes, YOLOv8s, Revcol, ASFPN, Revcol+ASFPN, and AMRCNet, respectively.

TABLE VII
DETECTION RESULTS OF DIFFERENT FPN-BASED METHODS ON HRSID DATASET

Method	AP	AP50	AP75	APs	APm	APl
FPN-PAN	64.2	88.9	73.3	52.9	79.5	30.9
RepGFPN	67.1	90.9	77.7	58.9	80.4	27.9
ASFPN	67.8	91.0	79.0	59.9	80.6	27.4

TABLE VIII
DETECTION RESULTS OF MULTIPLE METHODS MIXED ON HRSID DATASET

Method	AP	AP50	AP75	APs	APm	APl
Baseline	64.2	88.9	73.3	52.9	79.5	30.9
+Revcol-C2f	67.2	90.5	77.2	54.9	81.1	55.2
+ASFPN	68.9	91.9	79.6	59.7	80.8	38.9
+MASPPF	69.8	92.7	80.7	60.6	81.0	52.5

TABLE IX
DETECTION RESULTS OF MULTIPLE METHODS MIXED ON SSDD DATASET

Method	AP	AP50	AP75	APs	APm	APl
Baseline	68.8	97.4	81.5	67.6	73.9	65.9
+Revcol-C2f	70.2	97.5	83.5	69.0	75.5	72.5
+ASFPN	70.3	97.5	84.2	69.3	77.0	71.7
+MASPPF	71.1	98.1	84.3	69.5	76.9	68.4

improvement of AP75. This is due to ADM that employs an attention-based DS operation, ensuring the preservation of information from ships. Furthermore, the combination of ADM and bidirectional information fusion operations effectively integrates feature information from ships at different scales, resulting in the improvement in APs, APm, and APl. To verify the effectiveness of ADM, experiments were conducted by adding ADM to RepGFPN. Comparison experiments were performed on the HRSID dataset with PAN-FPN [49] and RepGFPN [57]. The results, shown in Table VII, show that the AP75 metrics are improved by 5.7% and 1.3%, respectively, compared to SPPF and RepGFPN. This proves that ADM can capture more relevant information during DS, which enables it to achieve better performance in more demanding environments.

Part 2: For the purpose of verifying the effect of fusing several modules, Revcol-C2f, MASPPF, and SAFPN are added to the network in the order of the backbone network, neck, and MASPPF sequentially, and the results are shown in Tables VIII and IX. By incorporating both Revcol-C2f and ASFPN into the network simultaneously, improvements are observed in experimental results. Specifically, on the HRSID dataset, the
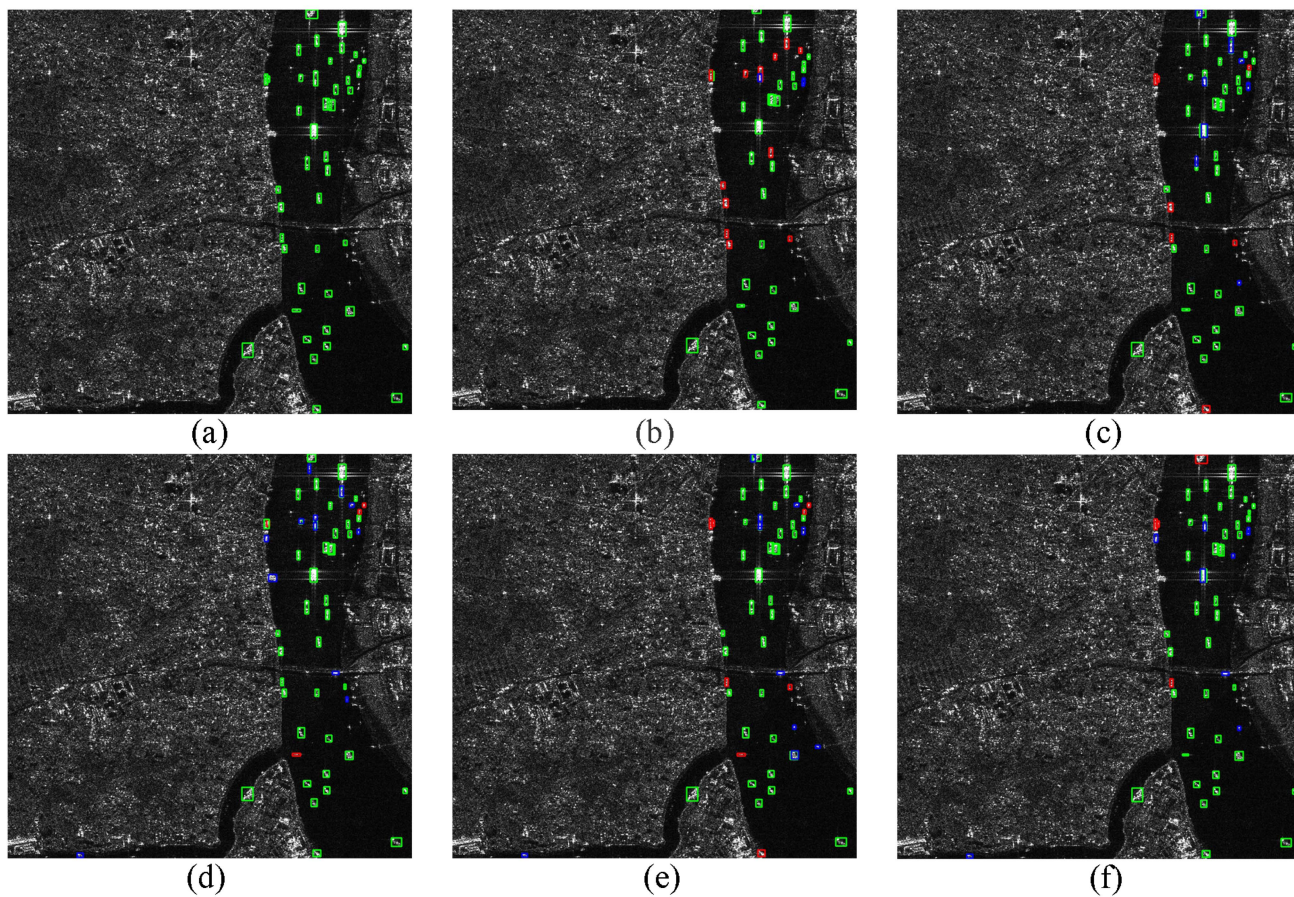
Fig. 8. Detection results for different methods in complex inland environments on the HRSID dataset. Green, blue, and red boxes represent TP, FP, and FN,respectively. (a), (b), (c), (d), (e), and (f) represent real boxes, YOLOv8s, Revcol, ASFPN, Revcol+ASFPN, and AMRCNet, respectively.
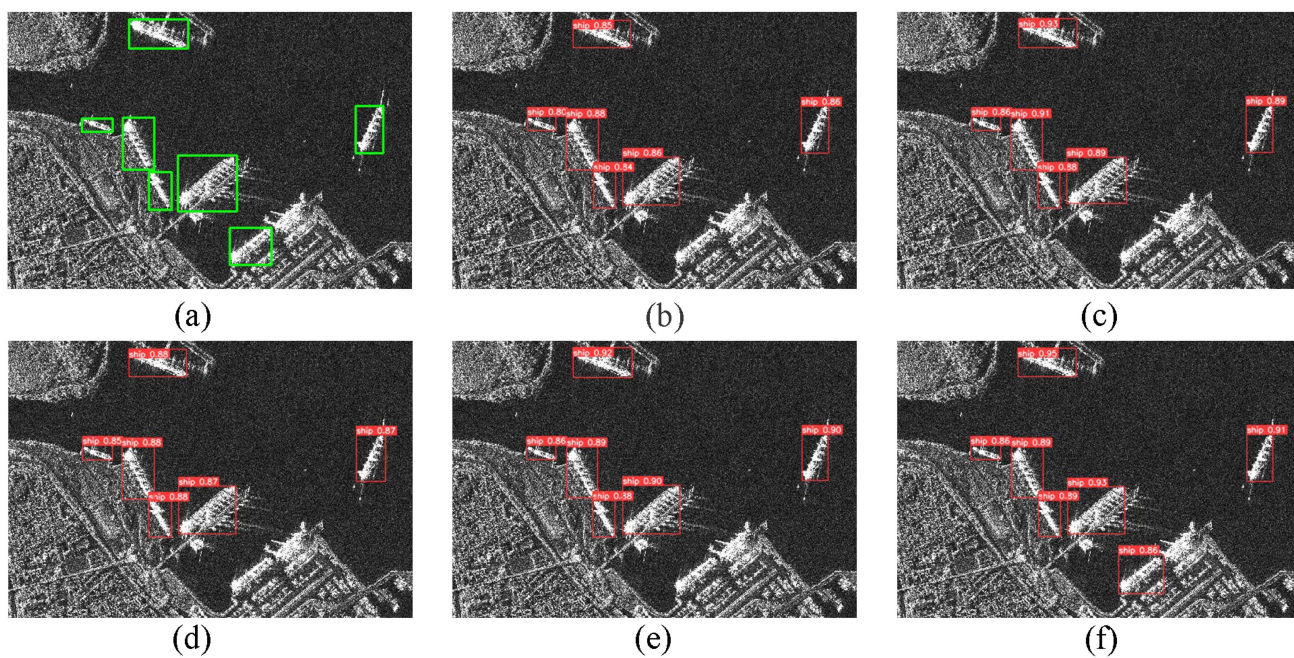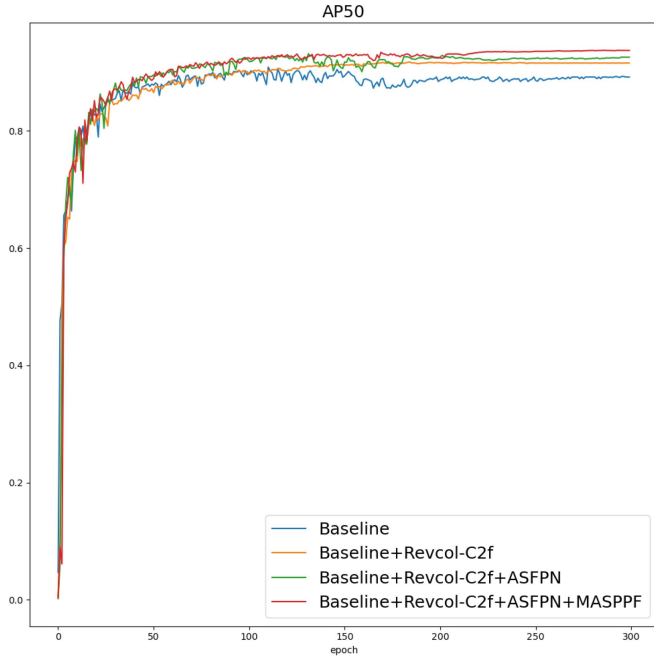

Fig. 9. Detection results of harsh inshore noise environments on the SSDD dataset. (a), (b), (c), (d), (e), and (f) represent real boxes, YOLOv8s, Revcol, ASFPN, Revcol+ASFPN, and AMRCNet, respectively.

Fig. 10. AP50 curves of ablation experiments on the SSDD dataset.
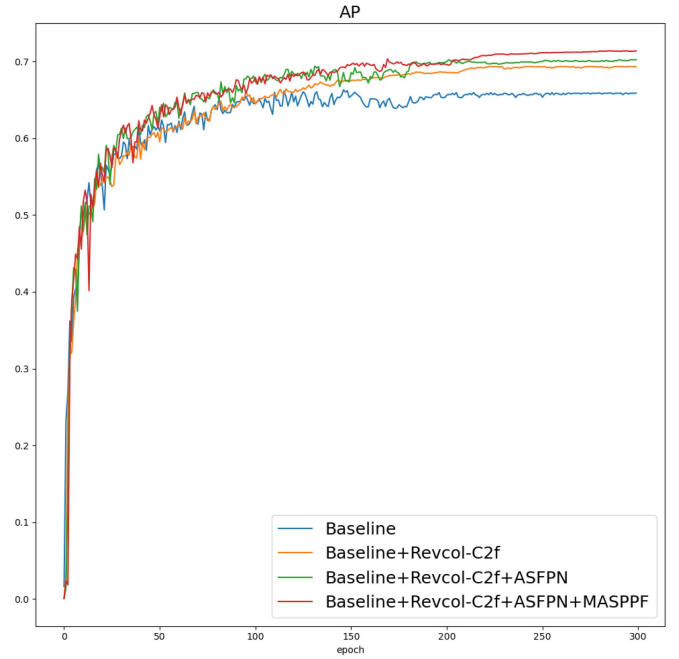
Fig. 11. AP curves of ablation experiments on the HRSID dataset.

metrics for AP, AP50, AP75, and APs reach 68.9%, 91.9%, 79.6%, and 59.7%, respectively. The overall performance is superior, especially in small-object detection, compared to using the Revcol-C2f module alone. Furthermore, improvements can be found when switching to the SSDD dataset. When MASPPF is added to complete AMRCNet, AP, AP50, AP75, APs, APm, and APl are improved by 0.9%, 0.8%, 1.1%, 0.9%, 0.2%, and 13.6%, respectively, on the HRSID dataset. For a more intuitive illustration of the effect of each proposed module, the ship detection results of different combinations are separately displayed on the HRSID dataset, which encounters complex marine environments, as shown in Figs. 7 and 8. It is shown that the leakage detection rate is significantly reduced after the addition of each module compared to the baseline. On the SSDD dataset, the visualization of the detection results for various methods is also shown in Fig. 9. As can be seen from the visualized images, the proposed combination of different modules results in higher ship scores. Moreover, based on the experimental results, Figs. 10 and 11 were plotted to show the comparison of ablation experiments on AP50 and AP in each epoch. It can be seen that the proposed methods significantly outperform the baseline in the assessment criteria. Furthermore, a comparison of the precision–recall curves of the two datasets used in the ablation experiments is shown in Figs. 12 and 13. The curves show that stacking the three methods also leads to increases in the included area of the curves. These experiments validate the effectiveness of the proposed modules in mitigating various challenges.

2) Comparative Experiment: This section presents a comparative performance evaluation of the proposed AMRCNet against several state-of-the-art methods. Target detection methods with excellent performance have been selected. These selected methods include one-stage network, two-stage network,
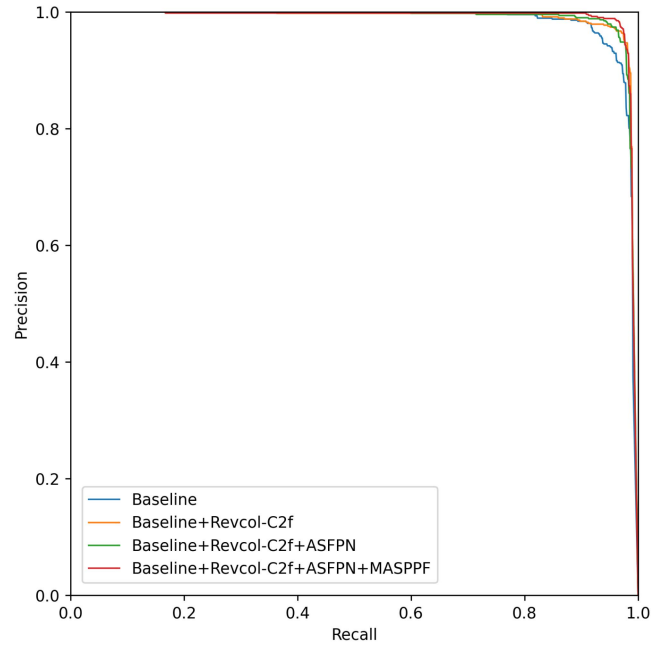

Fig. 12. PR curves of ablation experiments on the SSDD dataset.

transformer-based target detection network, and anchor-free network to ensure coverage of different types. On the SSDD dataset, the selected include Faster R-CNN [20], FoveaBox [21], MSSDNet [27], FEPS-Net [30], HRSDNet [48], YOLOv5s [52], FCOS [58], and Deformable DETR [59]. It is obvious to observe that AMRCNets AP, AP50, AP75, APs, and APm have surpassed other methods, as revealed in Table X.

Specifically, the AP(71.1%) and AP75(84.3%) for AMRC-Net are superior to other methods. This indicates that even in

TABLE X
RESULTS OF DIFFERENT METHODS ON THE SSDD DATASET

Method	AP	AP50	AP75	APs	APm	APl	Param (M)	Inference Time (ms)
FoveaBox	52.0	86.5	58.6	43.6	65.0	62.0	36.01	23.0
Faster R-CNN	54.4	85.2	61.1	47.2	66.1	67.4	41.12	21.5
Deformable DETR	55.7	92.2	61.6	48.8	68.2	68.2	40.00	40.5
HRSDNet	55.7	90.7	0.603	46.7	67.5	63.5	37.20	47.6
FCOS	56.0	91.9	61.7	49.3	67.1	68.4	31.84	19.8
MSSDNet	61.1	95.6	70.9	55.4	70.0	70.4	12.95	24.2
FEPS-Net	59.9	96.0	67.5	55.1	68.2	70.6	37.31	31.7
YOLOv5s	69.2	97.5	80.5	67.5	75.3	66.9	7.12	10.9
AMRCNet	71.1	98.1	84.3	69.5	76.9	68.4	18.73	23.6
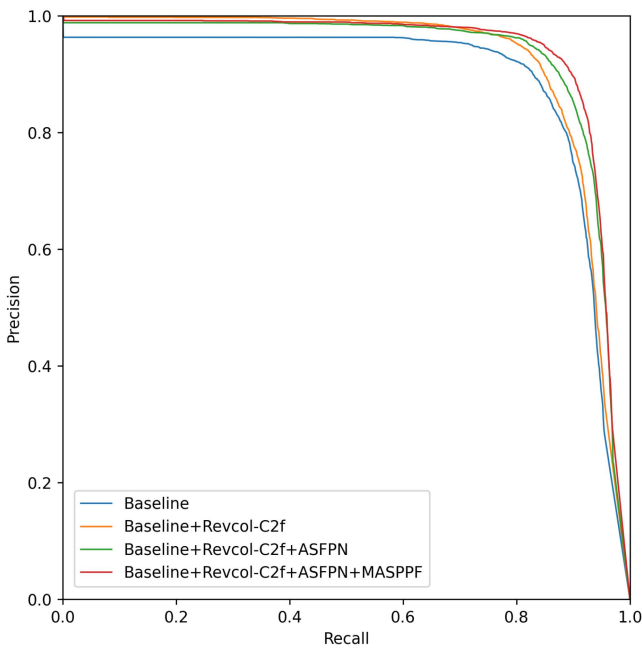
Fig. 13.   PR curves of ablation experiments on the HRSID dataset.

more challenging evaluation environments, AMRCNet is able to accurately locate and detect ships in SAR images. Despite the fact that the SSDD dataset itself has a wide range of SAR ship scales, the method achieves better multiscale detection results than other methods in terms of APs, APm, and APl. The improvement indicates that AMRCNet is more applicable to the task of multiscale SAR ship target detection.

In order to measure the model's performance for SAR ship detection under complex scenarios as well as to validate the robustness and generalization ability, comparative experiments with other state-of-the-art methods are conducted on the HRSID dataset. The experimental results are shown in Table XI. Experimental methods include Faster R-CNN, FoveaBox, Libra R-CNN [60], Deformable DETR, HRSDNet, FCOS, Cascade R-CNN [61], and FEPS-Net. The results consistently demonstrate that the proposed method outperforms other state-of-the-art methods across various evaluation metrics. On the HRSID dataset, the proposed method achieves the highest AP (69.8%) and AP75 (80.7%) among the compared methods, indicating

its superior robustness and generalization capabilities in complex scenarios. Moreover, the proposed method achieves a well-balanced performance in terms of APs, APm, and APl, showcasing its strong detection ability across ships of different scales. Notably, APm and APl reach 81.0% and 52.5% respectively. Although APs are slightly lower compared to suboptimal FEPS-Net, the overall improvements in AP and other metrics are significant, especially in inference Time. In Tables X and XI, it can be clearly observed that the proposed method outperforms Deformable DETR, which is based on the transformer architecture, across various evaluation metrics. This is primarily attributed to the proposed modules, which aim to mitigate the impact of background clutter and make effective use of local information and multilevel feature representation. Typically, anchor-free networks, such as FoveaBox and FCOS in Tables X and XI, face difficulties in achieving precise target localization. However, it is noteworthy that the AP of the proposed method much is higher than that of FoveaBox and FCOS, which means that the shortcomings of the traditional anchor-free methods for reduced localization accuracy can be effectively addressed. In addition, compared to traditional two-stage networks such as Faster R-CNN, Libra R-CNN, and Cascade R-CNN, the proposed uses only half the number of parameters and achieves better performance. YOLOv5s serves as a lightweight model with a short inference time (10.9 ms) and a relatively small number of parameters (7.12 million), but the proposed method outperforms across all key metrics. These experimental results emphasize that the proposed method strikes a commendable balance between speed and parameter efficiency while maintaining superior detection performance.

IV. DISCUSSION

In this section, the analysis of the performance of Revcol-C2f, MASPPF, ASFPN, and the overall network are provided with the aid of experimental results mentioned before.

A. Reversible Column Network With C2f for Backbone Extraction (Revcol-C2f)

The proposed Revcol-C2f is primarily designed to alleviate the issue of information loss during feature extraction and enhance the extraction capability of multiscale target information. The experimental results in Tables V and VI demonstrate the

TABLE XI
RESULTS OF DIFFERENT METHODS ON HRSID DATASET

Method	AP	AP50	AP75	APs	APm	APl	Param (M)	Inference Time (ms)
Faster R-CNN	56.0	79.1	63.6	56.6	60.6	13.4	41.12	43.5
FoveaBox	54.9	79.4	61.9	55.7	62.6	30.9	36.01	41.3
Libra R-CNN	55.6	77.4	63.6	56.0	61.4	16.2	41.39	49.5
Deformable DETR	45.8	71.9	51.6	45.7	52.5	10.1	40.00	72.4
HRSDNet	57.3	81.8	63.4	58.3	61.5	24.4	37.2	65.4
FCOS	58.9	87.0	65.9	60.7	63.4	25.2	31.84	37.3
Cascade R-CNN	59.1	80.5	67.1	59.9	63.2	20.1	53.6	43.5
FEPS-Net	65.7	90.7	74.3	66.8	65.2	31.6	37.31	94.4
AMRCNet	69.8	92.7	80.7	60.6	81.0	52.5	18.73	25.0

superiority of Revcol-C2f over the IB-based method for feature extraction backbone. Furthermore, Figs. 7(c), 8(c), and 9(c) show more accurate detection results for targets of different scales in different environments. The improvement is attributed to the adoption of the disentangled feature learning concept and multisubnetworks for feature extraction. The reversible transformation in Revcol with the disentangled feature learning concept preserves more target information and mitigates the issue of information loss. Moreover, multisubnetwork feature extraction is able to capture richer feature information for multiscale ships with the help of reversible transformation.

B. Multiplexed Adaptive Spatial Pyramid Pooling

MASPPF has proved to be an effective solution for background clutters through ablation analysis presented in Fig. 4 and Table V. It is primarily attributed to the implementation of multiplexed large-kernel pooling operations and adaptive fusion. Multiplexed large-kernel pooling operations capture richer receptive fields of different sizes, catering to different scales of ships. Meanwhile, the adaptive fusion process selects more suitable receptive fields for targets, effectively excluding background clutter. Compared to other similar methods, MASPPF demonstrates outstanding performance, as shown in Table VI and Fig. 6.

C. Adaptive Sampling FPN

This method is specifically designed to further meet the requirements of multiscale target feature fusion. In SAR images, the presence of significant noise and weak semantic information makes it challenging to perform effective information extraction through ordinary DS operations and information fusion without losing small-scale information and introducing noise. ASFPN addresses these challenges by first utilizing ADM to mitigate the loss of accurate target information. Subsequently, multiple bidirectional information fusion operations are employed to overcome scale differences. The effectiveness of ASFPN is demonstrated by ablation experiment results, as revealed in Tables IV and V. In comparison to other FPN-based methods, the proposed methods emphasize the detection of small and medium targets. The improvements in APs, APm, and AP75 in Table VII further attest to the effectiveness of ADM in reducing

information loss during the DS. Visual results in Figs. 7(d), 8(d), and 9(d) support these conclusions.

D. Overall Network

The overall experimental results in Tables X and XI demonstrate the effectiveness of the proposed method for ship detection in SAR images comprehensively. In comparison to anchor-based methods, such as YOLOv5s, MSSDNet, and HRSDNet, the proposed methods achieve superior overall detection results. When compared to anchor-free methods such as FCOS and FoveaBox, the issues of target omission and inaccurate positioning caused by the absence of predefined anchors are mitigated through the use of the disentangled feature learning concept and ASFPN. Compared to traditional two-stage methods such as Faster R-CNN, Libra R-CNN, and Cascade R-CNN, the proposed method is more lightweight and exhibits better detection performance. Relative to transformer-based methods such as Deformable DETR, the proposed method is more effective for multiscale ship detection, especially small-scale ships. Overall, the proposed method achieves a more effective balance between computational complexity and detection performance.

V. CONCLUSION

In this article, an integrated approach enclosing Revcol-C2f, MASPPF, and ASFPN is proposed to address challenges related to the information loss caused by sampling, impact of complex background clutter, and multiscale ship detection from SAR images. Specifically, the Revcol-C2f module reconfigures the YOLOv8 backbone using a reversible column network, effectively preserving target information and improving ship detection accuracy. Subsequently, the proposed MASPPF module mitigates the influence of background clutter. In addition, to address the issues of scale difference and semantic information loss, the ASFPN module fuses shallow and deep feature maps to enhance multiscale target detection capability at different stages. Ultimately, these modules are integrated into the AMRCNet architecture, demonstrating excellent SAR ship detection performance in complex environments, and mitigating the negative impact of scattering noise. Experimental results validate the effectiveness of the proposed method, showing improved detection performance for multiscale ship targets and reduced false and missed alarms in complex background clutter. In particular, the

AP reaches 71.1% and 69.8% on the SSDD and HRSID datasets, respectively.

Future work in SAR ship detection can prioritize real-time implementation, explore multimodal fusion techniques, and expand the application of transfer learning to enhance overall detection performance and promote generalization across radar domains.

REFERENCES

[1] Q. Sun, M. Liu, S. Chen, F. Lu, and M. Xing, "Ship detection in SAR images based on multilevel superpixel segmentation and fuzzy fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5206215.

[2] C. Qin, X. Wang, G. Li, and Y. He, "A semi-soft label-guided network with self-distillation for SAR inshore ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jul. 2023, Art. no. 5211814.

[3] B. Pan, Z. Xu, T. Shi, T. Li, and Z. Shi, "An imbalanced discriminant alignment approach for domain adaptive SAR ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Aug. 2023, Art. no. 5108111.

[4] G. Gao, L. Yao, W. Li, L. Zhang, and M. Zhang, "Onboard information fusion for multisatellite collaborative observation: Summary, challenges, and perspectives," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 2, pp. 40–59, Jun. 2023.

[5] L. Chen, R. Luo, J. Xing, Z. Li, Z. Yuan, and X. Cai, "Geospatial transformer is what you need for aircraft detection in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5225715.

[6] F. Ma, X. Sun, F. Zhang, Y. Zhou, and H.-C. Li, "What catch your attention in SAR images: Saliency detection based on soft-superpixel lacunarity cue," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Dec. 2023, Art. no. 5200817.

[7] D. Marzi and P. Gamba, "Inland water body mapping using multitemporal sentinel-1 SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11789–11799, Dec. 2021, doi: 10.1109/JSTARS.2021.3127748.

[8] R. Shang, M. Liu, L. Jiao, J. Feng, Y. Li, and R. Stolkin, "Region-level SAR image segmentation based on edge feature and label assistance," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5237216.

[9] F. Gao et al., "Cross-modality features fusion for synthetic aperture radar image segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 61, Aug. 2023, Art. no. 5214814.

[10] Y. Guan et al., "Fishing vessel classification in SAR images using a novel deep learning model," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Sep. 2023, Art. no. 5215821.

[11] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-Net in U-Net for infrared small object detection," *IEEE Trans. Image Process.*, vol. 32, pp. 364–376, Dec. 2023.

[12] T. Wu et al., "MTU-Net: Multilevel TransUNet for space-based infrared tiny ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jan. 2023, Art. no. 5601015.

[13] L. Ying, D. Miao, and Z. Zhang, "3WM-AugNet: A feature augmentation network for remote sensing ship detection based on three-way decisions and multigranularity," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Sep. 2023, Art. no. 1001219.

[14] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[15] J. Ai, Z. Cao, Y. Mao, Z. Wang, F. Wang, and J. Jin, "An improved bilateral CFAR ship detection algorithm for SAR image in complex environment," *J. Radars*, vol. 10, no. 4, pp. 499–515, Mar. 2021.

[16] X. Leng, K. Ji, K. Yang, and H. Zou, "A bilateral CFAR algorithm for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1536–1540, Jul. 2015.

[17] H. Dai, L. Du, Y. Wang, and Z. Wang, "A modified CFAR algorithm based on object proposals for ship target detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1925–1929, Dec. 2016.

[18] T. Li, Z. Liu, R. Xie, and L. Ran, "An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 184–194, Jan. 2017.

[19] C. Wang, F. Bi, W. Zhang, and L. Chen, "An intensity-space domain CFAR method for ship detection in HR SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 529–533, Apr. 2017.

[20] T.-Y. Lin, P. Dollr, 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.

[21] J. Jiao et al., "A densely connected end-to-end neural network for multiscale and multiscene SAR ship detection," *IEEE Access*, vol. 6, pp. 20881–20892, 2018.

[22] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.

[23] T. Zhang, X. Zhang, and X. Ke, "Quad-FPN: A novel quad feature pyramid network for SAR ship detection," *Remote Sens.*, vol. 13, no. 14, Jul. 2021, Art. no. 2771.

[24] D. Li, Q. Liang, H. Liu, Q. Liu, H. Liu, and G. Liao, "A novel multidimensional domain deep learning network for SAR ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5203213.

[25] K. Zhou, M. Zhang, H. Wang, and J. Tan, "Ship detection in SAR images based on multi-scale feature extraction and adaptive feature fusion," *Remote Sens.*, vol. 14, no. 3, Feb. 2022, Art. no. 755.

[26] L. Yu, H. Wu, Z. Zhong, L. Zheng, Q. Deng, and H. Hu, "TWC-Net: A. SAR ship detection using two-way convolution and multiscale feature mapping," *Remote Sens.*, vol. 13, no. 13, pp. 2558–2558, Jun. 2021.

[27] M. Zhao, X. Zhang, and A. Kaup, "Multitask learning for SAR ship detection with Gaussian-mask joint segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Aug. 2023, Art. no. 5214516.

[28] L. Zhang, Y. Liu, W. Zhao, X. Wang, G. Li, and Y. He, "Frequency-adaptive learning for SAR ship detection in clutter scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5215514.

[29] S. Wang, Z. Cai, and J. Yuan, "Automatic SAR ship detection based on multifeature fusion network in spatial and frequency domains," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 4102111.

[30] L. Bai, C. Yao, Z. Ye, D. Xue, X. Lin, and M. Hui, "Feature enhancement pyramid and shallow feature reconstruction network for SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1042–1056, Jan. 2023.

[31] Y. Zhou, H. Liu, F. Ma, Z. Pan, and F. Zhang, "A sidelobe-aware small ship detection network for synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5205516.

[32] 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, Mar. 2023, Art. no. 5203922.

[33] Y. Jiang, W. Li, and L. Liu, "R-CenterNet+: Anchor-free detector for ship detection in SAR images," *Sensors*, vol. 21, no. 17, Aug. 2021, Art. no. 5693.

[34] Q. Hu, S. Hu, and S. Liu, "BANet: A balance attention network for anchor-free ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5222212.

[35] G. Gao, Y. Dai, X. Zhang, D. Duan, and F. Guo, "ADCG: A cross-modality domain transfer learning method for synthetic aperture radar in ship automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5109114.

[36] R. Xia et al., "CRTransSar: A visual transformer based on contextual joint representation learning for SAR ship detection," *Remote Sens.*, vol. 14, no. 6, 2022, Art. no. 1488.

[37] 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.

[38] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, and J. Shi, "FoveaBox: Beyond anchor-based object detection," *IEEE Trans. Image Process.*, vol. 29, pp. 7389–7398, Jun. 2020.

[39] D. Hong et al., "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sens. Environ.*, vol. 299, 2023, Art. no. 113856.

[40] D. Hong et al., "SpectralGPT: Spectral foundation model," 2024, *arXiv:2311.07113*.

[41] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," 2000, *arXiv:physics/0004057*.

[42] N. Tishby and N. Zaslavsky, "Deep learning and the information bottleneck principle," in *Proc. IEEE Inf. Theory Workshop*, 2015, pp. 1–5, doi: 10.1109/ITW.2015.7133169.

[43] G. Desjardins, A. Courville, and Y. Bengio, "Disentangling factors of variation via generative entangling," 2012, *arXiv:1210.5474*.

[44] G. Hinton, "How to represent part-whole hierarchies in a neural network," *Neural Comput.*, vol. 35, no. 3, pp. 413–452, 2023.

[45] Y. Cai et al., "Reversible column networks," in *Proc. Int. Conf. Learn. Representation*, 2023. [Online]. Available: https://openreview.net/forum?id=Oc2vlWU0jFY

[46] J. Glenn, "Ultralytics YOLOv8," GitHub, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[47] T. Zhang et al., "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3690.

[48] 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.

[49] Z. Zhang, K. Zhang, Z. Li, and Y. Qiao, "PANet: Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8759–8768.

[50] 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. 13713–13722.

[51] 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.

[52] J. Glenn, "YOLOv5 release v6.1" GitHub, Accessed: Jan. 1, 2022. [Online]. Available: https://github.com/ultralytics/yolov5/releases/tag/v6.1

[53] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[54] 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.

[55] C. Li et al., "YOLOv6 v3.0: A full-scale reloading," 2023, *arXiv:2301.05586*.

[56] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[57] X. Xu, Y. Jiang, W. Chen, Y. Huang, Y. Zhang, and X. Sun, "DAMO-YOLO: A report on real-time object detection design," 2022, *arXiv:2211.15444v2*.

[58] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9626–9635.

[59] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," 2020, *arXiv:2010.04159*.

[60] J. Pang et al., "Libra R-CNN: Towards balanced learning for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 821–830.

[61] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.

Tianxiang Wang received the Bachelor of Engineering degree in software engineering from Yibin University, Yibin, China, in 2022. He is currently working toward the Master of Engineering degree in software engineering with the School of Computer and Information Engineering, Hubei University.

His research interests include computer vision and remote-sensing image target detection and recognition.

Zhangfan Zeng received the Bachelor of Science (B.S.) degree in electrical and information engineering from Wuhan University, Wuhan, China, in 2006, the Master of Science (M.S.) degree in communications engineering from The University of Manchester, Manchester, U.K., in 2007, and the Ph.D. degree in communications engineering from the University of Birmingham, Birmingham, U.K., in 2013.

From 2008 to 2009, he was an Algorithm Engineer with Guangdong Nortel Network, China. From 2013 to 2014, he was a ProtocolStack Software Engineer with NextG-Com, Ltd., Staines-upon-Thames, U.K. From January 2015 to September 2015, he was a Senior Engineer with Cobham Wireless, Stevenage, U.K. Since December 2015, he has been a full-time Associate Professor with Hubei University, Wuhan. His research interests include wireless communication and digital signal processing.