Multigranularity-Aware Network for SAR Ship Detection in Complex Backgrounds

Li Ying, Yizhang Liu, Zhifei Zhang, *Member, IEEE*, and Duoqian Miao

***Abstract*— Synthetic aperture radar (SAR) is a vital tool for ship detection, as it acquires high-resolution remote sensing images when optical images cannot penetrate. However, two primary challenges confronting SAR ship detection are complex backgrounds with islands, clutter, and land, as well as diverse scales of ship targets, particularly small ones, leading to numerous missed detections and false alarms. To overcome these challenges, we propose a multigranularity-aware network (MGA-Net). Specifically, we design a multigranularity hybrid feature fusion module (MGHF^{2}M) to extract more representative local detail and global semantic information, enhancing the model's capability to represent ship features to adapt to complex backgrounds. In addition, we design a multigranularity feature synergy enhancement module (MGFSEM), which uses depthwise separable convolutions with different kernel sizes to extract features at different granularities and retain the original features, significantly improving the model's representation of ship features at different scales. Experimental results show that our MGA-Net achieves the highest mAP and $F1$-score, surpassing eight advanced methods on three public datasets.**

***Index Terms*— Complex backgrounds, multigranularity, ship detection, synthetic aperture radar (SAR).**

I. Introduction

With the continuous development of maritime strategy, maritime ship targets play a crucial role in marine monitoring and maritime management. Compared with other sensors, synthetic aperture radar (SAR) has the advantage of being all-weather, all-day, and unaffected by natural factors [1], making it an essential tool for maritime ship detection.

Manuscript received 29 November 2023; revised 28 December 2023; accepted 8 January 2024. Date of publication 15 January 2024; date of current version 17 May 2024. This work was supported in part by the National Key Research and Development Program under Grant 2022YFB3104700; in part by the National Natural Science Foundation of China under Grant 62376198, Grant 61906137, and Grant 62076040; in part by China National Scientific Seafloor Observatory; in part by the Natural Science Foundation of Shanghai under Grant 22ZR1466700; and in part by the Interdisciplinary Project in Ocean Research of Tongji University. *(Li Ying and Yizhang Liu contributed equally to this work.) (Corresponding author: Zhifei Zhang.)*

Li Ying and Zhifei Zhang are with the Department of Computer Science and Technology, Tongji University, Shanghai 201804, China, and also with the Project Management Office of China National Scientific Seafloor Observatory, Tongji University, Shanghai 200092, China (e-mail: 1910663@tongji.edu.cn; zhifeizhang@tongji.edu.cn).

Yizhang Liu is with the School of Software Engineering, Tongji University, Shanghai 201804, China (e-mail: lyz8023lyp@gmail.com).

Duoqian Miao is with the Department of Computer Science and Technology, Tongji University, Shanghai 201804, China (e-mail: dqmiao@tongji.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/LGRS.2024.3352633, provided by the authors.

Digital Object Identifier 10.1109/LGRS.2024.3352633

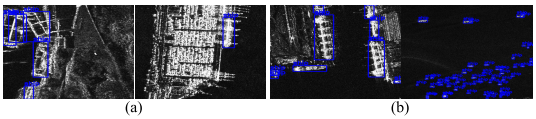


Fig. 1. Examples of SAR images in HRSID include (a) complex backgrounds and (b) multiscale ship targets. Blue boxes represent ground truth.

However, SAR ship detection mainly faces the challenges of complex backgrounds and multiscale ship targets [2]. On the one hand, SAR images often contain substantial noise [see Fig. 1(a)], such as land, islands, and clutter, especially in the shore area, which leads to the complex backgrounds in ship detection. This blurs the details of ships, which impairs clear recognition and precise localization of ships and potentially leads to false alarms. On the other hand, SAR adopts a multiresolution imaging mode, resulting in various scales of ship targets [see Fig. 1(b)]. Accurate detection becomes more difficult, particularly for small ships, leading to missed detections and false alarms, diminishing the detection performance.

Traditional SAR ship detection methods mainly rely on CFAR algorithms [3], yet their detection lacks efficiency and accuracy. Researchers have improved ship detection performance by introducing CNN-based models [4], [5], [6], such as Faster R-CNN [7], MSRIHL-CNN [8], Swin-RetinaNet [9], and GFECSI-Net [10], as well as the attention mechanism [11].

Researchers have proposed various methods to address multiscale SAR ship detection in complex backgrounds. Zhang et al. [12] designed BL-Net, improving the accuracy of SAR ship detection in imbalanced classes and complex backgrounds by adjusting sample weights and incorporating reinforcement learning. Shi et al. [2] proposed ASAFE to improve accuracy in detecting multiscale SAR ships in complex backgrounds by combining feature enhancement and adaptive sample allocation. He et al. [13] designed CMFT, transferring features from different images to improve multiscale ship detection in complex backgrounds. Although these methods make progress in improving SAR ship detection performance, they either require more computational resources or lack robustness in complex backgrounds.

To tackle the above challenges, this letter proposes a novel multigranularity-aware network, MGA-Net. We introduce the ideas of multigranularity [14] and design two modules to enhance SAR ship detection. The multigranularity hybrid feature fusion module (MGHF^{2}M) reduces the interference of complex backgrounds. Meanwhile, the multigranularity feature synergy enhancement module (MGFSEM) enhances

the feature. The primary contributions are outlined in this letter.

1) For complex backgrounds, MGHF^{2}M is designed and integrated into the backbone CSPDarknet53, adeptly merges local attention features of different granularity with the original feature to obtain stronger semantic features.
2) For multiscale SAR ship detection, MGFSEM introduces diverse kernel-sized depthwise separable convolutions to capture ship features at various granularities, enhancing the model's receptive field. Furthermore, it adaptively fuses these features with the original ones, which enriches its representation of multiscale ship targets and boosts detection accuracy.

II. Methodology

The framework of the MGA-Net is introduced (see Fig. 2). We integrate the MGHF^{2}M into the backbone CSPDarknet53 [1] to form multigranularity CSPDarknet53 (MG-CSPDarknet53) to capture the information of all ships in complex backgrounds more comprehensively. Next, the multigranularity feature pyramid network (MG-FPN) is an enhanced FPN [7], including MGFSEM, to extract multigranularity context information. Additionally, the BCE loss [1] for classification and objectness, and the GIoU loss [1] for regression.

A. Multigranularity Hybrid Feature Fusion Module

Complex backgrounds in SAR images obscure details and notably affect ship detection accuracy. To tackle this, we propose an MGHF^{2}M (see Fig. 3), inspired by attention mechanisms [11], [15], [16]. MGHF^{2}M is designed to focus on the key channel, spatial, and coordinate information. It employs local channel, local spatial, and local coordinate attention to capture local feature information at different granularities. Simultaneously, it retains the original feature, which includes the details and global features of the image. Next, these features are fused using the adaptive spatial feature fusion module (ASF^{2}M) [17] to enhance the model's representation of ship features under complex backgrounds [18], improving the accuracy of SAR ship detection. The operation of local attention in MGHF^{2}M will be elaborated below.

1) Local Channel Attention: Given an input feature $X \in R^{C\times H\times W}$ (C, H, and W represent the number of channels, height, and width, respectively), we use a 1-D convolution with a k kernel to convolve X to generate a feature map $C1\mathrm{D}_k(X)$. Next, the sigmoid activation function is applied to each channel of $C1\mathrm{D}_k(X)$ to generate channel attention weights. Thus, the feature map X_0 is generated, as follows:

$$\omega_C(X) = \sigma(C1\mathrm{D}_k(X)) \ \left(k = \left|\left(\log_2 C + b\right)/\gamma\right|_{\mathrm{odd}}\right) \quad (1)$$

$$X_0 = X \odot \omega_C(X) \quad (2)$$

where $C1\mathrm{D}$ is 1-D convolution. σ is the sigmoid activation function. b is set to 1, while γ is set to 2. $|\cdot|_{\mathrm{odd}}$ means take the odd number closest to the operation result. $\odot$ represents element-wise multiplication.

2) Local Spatial Attention: Given an input feature $X \in R^{C\times H\times W}$, we initially conduct average and max-pooling operations on X to generate two different spatial features. Then, two features are concatenated along the channel dimension and undergo a 7×7 convolution operation. Finally, the sigmoid activation function generates spatial attention weights applied to X, yielding the feature map X_1, represented as

$$M_s(X) = \sigma\left(f^{7\times 7}(\mathrm{Concat}(\mathrm{MP}(X), \mathrm{AP}(X)))\right) \quad (3)$$

$$X_1 = X \odot M_s(X) \quad (4)$$

where σ is the sigmoid activation function. $f^{7\times 7}$ performs convolution with a 7×7 kernel. Concat is tensor concatenation. AP and MP denote average and max-pooling, respectively. $\odot$ indicates element-wise multiplication.

3) Local Coordinate Attention: Given an input feature $X \in R^{C\times H\times W}$, two 1-D convolutions with kernels $(H, 1)$ and $(1, W)$ convolve X along the horizontal and vertical directions for each channel, yielding two 1-D feature maps $X^h \in \mathbb{R}^{C\times H\times 1}$ and $X^w \in \mathbb{R}^{C\times 1\times W}$. Next, a 1×1 pointwise convolution $P_1 \in \mathbb{R}^{C\times(C/r)\times 1\times 1}$ performs shared operations on X^h and X^w. Finally, the intermediate features obtained after the P_1 operation are processed using two 1×1 pointwise convolutions, $P_h \in \mathbb{R}^{(C/r)\times C\times 1\times 1}$ and $P_w \in \mathbb{R}^{(C/r)\times C\times 1\times 1}$, maintaining the same number of channels as X. r is the channel reduction ratio of 32. The calculation process for feature map X_2 by local coordinate attention is below

$$O^h(X) = \sigma\left(P_h\left(\delta\left(B\left(P_1\left(X^h\right)\right)\right)\right)\right) \quad (5)$$

$$O^w(X) = \sigma(P_w(\delta(B(P_1(X^w))))) \quad (6)$$

$$X_2 = X \odot O^h(X) \odot O^w(X) \quad (7)$$

where $B(\cdot)$ denotes batch normalization. δ and σ represent ReLU and sigmoid activation functions, respectively. O^h and O^w are the attention weights for local coordinates along the horizontal and vertical axes, respectively. $\odot$ represents element-wise multiplication.

To enhance the model's representation of ships with complex backgrounds in SAR images, we employ the ASF^{2}M to fuse X_0, X_1, X_2, and X, effectively leveraging both local and global information. We set X_0, X_1, X_2, and X as levels 0, 1, 2, and 3, respectively. Next, we learn spatial feature weights from each input layer and perform feature fusion calculations, yielding the fused feature map X', as follows:

$$X' = \alpha^l \cdot X^{0\to l} + \beta^l \cdot X^{1\to l} + \phi^l \cdot X^{2\to l} + \eta^l \cdot X^{3\to l} \quad (8)$$

where α^l, β^l, ϕ^l, and η^l are weights from different granularity layers. $X^{0\to l}$, $X^{1\to l}$, $X^{2\to l}$, and $X^{3\to l}$ are outputs from different granularity layers. The sum of α^l, β^l, ϕ^l, and η^l is 1 and compressed between [0, 1] by Softmax, as follows:

$$\alpha^l + \beta^l + \phi^l + \eta^l = 1\left(\alpha^l, \beta^l, \phi^l, \eta^l \in [0, 1]\right) \quad (9)$$

$$\alpha^l = \frac{e^{\lambda_\alpha^l}}{e^{\lambda_\alpha^l} + e^{\lambda_\beta^l} + e^{\lambda_\phi^l} + +e^{\lambda_\eta^l}} \quad (10)$$

where λ_α, λ_β, λ_ϕ, and λ_η are achieved by computing the Softmax layer's parameters during network updates with the backpropagation method.

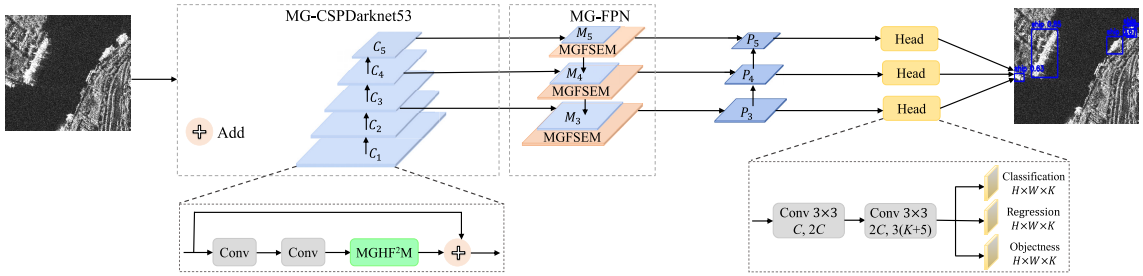


Fig. 2. Framework of MGA-Net. The input image is fed into the improved CSPDarknet53 by the designed MGHF²M for feature extraction. Next, the improved FPN with the proposed MFGSEM extracts multiscale context information. Finally, the ship detection result is output. "Conv" denotes convolution.

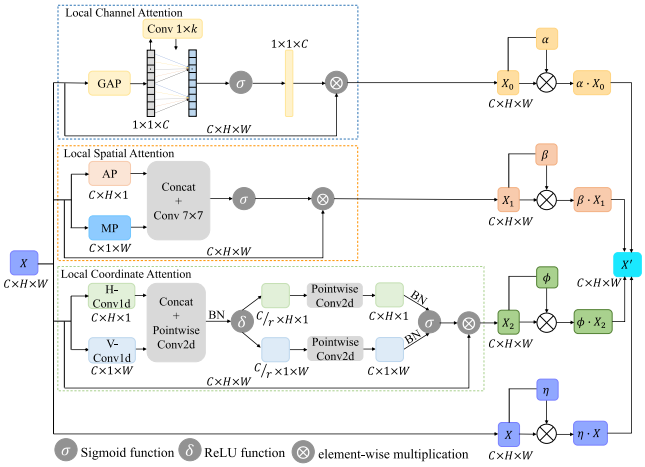


Fig. 3. Structural diagram of the MGHF²M. An ASF²M is employed to merge LCA, LSA, and LCoA features with the original feature, significantly enhancing the model's representation of ships in complex backgrounds. "GAP," "AP," "MP," "H-Conv1d," and "V-Conv1d" refer to global average pooling, average pooling, max-pooling, 1-D convolution with an $(H, 1)$ kernel, and 1-D convolution with a $(1, W)$ kernel, respectively.

B. Multigranularity Feature Synergy Enhancement Module

In SAR images, ships vary in scale, posing detection challenges, especially for small ships. Maximizing multigranularity information is crucial in SAR ship detection. To achieve this without excessive parameter increase, we design MGFSEM (Fig. 4). It uses three parallel branches, each utilizing depthwise separable convolutions [19] of different kernel sizes to extract features at various granularities, enriching the model's representation of ship targets at different scales. We also retain original features to counter potential information loss during extraction. In addition, we adopt an adaptive spatial feature fusion method to avoid ignoring the differences between ship features at different scales.

Specifically, given a feature map Y, we apply separate depthwise convolution to each channel using three different convolution kernels. Next, a pointwise convolution is performed on each pixel of the feature map after depthwise convolution, resulting in three feature maps of different granularities. These operations can be expressed as

$$Y_i = \delta(B(\mathrm{PConv}_1(\mathrm{DConv}_i(Y)))) \quad (i = 0, 1, 2) \tag{11}$$

where DConv_i $(i = 0, 1, 2)$ are depthwise convolutions with 3×3, 5×5, and 7×7 kernels, respectively. PConv_1 is a 1×1 pointwise convolution. $B(\cdot)$ and δ are batch normalization and ReLU activation functions, respectively.

To balance the impact of different granularity features during training, we adopt the ASF²M to fuse representations of each granularity feature. We set Y_0, Y_1, Y_2, and Y as levels 0, 1, 2, and 3, respectively. Similarly, the calculation process of the fused feature map Y' is as follows:

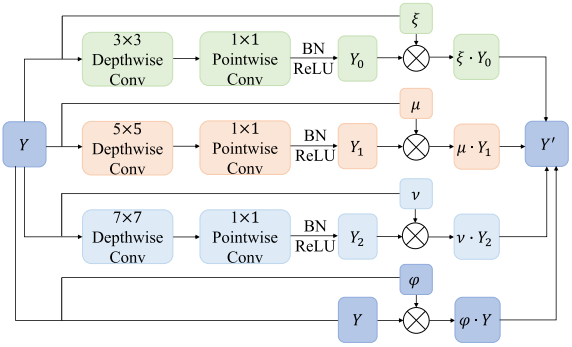


Fig. 4. Structure diagram of the MGFSEM, comprising three parallel depthwise separable convolutional layers, along with an original feature map.

$$Y' = \xi^l \cdot Y^{0\to l} + \mu^l \cdot Y^{1\to l} + \nu^l \cdot Y^{2\to l} + \varphi^l \cdot Y^{3\to l} \tag{12}$$

where ξ^l, μ^l, ν^l, and φ^l are weights from different granularity layers. $Y^{0\to l}$, $Y^{1\to l}$, $Y^{2\to l}$, and $Y^{3\to l}$ are outputs from different granularity layers. The sum of ξ^l, μ^l, ν^l, and φ^l is 1 and compressed between [0, 1] by Softmax, as follows:

$$\xi^l + \mu^l + \nu^l + \varphi^l = 1\left(\xi^l, \mu^l, \nu^l, \varphi^l \in [0, 1]\right) \tag{13}$$

$$\alpha^l = \frac{e^{\lambda_\xi^l}}{e^{\lambda_\xi^l} + e^{\lambda_\mu^l} + e^{\lambda_\nu^l} + + e^{\lambda_\varphi^l}} \tag{14}$$

where the parameters λ_ξ, λ_μ, λ_ν, and λ_φ are achieved by computing the Softmax layer's parameters during network updates with the backpropagation method.

III. Experimental Results

A. Implementation Details

1) Datasets: HRSID [20] has 5604 SAR images resized to 800 × 800, with 65% for training and 35% for testing. SAR-Ship-Dataset [21] comprises 43 819 images sized at 256 × 256, with a ratio of 7:2:1 for training, validation, and test sets. SSDD [22] has 1160 images with a size of 512 × 512, with those ending in 1 or 9 for testing and the rest for training.

2) Experimental Settings and Evaluation Metrics: All experiments use PyTorch on GPU Tesla V100. HRSID is utilized for ablation study and parameter analysis. We use the stochastic gradient descent optimizer with an initial learning rate of 0.01, batch size of 4, momentum value of 0.9, and weight decay of 0.0001. The training epochs are 60 for HRSID, 45 for SAR-Ship-Dataset, and 80 for SSDD.

TABLE I

ABLATION STUDY OF INDIVIDUAL MODULES ON HRSID. ✓ MEANS TO USE THIS MODULE. MGHF^{2}M CONSISTS OF LCA, LSA, LCOA, AND ASF^{2}M. MGFSEM MEANS MULTIGRANULARITY FEATURE SYNERGY ENHANCEMENT MODULE

LCA	LSA	LCoA	ASF^{2}M	MGFSEM	P	R	F1	mAP	FPS	Param(M)	FLOPs(G)
✗	✗	✗	✗	✗	88.83	89.06	88.94	89.96	13.67	46.50	36.97
✓	✗	✗	✗	✗	88.95	89.15	89.05	90.31	13.65	46.50	37.01
✓	✓	✗	✓	✗	89.84	89.43	89.63	90.88	13.34	46.63	37.11
✓	✗	✓	✓	✗	91.01	89.82	90.41	91.52	12.25	47.61	37.89
✗	✓	✓	✓	✗	90.53	89.77	90.15	91.44	12.27	47.61	37.93
✓	✓	✓	✓	✗	91.23	90.37	90.80	92.38	11.82	47.74	38.02
✗	✗	✗	✗	✓	91.93	89.24	90.57	91.31	10.44	49.33	39.26
✓	✓	✓	✓	✓	93.87	91.14	92.48	93.45	10.01	50.57	40.25

TABLE II

COMPARISON OF EVALUATION METRICS FOR VARIOUS METHODS ON HRSID, WITH THE BEST RESULTS IN BOLD

Method	Entire Scenes				Inshore Scenes				Offshore Scenes				FPS	Param(M)
	P	R	F1	mAP	P	R	F1	mAP	P	R	F1	mAP		
Faster R-CNN [7]	81.45	81.97	81.71	80.66	65.55	65.51	65.53	60.10	95.93	97.17	96.55	97.09	14.05	41.30
Guided Anchoring [4]	90.41	84.62	87.42	83.72	81.22	70.64	75.56	66.99	97.82	97.53	97.67	97.46	10.49	41.89
BL-Net [12]	91.58	89.74	90.65	88.67	84.35	80.99	82.64	77.71	98.01	97.82	97.91	97.77	5.21	47.81
Swin-RetinaNet [9]	69.60	87.10	77.37	85.94	50.20	75.00	60.14	65.43	91.20	96.70	93.87	97.32	12.80	36.82
ASAFE [2]	83.46	86.75	85.07	85.18	71.67	71.95	71.81	68.91	98.67	96.59	97.62	96.92	13.68	42.43
Improved FCOS [5]	81.70	90.10	85.69	88.60	67.60	80.90	73.65	75.17	94.70	97.30	95.98	97.09	10.23	40.60
GFECSI-Net [10]	85.30	89.30	87.25	90.84	72.60	80.40	76.30	79.65	96.50	96.30	96.40	97.34	**16.10**	**36.82**
CMFT [13]	81.30	91.10	85.92	89.60	80.16	86.77	83.33	78.63	95.88	98.01	96.93	96.94	12.58	46.12
MGA-Net (Ours)	93.87	91.14	**92.48**	**93.45**	85.13	87.02	**86.06**	**83.15**	97.91	98.02	**97.96**	**98.53**	10.01	50.57

TABLE III

COMPARISON OF EVALUATION METRICS FOR VARIOUS METHODS ON SAR-SHIP-DATASET, WITH THE BEST RESULTS IN BOLD

Method	Entire Scenes				Inshore Scenes				Offshore Scenes				FPS	Param(M)
	P	R	F1	mAP	P	R	F1	mAP	P	R	F1	mAP		
Faster R-CNN [7]	86.85	93.24	89.93	91.73	69.55	86.80	77.22	79.47	93.85	95.36	94.60	94.65	23.74	41.30
Guided Anchoring [4]	92.59	93.80	93.19	92.73	82.54	86.12	84.29	81.74	96.03	96.33	96.18	95.79	17.41	41.89
BL-Net [12]	91.58	89.74	90.65	94.25	85.91	91.75	88.73	88.65	96.93	97.24	97.08	96.67	8.26	47.81
Swin-RetinaNet [9]	86.05	94.10	89.90	94.11	70.71	90.25	79.29	83.93	95.02	97.33	96.16	96.91	20.53	36.82
ASAFE [2]	86.12	93.71	89.75	94.01	80.93	87.05	83.88	84.77	96.93	97.04	96.98	96.78	23.11	42.43
Improved FCOS [5]	89.80	95.20	92.42	94.09	78.30	90.20	83.83	85.90	95.80	97.50	96.64	97.15	12.20	40.60
GFECSI-Net [10]	90.24	93.01	91.60	94.33	83.03	93.45	87.93	88.65	94.59	97.07	95.81	97.00	**26.12**	**16.08**
CMFT [13]	90.07	94.41	92.19	94.54	86.03	92.45	89.12	89.33	95.02	97.31	96.15	97.02	21.26	46.12
MGA-Net (Ours)	92.61	96.73	**94.63**	**96.47**	87.65	93.84	**90.64**	**91.53**	95.82	98.34	**97.06**	**98.21**	17.57	50.57

Evaluation metrics include precision (P), recall (R), $F1$-score, mean average precision (mAP), frames per second (FPS), model parameters (Param), and floating point of operations (FLOPs).

B. Ablation Study

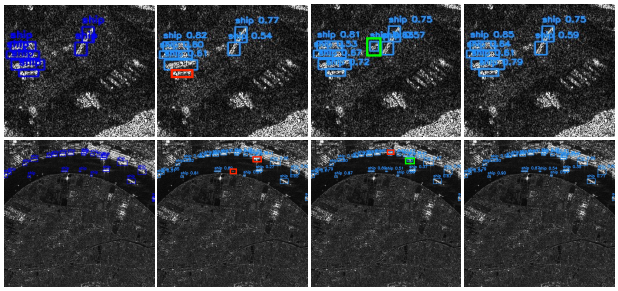
We assess MGHF^{2}M (including local channel attention (LCA), local spatial attention (LSA), local coordinate attention (LCoA), and ASF^{2}M) and MGFSEM by ablation study, presented in Table I. Compared to the baseline, only the combined effect of these four submodules increases mAP by 2.42%, adding only 1.05G FLOPs. This combination pays more attention to local information of different granularities related to the ship target, thereby reducing interference from complex backgrounds and improving detection accuracy. After adding MGFSEM, mAP increases by 1.35% from the baseline to reach 91.31%, enhancing context information in various scales. Integrating these modules yields a 93.45% mAP, achieving the best result.

C. Comparison With the State-of-the-Art

The proposed MGA-Net is validated on HRSID, SAR-Ship-Dataset, and SSDD, comparing with other state-of-the-art methods in Tables II–IV. Table II lists the outcomes of various methods in various scenes on HRSID. In the entire and inshore scenes, MGA-Net attains 93.45% mAP and 83.15% mAP, surpassing the suboptimal GFECSI-Net by 2.61% and 3.5%. In the offshore scene, MGA-Net achieves a 98.53% mAP, surpassing the suboptimal BL-Net by 0.76%. Furthermore, it achieves the highest $F1$-score of 92.48% (entire), 86.06% (inshore), and 97.96% (offshore), respectively. Table III shows the outcomes of diverse methods on SAR-Ship-Dataset. In the entire and inshore scenes, MGA-Net achieves 96.47% mAP and 91.53% mAP, surpassing the suboptimal CMFT by 1.93% and 2.2%. In the offshore scene, MGA-Net achieves 98.21% mAP, which is 1.06% higher than the suboptimal Improved FCOS. Moreover, it achieves the highest $F1$-score in various scenes. Similarly, Table IV displays the results of diverse methods on SSDD. In diverse scenes, MGA-Net achieves the highest mAP and $F1$-score. This is due to MGHF^{2}M and MGFSEM in MGA-Net. MGHF^{2}M extracts local attention features with different granularities, alleviating the impact of complex backgrounds on ship detection. Meanwhile, MGFSEM enriches multigranularity contextual information, mitigating detection challenges for ships of different scales, especially for small ones.

TABLE IV
COMPARISON OF EVALUATION METRICS FOR VARIOUS METHODS ON SSDD, WITH THE BEST RESULTS IN BOLD

Method	Entire Scenes				Inshore Scenes				Offshore Scenes				FPS	Param(M)
	P	R	F1	mAP	P	R	F1	mAP	P	R	F1	mAP		
Faster R-CNN [7]	87.08	90.44	88.73	89.74	68.98	75.00	71.86	71.39	96.03	97.58	96.80	97.37	11.87	41.30
Guided Anchoring [4]	94.62	90.44	92.48	90.01	86.90	73.26	79.50	71.59	97.60	98.39	97.99	98.22	9.64	41.89
BL-Net [12]	91.27	96.14	93.64	95.25	80.00	88.37	83.98	84.79	96.87	99.73	98.28	99.62	5.02	47.81
Swin-RetinaNet [9]	89.52	95.07	92.21	95.23	74.75	81.19	77.84	80.25	97.85	98.97	98.41	98.84	10.14	36.82
ASAFE [2]	88.54	95.94	92.09	95.19	75.66	81.01	78.24	80.82	97.87	98.89	98.38	98.78	11.55	42.43
Improved FCOS [5]	89.05	96.73	92.73	95.01	73.27	87.21	79.63	82.18	95.27	98.61	96.91	98.43	8.15	40.60
GFECSI-Net [10]	93.10	96.91	94.97	97.18	89.21	80.42	84.59	85.75	97.52	98.79	98.15	99.11	**14.02**	**16.08**
CMFT [13]	92.40	98.10	95.16	97.30	90.11	83.37	86.61	87.23	97.65	99.53	98.58	99.63	10.63	46.12
MGA-Net (Ours)	94.79	99.01	**96.85**	**98.63**	91.07	87.44	**89.22**	**89.91**	97.87	99.56	**98.71**	**99.67**	10.71	50.57



(a) (b) (c) (d)

Fig. 5. Comparison of visualization results from various methods on HRSID. Each column displays the detection results of ground truth, CMFT, GFECSI-Net, and MGA-Net. Navy blue, light blue, green, and red boxes are ground truth, detection results, false alarms, and missed detections, respectively. (a) Ground truth. (b) CMFT. (c) GFECSI-Net. (d) MGA-Net.

To better showcase the robustness of MGA-Net, we exhibit the visualization results compared with several methods in Fig. 5. The first and second rows depict the detection results of various methods under complex backgrounds and small ship targets, respectively. Obviously, MGA-Net can detect ships accurately, minimizing false alarms and missed detections.

IV. CONCLUSION

This letter proposes a novel MGA-Net to tackle challenges posed by complex backgrounds and multiscale ship targets in SAR images. MGHF^{2}M is designed to acquire contextual information ranging from coarse to fine, strengthening the model's resistance to interference from complex backgrounds. Meanwhile, MGFSEM is proposed to enhance the model's capability to extract features at different scales and effectively handle ship targets of different scales in SAR images. Experimental results indicate that our MGA-Net achieves superior performance in two SAR ship datasets. Since this letter uses rectangular bounding boxes, it restricts the precise localization and classification of SAR ship targets. We plan to develop an efficient MGA-Net to achieve SAR ship detection in arbitrary directions.

REFERENCES

[1] 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, 2023, Art. no. 5215514.

[2] H. Shi, Z. Fang, Y. Wang, and L. Chen, "An adaptive sample assignment strategy based on feature enhancement for ship detection in SAR images," *Remote Sens.*, vol. 14, no. 9, p. 2238, May 2022.

[3] H. Lin, H. Chen, K. Jin, L. Zeng, and J. Yang, "Ship detection with superpixel-level Fisher vector in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 247–251, Feb. 2020.

[4] J. Wang, K. Chen, S. Yang, C. C. Loy, and D. Lin, "Region proposal by guided anchoring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2960–2969.

[5] S. Yang, W. An, S. Li, G. Wei, and B. Zou, "An improved FCOS method for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8910–8927, 2022.

[6] J. Ai, Y. Mao, Q. Luo, L. Jia, and M. Xing, "SAR target classification using the multikernel-size feature fusion-based convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5214313.

[7] 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.

[8] J. Ai, R. Tian, Q. Luo, J. Jin, and B. Tang, "Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10070–10087, Dec. 2019.

[9] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[10] S. Yang, W. An, S. Li, S. Zhang, and B. Zou, "An inshore SAR ship detection method based on ghost feature extraction and cross-scale interaction," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[11] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

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

[13] J. He et al., "A cross-modality feature transfer method for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5213615.

[14] 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, 2023, Art. no. 1001219.

[15] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.

[16] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13713–13722.

[17] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," 2019, *arXiv:1911.09516*.

[18] Y. Liu, B. N. Zhao, S. Zhao, and L. Zhang, "Progressive motion coherence for remote sensing image matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5631113.

[19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[20] 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.

[21] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019.

[22] 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. (BIGSARMM)*, Nov. 2017, pp. 1–6.