3WM-AugNet: A Feature Augmentation Network for Remote Sensing Ship Detection Based on Three-Way Decisions and Multigranularity

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

***Abstract*— With the continuous advancement of remote sensing (RS) technology, RS ship detection plays a crucial role in ensuring maritime safety and the oceanic economy, but it also faces various challenges. Most existing RS ship detection methods typically apply deblurring processing to all input images before using a feature pyramid network (FPN) to detect ships of different sizes. However, this indiscriminate operation may cause image quality degradation due to excessive deblurring. Moreover, FPN has limitations in fully utilizing multigranularity features, which is particularly severe in RS ship detection tasks. These issues severely affect the accuracy of RS ship detection. To address these problems, this article proposes an effective feature augmentation network, 3WM-AugNet, based on the three-way decisions (3WDs) and multigranularity feature learning for RS ship detection. It consists of two modules: a blurred classification and deblurring module (BCDM) and a multigranularity feature augmentation module (MFAM). BCDM aims to combine 3WD and support vector machine (SVM) to design an image clarity classification algorithm and use the multi-temporal recurrent neural network (MT-RNN) algorithm to process the blurry images classified, effectively avoiding excessive deblurring of clear images. MFAM is used to enhance the richness and robustness of feature representations for ships of different sizes by introducing the bottom-up feature fusion layer and designing an adaptive coordinate attention module. Experimental results on three commonly used datasets, FGSD2021, HRSC2016, and UCAS-AOD, show that our proposed 3WM-AugNet achieves state-of-the-art performance in RS ship detection.**

***Index Terms*— Adaptive coordinate attention (ACA), multigranularity, remote sensing (RS), ship detection, three-way decisions (3WDs).**

I. Introduction

SHIPS are important tools and carriers for maritime transportation and the utilization of marine resources, playing a crucial role in both the marine economy and national security. Remote sensing (RS) technology can provide high-resolution and wide-coverage marine image data, which is an indispensable data foundation for real-time monitoring and analysis of ship detection. Therefore, ship detection in RS images has been increasingly receiving attention, as confirmed by numerous studies [1], [2], [3].

However, efficient ship detection in RS images is challenging. On the one hand, several factors such as shaking of the satellite imaging equipment, atmospheric disturbances, and the movement of the ships themselves can affect RS ship image acquisition, resulting in blurry representations of ships in the image, as shown in Fig. 1(a). This makes it difficult to accurately extract object features and affects the accuracy of the detection algorithm. On the other hand, factors such as the size of the ship itself, voyage distance, and shooting angle make the ship targets in the RS image show different sizes, as shown in Fig. 1(b). This increases the difficulty of ship detection. Therefore, an adaptive and accurate ship detection method is required to address issues such as image blurring and varying sizes of ship targets to improve the reliability and accuracy of ship detection.

In recent years, advancements in deep learning have greatly enhanced the performance of RS ship detection networks. Many studies have continuously introduced various excellent detection networks that enhance the richness and robustness of features for ships, thereby further improving the accuracy of RS ship detection. During RS ship detection, deep learning-based algorithms are often used to deblur images, addressing the problem of blurry RS images. However, this strategy can lead to the loss of details in clear images due to excessive deblurring, which can affect the accuracy and reliability of RS ship detection. In addition, when dealing with the problem of different sizes of ship targets in RS ship detection, some studies have adopted feature pyramid network (FPN)-based frameworks [4], [5] to extract ship object features of different granularities. FPN constructs a feature pyramid structure by propagating high-level semantically strong features to low-level features, thereby achieving accurate localization and classification of ship targets of different sizes in the ship detection network. However, there are some design flaws in FPN, as shown in Fig. 2. On the one hand, due to the adoption of layer-by-layer fusion, low-level features can only influence high-level features through top-down propagation, thus limiting the effective transmission and impact of low-level features on high-level features. On the

Manuscript received 15 May 2023; revised 10 July 2023, 27 July 2023; 26 August 2023; accepted 4 September 2023. Date of publication 11 September 2023; date of current version 28 September 2023. This work was supported in part by the National Key Research and Development Program of "Cyberspace Security Governance" Special Program under Grant 2022YFB3104700; in part by the National Nature Science Foundation of China under Grant 62376198, Grant 61906137, Grant 62076040, Grant 62076182, Grant 62163016, and Grant 62006172; in part by the Jiangxi "Double Thousand Plan," the Jiangxi Provincial Natural Science Fund under Grant 20212ACB202001; in part by the China National Scientific Sea-Floor Observatory, the Natural Science Foundation of Shanghai under Grant 22ZR1466700; and in part by the Interdisciplinary Project in Ocean Research of Tongji University. *(Corresponding authors: Duoqian Miao; Zhifei Zhang.)*

Li Ying and Zhifei Zhang are with the Department of Computer Science and Technology and 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).

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

Digital Object Identifier 10.1109/TGRS.2023.3313603

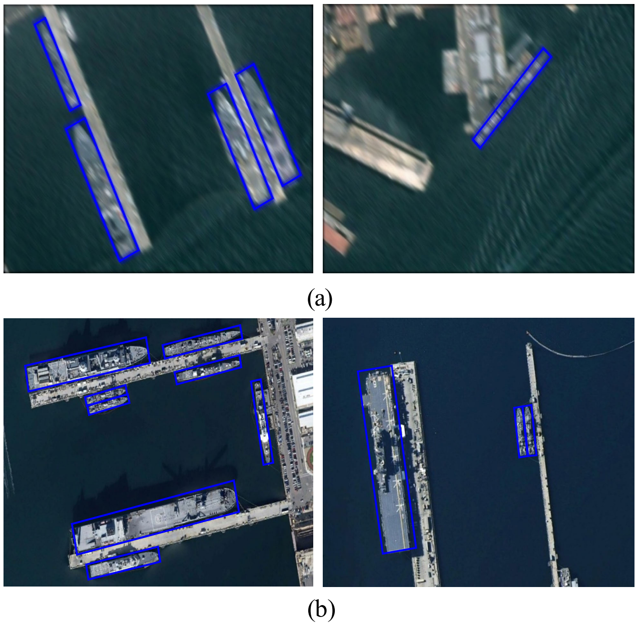


Fig. 1. Challenges of RS images in FGSD2021. (a) Motion blur and (b) ship targets of different sizes. Blue boxes represent ground truth.

other hand, during feature fusion, features propagate along a top-down path. The features at the top pyramid level suffer from information loss due to the reduced number of channels. These deficiencies limit FPN to a certain extent to make full use of multigranularity features, resulting in a decrease in RS ship detection performance.

To overcome the limitations of existing methods in RS ship detection, this article proposes a network called 3WM-AugNet. This network aims to address the problems of RS image blurring and inconsistent ship target sizes, thereby improving the accuracy of RS ship detection. Specifically, we adopt a preprocessing and detection strategy for RS ship images. First, based on the three-way decision (3WD) theory, we design the blurred classification and deblurring module (BCDM) to preprocess the RS images to avoid excessive blurring of clear images. The 3WD theory is a method for dealing with uncertain decision-making [6]. The preprocessing of RS images based on the 3WD theory can effectively classify clear and blurry images. This strategy allows for targeted deblurring processing solely for the blurry images while avoiding overprocessing the clear ones. Then, combining with the idea of multigranularity feature learning [7], [8], we design the multigranularity feature augmentation module (MFAM) to enhance the richness and robustness of feature representations for different-sized ships by introducing the bottom-up feature fusion layer (BF^{2}L) and designing an adaptive coordinate attention (ACA) module, thus solving the defects in FPN.

Our work mainly has the following contributions.

1) We propose a BCDM to differentially process the input RS images for data augmentation. BCDM can avoid excessive blurring of clear images while effectively deblurring blurry images, which greatly benefits subsequent feature extraction.

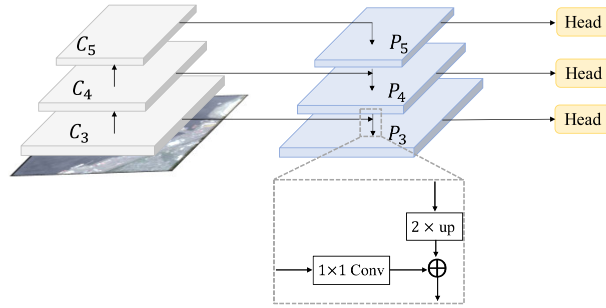


Fig. 2. FPN structure of the baseline. "1 × 1 Conv" refers to 1 × 1 convolution, "2× up" refers to bilinear difference for upsampling, and ⊕ means addition.

2) We design an MFAM to tackle the shortcomings of FPN in extracting RS ship features of different sizes. MFAM can facilitate the transfer of low-level features to the high level, thus minimizing the loss of high-level features while enriching their representation.

The rest of this article is structured as follows. Section II introduces the related works. Section III describes the details of the proposed 3WM-AugNet. Section IV presents the qualitative and quantitative comparisons with state-of-the-art methods and some ablation studies. Section V draws some conclusions and potential future work.

II. Related Works

In this section, we first briefly introduce the existing RS ship detection methods. Then, the related technologies used in the proposed method are introduced, including RS image deblurring and multigranularity feature learning.

A. RS Ship Detection

RS ship detection is an important branch of RS image processing, which has received widespread attention from many researchers. RS ship detection methods can be divided into traditional and deep learning-based methods. Traditional methods [9], [10] usually require extensive manual feature extraction, and the extracted features are then fed into a classifier for learning. However, the robustness and generalization ability of manually extracted features is limited, resulting in poor performance in complex scenes, such as detecting ships with varying sizes and blurry backgrounds. In recent years, the continuous development of deep learning technology [11], [12], [13] has made deep learning-based ship detection methods a research hotspot.

Many algorithms based on convolutional neural networks (CNNs) have been proposed for RS ship detection to enhance accuracy and robustness. In particular, Liu et al. [4] drew inspiration from the YOLOv3 algorithm in RS ship detection, dividing the detection task into coarse and fine detection stages and utilizing distinct network structures for each stage. Wang et al. [5] proposed a RetinaNet automatic ship detection method using multiresolution Gaofen-3 RS images and achieved object detection through a pyramidal classifier. Yang et al. [14] proposed a robust one-stage detector that can detect RS ships of different sizes in complex backgrounds.

In addition, to further improve the accuracy of ship detection in any direction within RS images, many researchers have proposed arbitrary-direction object detection methods, including both one- and two-stage methods.

Two-stage arbitrary-orientation object detection methods involve the extraction of candidate boxes followed by classification and regression on each candidate box to obtain accurate location and category information for the objects. Popular methods in this category include R^2CNN [15], RoI Transformer [16], SCRDet [17], Gliding Vert [18], and SCRDet++ [19]. While these methods can effectively detect objects in any orientation and provide precise location and category information, they come with high computational costs and are unsuitable for real-time applications.

In contrast, one-stage arbitrary-orientation object detection methods are simpler and faster, typically requiring only a single neural network to complete the task of object detection and orientation estimation. Some common methods include CSL [20], R^3Det-DCL [21], R^3Det [22], RSDet [23], BBAVectors [24], DAL [25], and Oriented R-CNN [26]. More recently, various methods for detecting objects with arbitrary orientations have emerged. For instance, Zhang et al. [27] designed the CHPDet detector that utilizes center point extraction to detect ships in any direction in RS images. This is achieved by combining the center point with the prediction of the head direction. Han et al. [28] proposed S^2A-Net that solves the problem of rotation variance in object detection by using aligned depth features, achieving more accurate object detection. Zhang et al. [29] proposed FFN that generates fountain features by reconstructing unsatisfactory detection unit features, significantly improving object detection accuracy in any direction in RS images. Li et al. [30] designed Oriented RepPoint that uses a deformable convolutional network to generate rotated boxes and represents the target through deformable points, realizing efficient detection of targets in any direction in RS images. Zhang et al. [31] proposed TCD that significantly improves the performance of detecting oriented objects in RS images through task-collaborative learning and information sharing. Li et al. [32] designed LSKNet that uses a spatial selection mechanism to dynamically adjust the receptive field of the feature extraction backbone for better RS object detection. Liang et al. [33] proposed DEA-Net that improves the robustness of object detection in RS images by adaptively optimizing the position and size of prior boxes through mutual interaction and information transfer between models. Wang et al. [34] proposed GF-CSL that optimizes the detection results of arbitrary-direction targets in RS images by introducing polarization angle prediction and Gaussian distribution strategy. These methods have demonstrated high accuracy and robustness in practical applications and can be used in RS image ship detection. However, these methods mainly improve the precision of ship detection by enhancing the representation of rotation boxes and do not fully consider the impact of image blur and varying ship sizes on detection results.

Therefore, this article proposes a feature augmentation network named 3WM-AugNet for RS ship detection, which is based on S^2A-Net and integrates the principles of 3WD theory and multigranularity feature learning. The proposed method aims to tackle the challenges posed by image blurring and various sizes of ship targets.

B. RS Image Deblurring

High-quality RS images [35], [36], [37] are essential for effectively operating many intelligent visual algorithms. However, such images often suffer from motion blur caused by camera shake or ship motion, which can severely impact image quality and utility. Therefore, motion deblurring of RS images has become a widely researched field.

To improve the clarity and visual quality of RS images, researchers have adopted techniques for motion deblurring, supporting the analysis and interpretation of RS images and providing an accurate and reliable data foundation. Traditional RS image deblurring algorithms usually use regularization techniques based on statistical priors, such as gradient sparsity prior [38], hyper-Laplacian prior [39], low-rank prior [40], and L_0-norm gradient prior [41]. However, these methods heavily rely on image priors, and their performance will significantly degrade if the image priors do not hold.

In recent years, a series of extension methods based on deep learning for RS image motion deblurring have shown remarkable progress [42], [43]. For example, Tao et al. [44] proposed SRN that utilizes a scale-recursion mechanism in CNN to address motion blur in images. This method significantly improves the clarity of images through multilevel feature representation and information propagation. Kupyn et al. [45] designed DeblurGAN-v2 that utilizes generative adversarial networks (GANs) for motion deblurring of images, resulting in significant improvements in both efficiency and effectiveness. Park et al. [46] designed a multi-temporal recurrent neural network (MT-RNN) for incremental time training in progressive nonuniform single-image deblurring, achieving more accurate motion deblurring effects. Cho et al. [47] proposed MIMO-UNet that incorporates multiscale attention mechanisms and deconvolution operations to better preserve the details and structural information of the image, thereby achieving more accurate motion deblurring effects in single-image deblurring. Ji et al. [48] proposed XYDeblur that utilizes image decomposition and deep CNN to partition single image deblurring into subtasks, achieving image motion deblurring. These methods fully exploit the advantages of deep learning, providing more powerful techniques in the field of RS image motion deblurring and improving image quality.

Although the above algorithms can solve the motion blur problem of RS images to a certain extent, their objective is to deblur all RS images, including those clear ones. This can disrupt subsequent image target detection. To address this issue, this article presents a classification-based deblurring strategy that specifically targets blurry images, preventing the overdeblurring of clear ones and ultimately resulting in high-quality RS images.

C. Multigranularity Feature Learning

Multigranularity feature learning is a key issue for object detection since there are many differences in the size, shape,

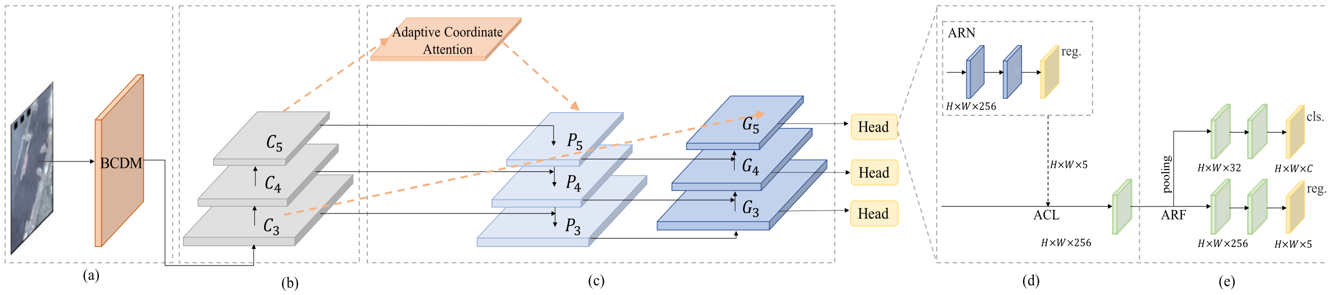


Fig. 3. Architecture of 3WM-AugNet. Compared with the baseline S^2A-Net, we mainly design a BCDM for RS image preprocessing to mitigate image overblurring. MFAM is designed to replace the original FPN for feature fusion and enhance the richness of multigranularity feature representation of ship targets. (a) BCDM. (b) Backbone. (c) MFAM. (d) FAM. (e) ODM.

and pose of objects in images. How to effectively learn multigranularity features has been extensively studied by many researchers. Previous methods employed single-granularity feature networks [49], [50] for prediction, but they struggled to effectively handle the variations of targets at different sizes. To improve the performance of multigranularity feature learning, a series of network frameworks dealing with multigranularity features have emerged. FPN [51] is a widely used multigranularity feature learning framework, which uses a top-down feature extraction method to obtain multigranularity features of objects through feature pyramids between different levels, enabling more accurate detection of objects with different sizes.

However, several studies have revealed the limitations of FPN, including the loss of high-level features and the difficulty of low-level features effectively influencing high-level features. To address the loss of high-level features, Ghiasi et al. [52] proposed NAS-FPN that employs neural architecture search to automatically learn the architecture of FPN. This approach better preserves and propagates the semantic information of high-level features, enabling more efficient and accurate multigranularity feature extraction. Zhao et al. [53] designed GraphFPN that incorporates graph convolutional networks to better propagate and fuse low-level features with high-level features, thereby improving the robustness and accuracy of object detection. In addition, to enhance the interactions between low- and high-level features, Liu et al. [54] proposed PANet that achieves cross-layer feature fusion through information aggregation between different branches in FPN and further utilizes low-level spatial information, thereby significantly improving the accuracy of object detection. Tan et al. [55] designed BiFPN that introduces a bidirectional FPN on top of PANet, integrating low- and high-level features, thereby significantly improving the accuracy and efficiency of object detection. Huang et al. [56] proposed FaPN to align feature pyramids at different levels by introducing a feature alignment mechanism to enhance information transfer and detail reservation. Jin et al. [57] introduced a cascaded attention mechanism to augment the fusion of low- and high-level features in FPN while simultaneously focusing on global context and fine-grained features. This significantly improves the accuracy and robustness of the model on objects of different sizes.

Nevertheless, these algorithms predominantly address one of the two shortcomings of FPN and have certain limitations. Therefore, this article proposes an MFAM that incorporates multigranularity feature learning to address the two main flaws in FPN, thereby improving the performance of object detection.

III. Method

This article proposes a novel RS ship detection network, 3WM-AugNet, built on the S^2A-Net [28] baseline. First, a BCDM is innovatively proposed by incorporating the 3WD theory to solve the problem of excessive blurring in RS images. Then, the MFAM is designed to replace the FPN, improve the richness of feature representations of ships with various sizes, and reconstruct the pyramid network with multigranularity features. The overall framework of 3WM-AugNet is shown in Fig. 3. The proposed 3WM-AugNet in this article will be explained in detail from the following three aspects.

A. S^2A-Net as Baseline

This article chooses the one-stage detector S^2A-Net [28] as the baseline model. It is an RS ship detection model based on rotating RetinaNet, which consists of the backbone network, FPN [51], the feature alignment module (FAM), and the orientation detection module (ODM). FAM uses an anchor refinement network (ARN) to generate rotated anchors and an aligned convolutional layer (ACL) to extract aligned features. ODM uses an active rotation filter (ARF) [58] to encode orientation information and obtain direction-sensitive features, which are fused to extract direction-invariant features. The model changes the RetinaNet regression output from horizontal bounding boxes (BBoxes) to rotating BBoxes, making it compatible with arbitrary-oriented RS ship detection.

B. Blurred Classification and Deblurring Module

In ship detection for RS images, image blurring can impact image clarity and reduce model detection rates. To enhance model robustness, RS images must be deblurred. However, existing RS image deblurring algorithms deblur all images, leading to excessive deblurring of clear images and reduced clarity, which is counterproductive to subsequent detection. Therefore, we propose a BCDM for deblurring RS ship images and obtaining high-quality RS image samples for subsequent training, as shown in Fig. 4. Specifically, we first design a 3WD-based RS image blur level classification algorithm,

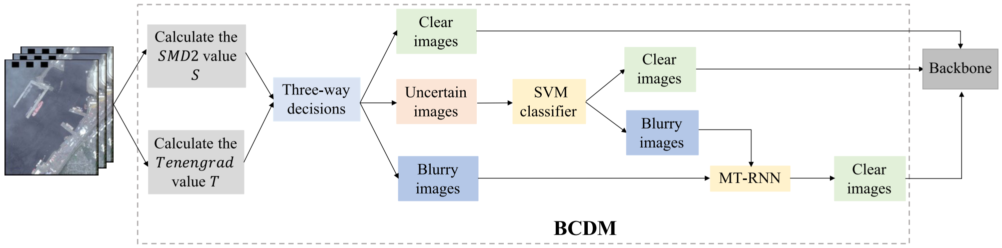


Fig. 4. Structure of the BCDM. First, the input images are categorized as clear, uncertain, and blurry by using the 3WD theory. Then, the SVM classifier is utilized to further classify the uncertain images. Finally, the MT-RNN algorithm is employed solely to deblur the blurry images, while the clear images remain unchanged.

which can effectively classify clear and blurry images. Then, we use the MT-RNN algorithm [46] to deblur the blurry images, achieving efficient deblurring while avoiding the overdeblurring of clear images.

The pseudocode of our 3WD-based RS image blur level classification algorithm is shown in Algorithm 1. Specifically, first, under the guidance of no-reference image blur assessment metrics, our classification algorithm mainly selects two ambiguity evaluation algorithms for all images $I = (i_1, i_2, \ldots, i_m)$. The first algorithm is based on the value of the sum of the modified differential squared (SMD2) function in pixel technology. The SMD2 function can be expressed as

$$\mathrm{SMD2}(g) = \sum_b \sum_a |g(a,b) - g(a+1,b)| * |g(a,b) - g(a,b+1)| \quad (1)$$

where $g(a,b)$ represents the gray-scale value of the image g corresponding to the pixel (a,b). The smaller the SMD2(g) is, the blurrier the image becomes, and vice versa. The second is based on the Tenengrad function value in the image gradient technology, which can be written as

$$\mathrm{Tenengrad}(g) = \sqrt{g_x^2(a,b) + g_y^2(a,b)} \quad (2)$$

$$g_x(a,b) = g(a,b) * K_x \quad (3)$$

$$g_y(a,b) = g(a,b) * K_y \quad (4)$$

$$K_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}, \quad K_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (5)$$

where g_x and g_y are the convolution of the Sobel horizontal convolution kernel K_x and the vertical convolution kernel K_y at the pixel point (a,b), respectively. The smaller the Tenengrad(g) is, the blurrier the image becomes, and vice versa. By calculating the value of the SMD2 function and the Tenengrad function, the blurring degree of each image can be obtained, namely, $s_j = \mathrm{SMD2}(i_j)$ and $t_j = \mathrm{Tenengrad}(i_j)$. Take the SMD2 values and Tenengrad values as data points $q_j = [s_j, t_j] (j = 1, 2, \ldots, m)$, and combine them into a dataset $Q = (q_1, q_2, \ldots, q_m)$. Then, the dataset $Q = (q_1, q_2, \ldots, q_m)$ is clustered using a Gaussian mixture clustering algorithm and 3WD theory. The steps are given as follows. Step 1, based on the 3WD theory, randomly initializes the Gaussian distribution parameters of each category, including mean value μ_k, covariance matrix Σ_k, and mixing coefficient χ_k. Step 2 calculates the posterior probability that each data point q_j belongs to each Gaussian distribution. The formula for the posterior probability η_{jk} produced by each mixed component of q_j is given as follows:

$$\eta_{jk} = p(z_j = k | q_j) = \frac{\chi_k \cdot p(q_j | \mu_k, \Sigma_k)}{\sum_{l=1}^{3} \chi_l \cdot p(q_j | \mu_l, \Sigma_l)}, \quad (1 \le k \le 3) \quad (6)$$

where $p(q_j|\mu_k, \Sigma_k)$ denotes the probability density function of each blend element in q_j. To better fit the clustering data, in Step 3, the mean μ_k', covariance matrix Σ_k', and mixing coefficient χ_k' of each Gaussian distribution are updated as follows:

$$\mu_k' = \frac{\sum_{j=1}^{m} \eta_{jk} q_j}{\sum_{j=1}^{m} \eta_{jk}} \quad (7)$$

$$\Sigma_k' = \frac{\sum_{j=1}^{m} \eta_{jk} (q_j - \mu_k')(q_j - \mu_k')^T}{\sum_{j=1}^{m} \eta_{jk}} \quad (8)$$

$$\chi_k' = \frac{\sum_{j=1}^{m} \eta_{jk}}{m}. \quad (9)$$

In Step 4, repeat Steps 2 and 3 until the Gaussian distribution parameters of each category remain unchanged. Step 5, for each data point q_j, calculates its posterior probability value belonging to each Gaussian distribution through Step 3, then assigns it to the category represented by the Gaussian distribution with the highest probability, and marks this category as a cluster mark λ_j. The cluster mark λ_j is calculated as follows:

$$\lambda_j = \operatorname{argmax}\ \eta_{jk}. \quad (10)$$

In this way, the category that each data point belongs to can be obtained. Finally, we use the clear and blurry images obtained in Step 5 as the training set and the uncertain images as the test set. By employing the support vector machine (SVM) classifier, we further classify the uncertain images into clear and blurry images to obtain the final clear and blurry images.

C. Multigranularity Feature Augmentation Module (MFAM)

In the RS ship detection task, significant variations in ship sizes and the frequent presence of small objects make it common to use the FPN structure for extracting features of different granularities. As shown in Fig. 2, the feature map of the top layer P_5 in FPN is propagated in a top-down manner and fused with the feature maps of the lower layers P_4, P_3, P_2

Algorithm 1 Automatic Classification of Image Blur Level Based on 3WD

Input: All images $I = (i_1, i_2, \ldots, i_m)$, mean vector μ_k, covariance matrix Σ_k, and mixing coefficient χ_k

Output: Clear images I_C, blurry images I_B

1: For all images, use Eqs. (1) to (5) compute the SMD2 and Tenengrad values for each image, get $s_j = SMD2(i_j)$, $t_j = Tenengrad(i_j)$, $q_j = [s_j, t_j]$ $(j = 1, 2, \ldots, m)$, note $Q = (q_1, q_2, \ldots, q_m)$
2: Based on 3WD theory, initialize the model parameters $\{(\mu_k, \Sigma_k, \chi_k) \mid 1 \leq k \leq 3\}$ of the Gaussian mixture distribution
3: Repeat
4: **for** $j = 1, 2, \ldots, m$ **do**
5: Calculate the posterior probability generated by each mixed component of q_j according to Eq. (6), that is $\eta_{jk} = p(z_j = k|q_j)$ $(1 \leq k \leq 3)$
6: **end for**
7: **for** $k = 1, 2, 3$ **do**
8: Update μ'_k, Σ'_k, and χ'_k according to Eqs. (7) to (9)
9: **end for**
10: Until the stop condition is met: the current μ_k, Σ_k and χ_k remain unchanged
11: Cluster division $C_k = \phi (1 \leq k \leq 3)$
12: **for** $j = 1, 2, \ldots, m$ **do**
13: Calculate the cluster mark λ_j of q_j according to Eq. (10) and assign q_j to the corresponding cluster $C_{\lambda_j} = C_{\lambda_j} \bigcup \{q_j\}$, and there is a one-to-one correspondence between q_j and i_j
14: **end for**
15: Get $C = \{C_1, C_2, C_3\}$, namely clear images C_1, uncertain images C_2 and blurry images C_3. And the uncertain images C_2 is divided into clear images C_{11} and blurry images C_{33} using a SVM classifier
16: **return** $I_C = C_1 \cup C_{11}$, $I_B = C_3 \cup C_{33}$

layer by layer. However, this layer-by-layer fusion strategy has two limitations. First, the features of the lower layers cannot directly and effectively affect the high-level features. Second, reducing feature dimensions leads to the loss of high-level feature P_5 information. As both low- and high-level features are beneficial for detecting small and large ships, respectively, these issues may significantly affect the detection performance of the detection model for ships of different sizes.

To address these issues, this article proposes an MFAM based on the idea of multigranularity feature learning. The low-level features cannot affect high-level features in FPN, hence a bottom-up feature fusion layer (BF^{2}L) is included, which can facilitate the transfer of information from the bottom to the top, as shown in Fig. 5(b). It enhances the interaction between low- and high-level features. In addition, to solve the information loss of high-level feature P_5, ACA is introduced as shown in Fig. 5(a), which improves the feature P_5 by incorporating distinct channel and position information into the original branch, thereby reducing the loss of channel and coordinate information of P_5. In this way, it preserves

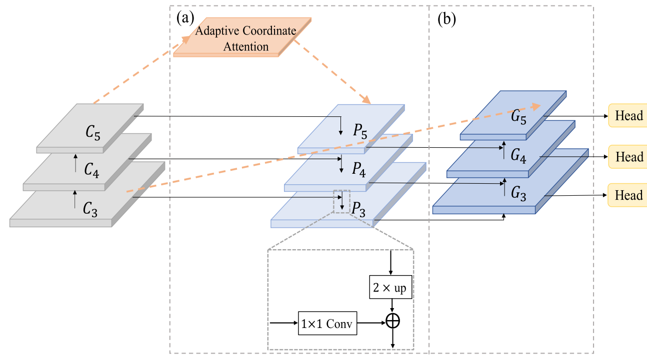


Fig. 5. Structure of the MFAM. (a) ACA. (b) BF^{2}L.

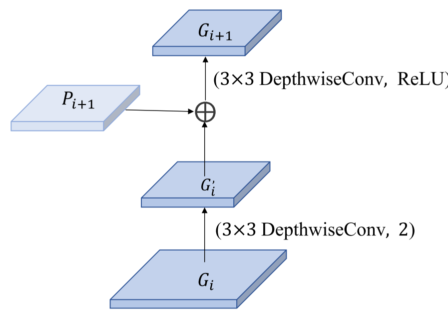


Fig. 6. Fusion method of BF^{2}L. "(3 × 3 DepthwiseConv, 2)" refers to 3 × 3 depthwise convolution with stride 2, "ReLU" is the activation function, and ⊕ means addition.

critical high-level features. Meanwhile, the adaptive feature fusion significantly improves the detection performance of ships of different sizes, making the MFAM more advantageous than FPN.

1) Bottom-Up Feature Fusion Layer: A BF^{2}L is incorporated into the FPN architecture. On the one hand, the lowest layer C_3 is directly transferred to the highest layer G_5. On the other hand, the shallower layer G_i and the deeper layer P_{i+1} are fused to generate the subsequent layer G_{i+1}. Therefore, three feature maps, namely, $\{G_3, G_4, G_5\}$, are obtained.

The fusion method of BF^{2}L is shown in Fig. 6. First, G_i is downsampled using a 3 × 3 depthwise convolution [59] with stride 2, yielding G_i'. Next, feature fusion is performed on P_{i+1} and G_i' using the addition operation. The resulting output is then passed through a 3 × 3 depthwise convolution with a stride of 1 to obtain G_{i+1}. Both G_4 and G_5 adopt the fusion method in Fig. 6, and G_3 is a direct copy of the value of P_3. Applying depthwise convolution to BF^{2}L can effectively fuse features and reduce the number of parameters during convolution.

2) Adaptive Coordinate Attention: As shown in Fig. 7, let $C_5 \in \mathbb{R}^{C \times H \times W}$ denote the input feature map, where C, H, and W denote the number of input channels, height, and width, respectively. First, it uses ratio-invariant adaptive pooling (RAP) to generate multigranularity feature maps with different scales $\{\beta_1 \times S, \beta_2 \times S, \ldots, \beta_n \times S\}$ and performs 1 × 1 convolution to obtain the same channel dimension as 256. Then, these feature maps are upsampled to the scale of $S = H \times W$ using bilinear interpolation. Finally, the fusion is

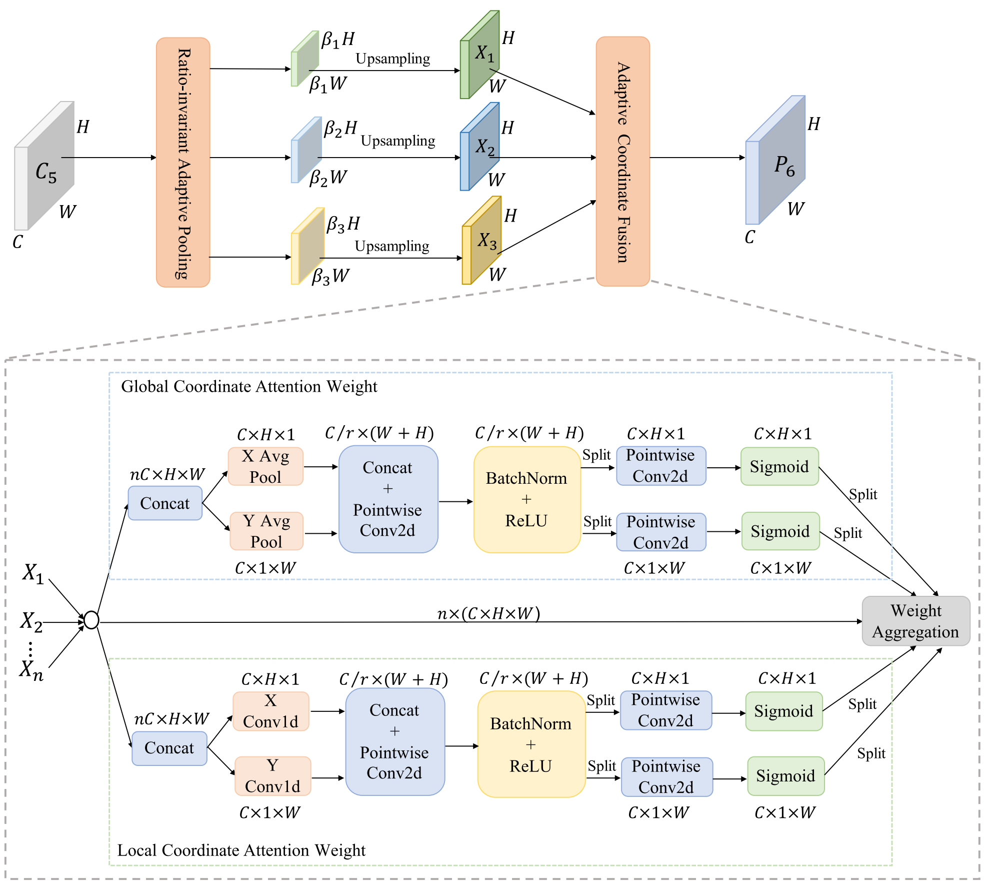


Fig. 7. Structure of the ACA. ACA consists of RAP and ACF. First, RAP uses adaptive pooling to process the input feature map C_5 to generate multigranularity features. Then, ACF can adaptively adjust the fusion weight and fuse multigranularity features with global coordinate attention weight and local coordinate attention weight to generate a richer and more robust feature representation. "X Avg Pool" and "Y Avg Pool" refer to 1-D horizontal global pooling and 1-D vertical global pooling, respectively. "X Conv1d" and "Y Conv1d" refer to 1-D convolution with convolution kernels $(H, 1)$ and $(1, W)$, respectively.

performed by the adaptive coordinate fusion (ACF) to obtain P_6 with multigranularity context information.

Our ACF is an upgrade over the CAM [60] to capture the local-to-global position information of RS ship targets by combining various pooling sizes and point convolutions, thereby capturing richer feature representations. The details of the ACF module are shown in Fig. 7. Specifically, the ACF module takes a single feature as input and generates global and local coordinate attention weights for each feature. Then, it uses the generated weights to aggregate contextual features into P_6 and assigns multigranularity global-to-local contextual information to P_6. To make the ACF module as lightweight as possible, we only add local context to the global context within the attention module and select pointwise convolution as the aggregator for global and local coordinate contexts.

a) Global coordinate attention weight: First, we use concatenation to perform channel fusion on feature maps $\{\beta_1 \times S, \beta_2 \times S, \ldots, \beta_n \times S\}$ of different granularities. Second, the fused feature map X uses two spatial pooling kernels $(H, 1)$ and $(1, W)$ to encode each channel along the horizontal and vertical coordinates, respectively. The output of the cth channel at width h can be expressed as

$$z_c^h = \frac{1}{W} \sum_{0 \le i \le W} x_c(h, i). \tag{11}$$

Similarly, the cth channel output of height w can also be written as

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j \le H} x_c(j, w). \tag{12}$$

We then aggregate features along two spatial directions through Eqs. (11) and (12) to generate a pair of direction- and position-aware feature maps $Z^h \in \mathbb{R}^{C \times H \times 1}$ and $Z^w \in \mathbb{R}^{C \times 1 \times W}$, which can effectively obtain global receptive fields and encode accurate position information. To fully leverage feature representations with global receptive fields and accurate location information, we apply a shared 1×1 pointwise convolutional layer $N_1 \in \mathbb{R}^{C \times (C/r) \times 1 \times 1}$ to both Z^h and Z^w. Finally, we use two 1×1 pointwise convolutional layers $N_h \in \mathbb{R}^{(C/r) \times C \times 1 \times 1}$ and $N_w \in \mathbb{R}^{(C/r) \times C \times 1 \times 1}$ to process the intermediate feature map obtained after the N_1 operation so that the transformed feature map has the same number of channels as the input feature map X. This allows us to calculate the global coordinate attention weight as follows:

$$g^h = \sigma\left(N_h\left(\delta\left(B\left(N_1\left(Z^h\right)\right)\right)\right)\right) \tag{13}$$

$$g^w = \sigma(N_w(\delta(B(N_1(Z^w))))) \tag{14}$$

$$g_c(i, \ j) = g_c^h(i) \times g_c^w(j) \tag{15}$$

where σ represents the sigmoid activation function, δ represents the ReLU activation function, and $B(\cdot)$ denotes batch normalization. g^h and g^w denote the global coordinate attention weights in the horizontal and vertical directions, respectively. $g_c(i, j)$ represents the global coordinate attention weight. The hyperparameter r is the channel reduction ratio, and we set r to 32.

b) Local coordinate attention weight: Similarly, we also use the concatenation to perform channel fusion on feature maps $\{\beta_1 \times S, \beta_2 \times S, \ldots, \beta_n \times S\}$ of different granularities. We then utilize two 1-D convolutions with kernel sizes of $(H, 1)$ and $(1, W)$ to convolve the fused feature map X in each channel horizontally and vertically, resulting in 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, we apply a shared 1×1 pointwise convolutional layer $N_2 \in \mathbb{R}^{C\times(C/r)\times 1\times 1}$ to both X^h and X^w. Finally, we use two 1×1 pointwise convolutional layers $\widetilde{N}_h \in \mathbb{R}^{(C/r)\times C\times 1\times 1}$ and $\widetilde{N}_w \in \mathbb{R}^{(C/r)\times C\times 1\times 1}$ to process the intermediate feature obtained after the N_2 operation, so that the transformed feature has the same number of channels as the input feature X. This allows us to calculate the local coordinate attention weight as follows:

$$l^h = \sigma\left(\widetilde{N}_h\left(\delta\left(B\left(N_2\left(X^h\right)\right)\right)\right)\right) \tag{16}$$

$$l^w = \sigma\left(\widetilde{N}_w(\delta(B(N_2(X^w))))\right) \tag{17}$$

$$l_c(i, j) = l_c^h(i) \times l_c^w(j) \tag{18}$$

where σ represents the sigmoid activation function, δ represents the ReLU activation function, and $B(\cdot)$ denotes batch normalization. l^h and l^w denote the local coordinate attention weights in the horizontal and vertical directions, respectively. $l_c(i, j)$ represents the local coordinate attention weight. The hyperparameter r is the channel reduction ratio, and we set r to 32.

After computing the global and local coordinate attention weights as described above, we obtain a new feature Y that contains both local and global coordinate attention. This can be written as

$$y = x_c(i,\ j) \otimes (g_c(i,\ j) \oplus l_c(i,\ j)) \tag{19}$$

where $\oplus$ and $\otimes$ denote broadcast addition and elementwise multiplication, respectively.

Therefore, the feature map P_6 generated by ACA contains rich multigranularity global and local contextual information. To alleviate the information loss due to the reduced number of channels, we combine P_6 with P_5 and fuse them with other lower level features. This fusion method enhances the perceptual ability of the model and the representation ability of multigranularity features, thereby improving the performance of the model.

IV. Experiments and Results

A. Datasets and Evaluation Metrics

1) Datasets: FGSD2021 [27] is a high-resolution ship dataset with fixed GSD obtained from publicly available Google Earth. It contains 636 normalized GSD images. It has a width of 157–7789 pixels, an average width of 1202 pixels, and a height of 224–6506 pixels. We use 424 training and 212 test images, respectively. For single-scale experiments, we resize the image to 512 × 512. For multiscale experiments, the original images are initially resized at three scales (0.5, 1.0, and 1.5) before being cropped into 1024 × 1024 patches using a stride of 512.

HRSC2016 [61] is a high-resolution RS ship dataset marked with a rotating box. It includes 1061 RS images obtained from Tianditu, and its size ranges from 300 × 300 to 1500 × 900. We train with a training set (436 images) and a validation set (181 images), test with a test set (444 images), and rescale all images to (512 × 512).

UCAS-AOD [62] is an RS dataset for aircraft and vehicle detection. It has 1510 images and 14 596 instances, including 510 car and 1000 airplane images. Previous research on ship detection [27] and studies that involve partial ship detection [19], [22] also use UCAS-AOD to verify the generalization ability of the model. Similar to the ship datasets FGSD2021 and HRSC2016, UCAS-AOD exhibits characteristics such as image blurring and inconsistent target sizes. In our experiments, we divide the dataset into training and testing sets on a 7:3 scale and cut the size of each image to 512 × 512.

2) Evaluation Metrics: We mainly utilize the intersection over union (IoU) between rotating BBoxes to distinguish detection results and adopt the widely used mean average precision (mAP) as the evaluation metric for RS ship detection methods. The IoU is calculated by dividing the overlapping area of the detection box with the ground-truth box by their union area. The detection box is labeled as true positive (TP) if the IoU between the two boxes exceeds a threshold. Otherwise, it is labeled as a false positive (FP). A ground-truth box is labeled as a false negative (FN) if it has no corresponding detections. The mAP is obtained by calculating the precision $P = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$ and the recall rate $R = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$, which can be expressed as $\mathrm{mAP} = (1/A)\sum_{a=1}^{A}\int P_a(R_a)dR_a$, where A represents the total number of categories, and P_a and R_a denote the precision and recall for each category a, respectively.

For FGSD2021, we choose to use the PASCAL VOC2007 metric with an IoU threshold of 0.5 to calculate the mAP. For HRSC2016 and UCAS-AOD, we select the PASCAL VOC2007 and PASCAL VOC2012 metrics with an IoU threshold of 0.5 to compute the mAP. In addition, we also consider model parameters (Param), giga floating-point operations per second (GFLOPs), runtime, and frames per second (FPS) to verify the efficiency of the methods.

B. Implementation Details

Our proposed 3WM-AugNet builds on the baseline model S^{2}A-Net, including its network architecture and most parameter settings. We keep the regression objective and loss function of 3WM-AugNet the same as S^{2}A-Net. During the inference, an image is passed through the entire network without complicated RoI operations, and we select top-2000 predictions and employ NMS to produce final detections. We use two Tesla V100s 32 GB for training and one Tesla V100 32 GB for testing. We choose ResNet50 and ResNet101 as the backbone networks for a fair comparison with the other methods. In Algorithm 1, mean vector $\mu_1 = q_6$, $\mu_2 = q_{22}$,

TABLE I

COMPARISON OF RESULTS UNDER DIFFERENT DEHAZING ALGORITHM SETTINGS IN BCDM. BCDM MEANS THE BLURRED CLASSIFICATION AND DEBLURRING MODULE. BOLD HIGHLIGHTS THE BEST RESULTS

Setting	Deblurring Algorithm	Param(M)	GFLOPs	mAP(%)
Baseline	-	35.02	189.71	80.19
BCDM	SRN	41.82	227.53	83.01
BCDM	DeblurGAN-v2	95.90	520.21	83.05
BCDM	**MT-RNN**	37.66	207.93	**83.08**
BCDM	MIMO-UNet	51.12	277.34	83.08
BCDM	XYDeblur	46.74	254.18	83.06

and $\mu_3 = q_{27}$, covariance matrix

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 0.1 & 0.0 \\ 0.0 & 0.1 \end{bmatrix}$$

and mixing coefficient $\chi_1 = \chi_2 = \chi_3 = (1/3)$, and choose to use a linear classifier SVM. In the loss function, we set the loss balance parameter λ to 1 and the hyperparameters α and γ of focal loss $\mathcal{L}_c$ to 0.25 and 2.0, respectively. The SGD optimizer is used with an initial learning rate of 0.01, the learning rate is divided by 10 for each decay step, and the batch size is 4. Momentum and weight decay are 0.9 and 0.0001, respectively. We train on FGSD2021, HRSC2016, and UCAS-AOD datasets for 100, 60, and 80 epochs, respectively. All experiments are conducted based on MMDetection [63] and default to a single-scale experiment unless otherwise specified.

C. Parametric Analysis

In this section, we conduct experiments on the FGSD2021 dataset to investigate the effect of different parameter settings on the performance of the proposed module.

1) Effect of Different Deblurring Algorithms on BCDM: To investigate the impact of different deblurring algorithms on the detection performance of BCDM, we compare MT-RNN with other deblurring algorithms on the FGSD2021 dataset and evaluate the influence of MT-RNN on BCDM. During the comparison process, we keep all other settings the same and only replace MT-RNN [46] with SRN [44], DeblurGAN-v2 [45], MIMO-UNet [47], or XYDeblur [48]. The results are shown in Table I. The experimental results indicate that incorporating MT-RNN in BCDM improves the baseline mAP by 2.89%. In addition, using MT-RNN enables the model to achieve the best accuracy, reaching 83.08% mAP, while its parameters and GFLOPs are only slightly higher than the baseline model, with an increase of merely 2.64M parameters and 18.22 GFLOPs in computational cost. Compared to SRN, DeblurGAN-v2, and XYDeblur, MT-RNN achieves a higher accuracy by 0.07%, 0.03%, and 0.02%, respectively. In addition, compared to MIMO-UNet, MT-RNN reduces parameters and GFLOPs by 13.46M and 69.41, respectively, while achieving the same level of accuracy as MIMO-UNet. This further proves the effectiveness of MT-RNN in improving the detection performance of the model.

TABLE II

COMPARISON OF RESULTS UNDER DIFFERENT POOLING SETTINGS IN ACF. ACF MEANS THE ADAPTIVE COORDINATE FUSION. SUM REFERS TO THE ELEMENTWISE SUMMATION. GMP, GAP, AND RAP REPRESENT GLOBAL MAX POOLING, GLOBAL AVERAGE POOLING, AND RATIO-INVARIANT ADAPTIVE POOLING, RESPECTIVELY

Setting	Pooling Type	β	mAP(%)
Baseline	-	-	80.19
sum	GMP	-	78.62
sum	GAP	-	82.31
sum	RAP	0.1,0.2,0.3	83.14
ACF	RAP	0.1	83.27
ACF	RAP	0.1,0.2	83.85
ACF	RAP	0.1,0.2,0.3	85.93
ACF	RAP	0.1,0.2,0.3,0.4	86.65
ACF	RAP	**0.1,0.2,0.4**	**86.57**
ACF	RAP	0.1,0.2,0.5	86.55
ACF	RAP	0.1,0.2,0.6	86.21
ACF	PSP	-	85.94

TABLE III

COMPARISON OF RESULTS UNDER DIFFERENT REDUCTION RATIO r SETTINGS IN ACF. ACF MEANS ADAPTIVE COORDINATE FUSION. BOLD HIGHLIGHTS THE BEST RESULTS

Setting	Reduction rate r	Param(M)	GFLOPs	mAP(%)
Baseline	-	35.02	189.71	80.19
ACF	8	37.51	207.35	85.20
ACF	16	35.93	200.78	85.99
ACF	**32**	35.58	198.12	**86.57**
ACF	64	35.19	194.86	85.73

2) Effect of Different Pooling Types on ACA: To study the impact of different pooling types on the detection performance of ACA, two different types of pooling are compared by replacing RAP, where the reduction ratio r is set to 32. Since there is only one branch, we use a summation operation for feature fusion. Table II shows that global max pooling (GMP) reduces the baseline by 1.57% mAP, while global average pooling (GAP) increases the baseline by 2.12% mAP. Therefore, GAP is found to be more effective than GMP in detecting RS ships. Then, we replace GAP with RAP and set three β values of 0.1, 0.2, and 0.3, respectively. In the fifth row of Table II, RAP improves by 2.95% mAP and 0.83% mAP compared to the baseline and GAP, respectively, indicating the effectiveness of RAP. Finally, combining ACF with RAP at the same β value results in an experimental outcome of 85.93% mAP, which is 5.74% mAP higher than the baseline.

We also explore the impact of different values of β on the model detection performance. Based on the results in Table II, we choose to set three values of β to achieve a better balance between model complexity and accuracy. The three β's are set to 0.1, 0.2, and 0.4, respectively. To further verify the

TABLE IV

COMPARISON OF RESULTS UNDER DIFFERENT TYPES OF CONVOLUTION SETTINGS IN BF^2L. BF^2L MEANS THE BOTTOM-UP FEATURE FUSION LAYER. CONV, DILATEDCONV, AND DEPTHWISECONV REPRESENT CONVOLUTION, DILATED CONVOLUTION, AND DEPTHWISE CONVOLUTION, RESPECTIVELY. BOLD HIGHLIGHTS THE BEST RESULTS

Setting	Convolution type	Param(M)	GFLOPs	mAP(%)
Baseline	-	35.02	189.71	80.19
BF^2L	Conv	36.90	203.53	80.70
BF^2L	DilatedConv	36.37	200.93	81.16
BF^2L	**DepthwiseConv**	35.23	195.09	**82.71**

TABLE V

RESULTS OF DIFFERENT ABLATION EXPERIMENTS ON FGSD2021. ✓MEANS TO USE THIS MODULE, BASELINE MEANS S^2A-NET, BCDM MEANS THE BLURRED CLASSIFICATION AND DEBLURRING MODULE, BF^2L MEANS THE BOTTOM-UP FEATURE FUSION LAYER, AND ACA MEANS ADAPTIVE COORDINATE ATTENTION. MFAM CONSISTS OF BF^2L AND ACA. ABLATION EXPERIMENTS USING SINGLE-SCALE TRAINING AND TESTING

Baseline	BCDM	BF^2L	ACA	Param(M)	GFLOPs	mAP(%)	FPS
✓				35.02	189.71	80.19	33.12
✓	✓			37.66	207.93	83.08	28.52
✓		✓		35.23	195.09	82.73	31.09
✓			✓	35.58	196.64	86.57	30.72
✓		✓	✓	35.79	198.12	89.45	30.50
✓	✓	✓	✓	38.43	212.35	91.53	26.49

effectiveness of RAP, we use PSP [64] with pooling kernel sizes of 1×1, 2×2, and 3×3 to replace RAP, and the results show that it is 0.63% mAP worse than RAP.

3) Effect of Different Reduction Ratios r on ACA: In Section III-C, we propose an ACA composed of RAP and ACF, where ACF introduces a hyperparameter reduction ratio r. Since different reduction ratios r have a certain impact on the performance of ACA, we conduct a series of experiments to determine the optimal r value and recorded the performance and parameter count under different r values, as shown in Table III. We found that, as r doubles, the parameter count of the model also significantly increases, but the performance initially improves and then deteriorates. On the contrary, a small r enables the convolutional layer to better eliminate redundant channel information, thus enhancing the performance of the model. However, increasing r may result in the loss of useful features, leading to a decline in performance. To better balance the number of parameters and the performance of the model, we set r to 32, achieving a better performance of 86.57% mAP. Compared to the baseline, the mAP of the model increases by 6.38% while only adding 0.56M parameters and 8.41 GFLOPs.

4) Effect of Different Convolution Types on BF^2L: We compare DepthwiseConv with other convolution types to evaluate its effectiveness in improving BF^2L detection performance. While keeping all other settings unchanged, we only replace

TABLE VI

RESULTS OF TIME EFFICIENCY OF BCDM IN TRAINING AND TESTING ON DIFFERENT DATASETS. BCDM MEANS THE BLURRED CLASSIFICATION AND DEBLURRING MODULE

Datasets	Setting	Params(M)	Runtime(s)	
			Train	Test
FGSD2021	BCDM	2.64	0.15	0.11
HRSC2016	BCDM	2.64	0.19	0.15
UCAS-AOD	BCDM	2.64	0.19	0.16

DepthwiseConv with Conv or DilatedConv, and the experimental results are shown in Table IV. The results indicate that DepthwiseConv significantly improves the detection performance of the model to 82.71% mAP, which is 2.52% mAP higher than the baseline and only increases 0.21M parameters and 5.38 GFLOPs. On the contrary, replacing DepthwiseConv with Conv and DilatedConv results in the inferior performance of the model, reaching only 80.70% mAP and 81.16% mAP, respectively, and both require more parameters and GFLOPs than DepthwiseConv. This demonstrates the effectiveness of DepthwiseConv in improving the detection performance of the model.

D. Ablation Study

Taking S^2A-Net as the baseline, we propose two novel modules, BCDM and MFAM, where MFAM is composed of ACA and BF^2L. To verify the effectiveness of different modules, we conduct an ablation study on FGSD2021, and the results are shown in Table V.

1) S^2A-Net as Baseline: As a one-stage alignment network, S^2A-Net uses the combination of FAM and ODM to detect rotating objects in RS images efficiently. Table V shows that S^2A-Net achieves 80.19% mAP on FGSD2021, which shows that our baseline is competitive.

2) Effectiveness of the BCDM: Before we add BCDM to the backbone of the baseline, other settings remained unchanged, and its effectiveness was verified on FGSD2021, as shown in Table V. Compared with the baseline, the detection result of the model adopting BCDM improves by 2.89%, from 80.19% mAP to 83.08% mAP. In addition, the parameter and GFLOPs of the model only increased by 2.64M and 18.22, respectively, indicating a significant performance improvement. Due to the blurred characteristics of RS images, existing algorithms perform deblurring on all images, resulting in some originally clear images being excessively deblurred, which reduces their clarity and is not conducive to subsequent detection. Therefore, incorporating BCDM into the model enables it to selectively deblur only the necessary images (blurry images), improving the detection accuracy.

In BCDM, all input unclassified RS images are first divided into clear, uncertain, and blurry images by combining the 3WD theory. Then, the SVM classifier is used to further classify uncertain images. Finally, only the blurry images are subjected to deblurring processing using the MT-RNN algorithm, while the clear images remain unchanged. Fig. 8 displays the results of the blur level classification of images in the FGSD2021

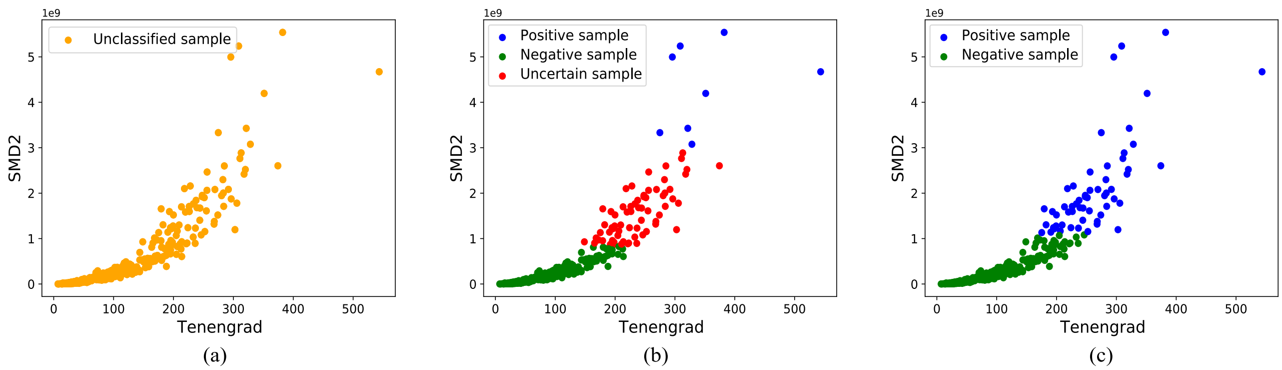


Fig. 8. Blur level classification results of images in the FGSD2021 dataset based on 3WD. First, we calculate the SMD2 and Tenengrad values for all input images, taking the obtained SMD2 values and Tenengrad values as data points and combining them to form a dataset, as shown in (a). Next, the dataset is processed using the Gaussian mixture clustering algorithm and 3WD theory, categorizing the images into three classes: clear, uncertain, and blurry, as shown in (b). Subsequently, clear and blurry images are obtained as the training set, while uncertain images are used as the testing set. By utilizing the SVM classifier, uncertain images are further categorized into clear and blurry images, yielding the final classification results depicted in (c). Unclassified samples (orange dots) represent all input images. Positive samples (blue dots) represent clear images, uncertain samples (red dots) represent uncertain images, and negative samples (green dots) represent blurry images.

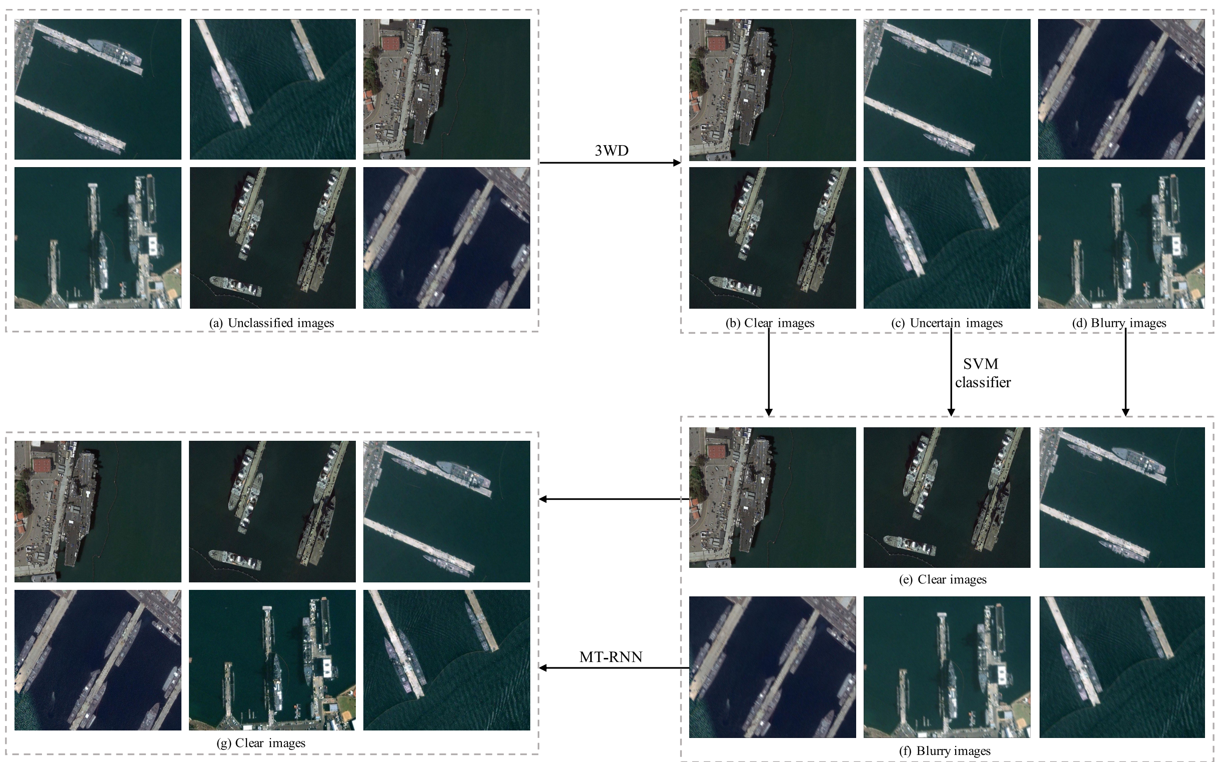


Fig. 9. Example of BCDM operation process in the FGSD2021 dataset. First, the 3WD theory is employed to classify all input unclassified images into clear, uncertain, and blurry categories. Second, the SVM classifier is used to further divide the uncertain images into clear and blurry images. Next, the MT-RNN algorithm is exclusively employed to deblur the blurry images, while the clear images remain unchanged. Finally, this process yields clear images. (a) Unclassified images. (b) Clear images. (c) Uncertain images. (d) Blurry images. (e) Clear images. (f) Blurry images. (g) Clear images.

dataset based on 3WD. Fig. 9 shows an example of the BCDM operation process in the FGSD2021 dataset and presents images classified as clear, uncertain, and blurry. Specifically, clear images refer to images in which most of the details and features of the ships in the image can be clearly distinguished. Uncertain images refer to images with moderate clarity, and blurry images refer to images with low clarity.

3) Efficiency of the BCDM: Given that BCDM should operate on each training and testing image separately in the training and testing stages, its time efficiency should also be studied to evaluate the practicality and feasibility of BCDM in practical applications. We conduct comprehensive experiments on three datasets (FGSD2021, HRSC2016, and UCAS-AOD) with varying scales and complexities. To ensure the reliability of the results, we run multiple trials on each dataset and calculate the average running time as the measure of time efficiency.

As shown in Table VI, the average time cost for BCDM training and testing on the three datasets is about 0.18 and 0.14 s, respectively. Compared with the test phase, training BCDM shows an additional time cost when training an SVM classifier using clear and blurry images. On the one hand, the

TABLE VII

DETECTION PERFORMANCE OF DIFFERENT METHODS ON FGSD2021. THE SHORT NAME OF THE CLASS IS DEFINED AS (ABBREVIATION-FULL NAME): AIR- AIRCRAFT CARRIERS, WAS-WASP CLASS, TAR-TARAWA CLASS, AUS-AUSTIN CLASS, WHI-WHIDBEY ISLAND CLASS, SAN-SAN ANTONIOCLASS, NEW-NEWPORT CLASS, TIC-TICONDEROGA CLASS, BUR-ARLEIGH BURKE CLASS, PER-PERRY CLASS, LEW-LEWIS, CLARK CLASS, SUP-SUPPLY CLASS, KAI-HENRY J. KAISER CLASS, HOP-BOB HOPE CLASS, MER-MERCY CLASS, FRE-FREEDOM CLASS, IND-INDEPENDENCE CLASS, AVE-AVENGER CLASS, SUB-SUBMARINE, AND OTH-OTHER. * MEANS THAT THE DATASET USES THE PREPROCESSING MODULE BCDM. † MEANS MULTISCALE TRAINING AND TESTING. *† MEANS THAT THE DATASET USES THE PREPROCESSING MODULE BCDM AND PERFORMS MULTISCALE TRAINING AND TESTING. mAP(07) REFERS TO THE mAP COMPUTED ON THE PASCAL VOC2007. BOLD HIGHLIGHTS THE BEST RESULTS

Method	Backbone	Air	Was	Tar	Aus	Whi	San	New	Tic	Bur	Per	Lew	Sup	Kai	Hop	Mer	Fre	Ind	Ave	Sub	Oth	mAP(07)(%)	FPS
Two-stage methods																							
R^2CNN [15]	ResNet50	89.91	80.90	80.48	79.41	87.01	87.77	44.20	89.03	89.62	79.45	80.40	47.73	81.52	87.36	**100**	82.44	**100**	66.37	50.91	57.19	78.09	10.31
ROI Transformer [16]	ResNet50	90.90	88.55	87.17	89.51	78.53	88.80	81.77	89.69	89.83	90.44	71.71	74.65	73.72	81.60	78.58	**100**	75.56	78.44	68.01	66.92	82.22	19.19
Oriented R-CNN [26]	ResNet50	90.85	89.71	81.46	81.06	79.57	88.23	98.92	89.90	90.58	87.83	60.44	73.88	81.81	86.73	**100**	60.03	**100**	79.37	66.85	63.74	82.55	27.43
DEA-Net [33]	ResNet50	90.43	91.37	84.61	93.50	88.71	94.54	92.11	90.73	92.35	88.92	60.62	81.63	85.37	90.32	99.66	83.10	98.52	76.63	68.45	69.18	86.04	12.08
SCRDet++ [19]	ResNet50	93.50	90.11	87.21	89.57	87.23	96.35	90.14	90.82	92.33	88.39	62.67	85.73	88.74	95.55	**100**	85.16	**100**	82.56	74.34	77.36	87.89	16.41
SCRDet++* [19]	ResNet50	95.34	91.70	88.56	90.47	88.49	97.06	93.32	91.37	92.33	90.69	75.47	86.93	89.93	95.79	**100**	87.00	**100**	86.71	78.83	80.76	90.04	12.52
One-stage methods																							
CSL [20]	ResNet50	89.67	81.25	77.23	80.19	71.43	77.24	52.68	87.71	87.74	74.15	57.07	**97.23**	77.64	80.46	**100**	72.71	**100**	32.58	36.98	40.74	73.74	10.38
R^3Det-DCL [21]	ResNet50	89.91	81.37	78.62	80.70	78.01	87.88	49.79	78.73	87.15	76.11	60.62	76.85	90.42	80.03	78.79	77.88	**100**	37.13	31.23	45.58	73.34	10.03
R^3Det [22]	ResNet50	90.93	80.90	81.51	90.05	79.26	87.52	29.47	77.36	89.44	69.73	59.87	67.30	80.66	76.84	72.69	83.31	90.85	38.43	23.14	40.04	70.47	14.01
RSDet [23]	ResNet50	89.90	80.43	75.76	77.28	78.64	88.82	26.06	84.73	87.63	75.16	55.11	74.37	89.67	89.33	**100**	86.36	**100**	27.64	37.61	50.63	73.76	15.38
RSDet* [23]	ResNet50	90.32	81.43	77.76	79.34	83.78	89.54	38.06	84.99	88.93	78.34	55.11	75.01	89.63	89.73	**100**	87.09	**100**	34.55	39.23	52.14	75.75	11.12
Anchor-free methods																							
BBAVectors [24]	ResNet50	**99.53**	90.92	75.86	94.27	90.93	52.88	88.47	90.03	80.41	72.17	76.93	88.19	**99.64**	**100**	93.97	**100**	74.54	58.91	63.12	81.84	83.63	18.53
CHPDet [27]	DLA34	90.90	90.40	89.60	89.30	89.60	99.10	99.40	90.20	90.20	90.30	70.70	87.90	89.20	96.50	**100**	85.10	**100**	84.40	68.50	56.90	87.91	15.40
Oriented RepPoint [30]	ResNet50	91.16	89.21	85.61	89.30	87.59	93.13	94.15	91.52	88.73	83.30	71.37	81.14	89.42	91.48	95.60	82.63	**100**	86.55	64.74	57.48	85.71	36.73
GF-CSL [34]	ResNet50	92.56	90.33	86.61	90.52	88.16	95.32	97.90	89.77	91.21	86.94	69.65	85.63	92.73	92.54	99.71	85.12	98.58	86.67	79.44	70.36	88.49	**40.32**
GF-CSL* [34]	ResNet50	96.04	95.15	89.73	96.75	89.98	96.22	98.73	90.62	91.18	88.94	72.65	87.63	94.65	92.87	**100**	88.08	**100**	88.31	80.94	73.54	90.60	35.64
GF-CSL*† [34]	ResNet50	97.31	95.82	90.63	96.98	93.45	**99.53**	**99.41**	92.11	93.03	95.13	85.44	95.77	91.25	96.73	**100**	92.01	**100**	**90.33**	81.06	75.32	93.07	14.31
S^2A-Net [28]	ResNet50	90.91	81.43	73.25	89.11	80.87	89.92	81.23	89.16	90.67	88.93	60.52	75.86	81.64	89.20	**100**	68.63	90.88	61.31	55.65	64.72	80.19	33.12
S^2A-Net* [28]	ResNet50	93.20	83.48	77.09	92.64	84.99	91.72	82.73	90.35	90.97	91.03	60.74	78.06	86.71	90.34	**100**	74.83	93.58	67.03	61.48	70.63	83.08	28.52
3WM-AugNet (Ours)	ResNet50	96.01	92.07	88.78	95.96	90.01	95.47	87.12	92.97	93.70	93.21	62.04	87.76	90.58	96.84	97.64	86.74	99.98	79.32	79.12	83.70	89.45	30.50
3WM-AugNet (Ours)*	ResNet50	98.52	94.67	90.65	97.62	91.36	96.88	89.78	94.12	93.82	95.09	67.82	88.74	93.78	97.34	**100**	89.52	**100**	81.14	85.72	84.02	91.53	26.49
3WM-AugNet (Ours)*†	ResNet50	98.64	**96.77**	**94.01**	**97.58**	**97.32**	97.01	92.85	**94.83**	**94.77**	**95.73**	**88.61**	89.02	95.32	98.55	**100**	94.20	**100**	84.23	**90.19**	**84.57**	**94.21**	9.02

TABLE VIII

COMPARISON OF THE NUMBER OF PARAMETERS AND COMPUTATION OF DIFFERENT METHODS ON FGSD2021. * MEANS THAT THE DATASET USES THE PREPROCESSING MODULE BCDM. mAP(07) REFERS TO THE mAP COMPUTED ON THE PASCAL VOC2007. BOLD HIGHLIGHTS THE BEST RESULTS

Method	BackBone	Param(M)	GFLOPs	mAP(07)(%)
DEA-Net [33]	ResNet50	59.90	237.21	86.04
R^3Det-DCL [21]	ResNet50	37.31	**152.85**	73.34
GF-CSL [34]	ResNet50	49.71	225.63	88.49
S^2A-Net [28]	ResNet50	**35.02**	189.71	80.19
3WM-AugNet* (Ours)	ResNet50	38.43	212.35	**91.53**

training phase of BCDM includes Gaussian mixture clustering and SVM training. Its time complexity is $O(m)$ (m represents the total number of images). On the other hand, the testing phase of BCDM includes testing of both SVM and MT-RNN. The testing time complexity of SVM is $O(m')$ (m' denotes the number of uncertain images and $m' < m$). The testing time complexity of MT-RNN is 18.22 GFLOPs. Given the large-scale and complex characteristics of RS images, these processing times are considered acceptable. Meanwhile, BCDM is designed to preprocess input images in a reasonable way to improve detection accuracy. Therefore, it is feasible to sacrifice a small amount of processing time to achieve improved performance.

4) Effectiveness of ACA and BF^{2}L: Table V demonstrates that incorporating ACA in the model results in a significant performance improvement, with the model achieving 86.57% mAP. This indicates that ACA is effective in reducing information loss during high-level feature map generation, thereby enhancing the accuracy of model detection. As shown in Table V, with the participation of ACA, the model reaches

TABLE IX

DETECTION PERFORMANCE OF DIFFERENT METHODS ON HRSC2016. MAP(07) AND MAP(12) REFER TO THE MAP COMPUTED ON THE PASCAL VOC2007 AND PASCAL VOC2012, RESPECTIVELY. * MEANS THAT THE DATASET USES THE PREPROCESSING MODULE BCDM. BOLD HIGHLIGHTS THE BEST RESULTS

Method	Backbone	mAP(07)(%)	mAP(12)(%)
Two-stage methods			
R^{2}CNN [15]	ResNet101	73.07	79.73
ROI Transformer [16]	ResNet101	86.20	
Gliding Vert [18]	ResNet101	88.20	-
Oriented R-CNN [26]	ResNet101	90.50	97.60
DEA-Net [33]	ResNet101	90.56	-
DEA-Net* [33]	ResNet101	90.62	-
SCRDet++ [19]	ResNet101	-	97.67
SCRDet++* [19]	ResNet101	-	97.75
One-stage methods			
CSL [20]	ResNet101	89.62	96.10
R^{3}Det [22]	ResNet101	89.26	96.01
RSDet [23]	ResNet152	86.50	-
R^{3}Det-DCL [21]	ResNet101	89.46	96.41
R^{3}Det-DCL* [21]	ResNet101	89.57	96.53
Anchor-free methods			
BBAVectors [24]	ResNet101	88.60	-
Oriented RepPoint [30]	ResNet50	90.38	97.26
CHPDet [27]	DLA34	88.81	-
GF-CSL [34]	ResNet101	90.53	97.90
GF-CSL* [34]	ResNet101	90.61	**97.97**
S^{2}A-Net [28]	ResNet101	90.17	95.01
S^{2}A-Net* [28]	ResNet101	90.28	95.30
3WM-AugNet (Ours)	ResNet101	90.60	96.94
3WM-AugNet (Ours)*	ResNet101	**90.69**	97.02

TABLE X

DETECTION PERFORMANCE OF DIFFERENT METHODS ON UCAS-AOD. MAP(07) AND MAP(12) REFER TO THE MAP COMPUTED ON THE PASCAL VOC2007 AND PASCAL VOC2012, RESPECTIVELY. * MEANS THAT THE DATASET USES THE PREPROCESSING MODULE BCDM. BOLD HIGHLIGHTS THE BEST RESULTS

Method	Backbone	Car	Plane	mAP(07)(%)	mAP(12)(%)
Two-stage methods					
ROI Transformer [16]	ResNet101	87.99	89.90	88.95	-
DEA-Net [33]	ResNet101	88.12	90.38	89.25	-
DEA-Net* [33]	ResNet101	88.34	90.41	89.38	-
SCRDet++ [19]	ResNet101	94.97	98.93	-	96.95
SCRDet++* [19]	ResNet101	95.09	98.97	-	97.03
One-stage methods					
YOLOv3 [4]	Darknet53	74.63	89.52	82.08	-
RetinaNet [5]	ResNet101	84.64	90.51	87.50	-
		93.61	97.30	-	95.46
CSL [20]	ResNet101	88.09	90.38	89.23	-
R^{3}Det-DCL [21]	ResNet101	88.15	90.57	89.36	-
DAL [25]	ResNet101	89.25	90.49	89.87	-
DAL* [25]	ResNet101	89.47	90.54	90.01	-
R^{3}Det [22]	ResNet101	94.14	98.20	-	96.17
R^{3}Det* [22]	ResNet101	94.25	98.26	-	96.26
Anchor-free methods					
CHPDet [27]	DAL34	88.58	90.64	89.61	-
Oriented RepPoint [30]	ResNet101	89.51	90.70	90.11	-
Oriented RepPoint* [30]	ResNet101	89.69	90.77	90.23	-
GF-CSL [34]	ResNet101	88.39	90.60	89.49	-
		93.05	98.53	-	95.79
GF-CSL* [34]	ResNet101	88.56	90.68	89.62	-
		93.22	98.55	-	95.89
S^{2}A-Net [28]	ResNet101	89.50	90.40	89.90	-
S^{2}A-Net* [28]	ResNet101	89.61	90.45	90.03	-
3WM-AugNet (Ours)	ResNet101	90.02	90.73	90.38	-
		95.31	98.93	-	97.12
3WM-AugNet (Ours)*	ResNet101	90.16	90.83	**90.50**	-
		95.43	98.98	-	**97.21**

86.57% mAP, indicating that ACA can effectively reduce information loss when generating high-level feature maps, thereby improving the accuracy of model detection. In addition, we also find that incorporating BF^{2}L can also achieve 82.73% mAP, indicating that feature fusion can fully use accurate positioning information of low level to improve high-level feature learning. Therefore, by combining ACA and BF^{2}L, we achieve a detection accuracy of 89.45% mAP. Compared with the baseline, this combination results in a 9.26% increase in the mAP of the model while only increasing 0.77M parameters and 8.41 GFLOPs. Therefore, ACA and BF^{2}L have been proven to be effective in improving the performance of the model.

E. Comparison With State-of-the-Art

This section compares our 3WM-AugNet with other state-of-the-art methods on the three RS datasets, i.e., FGSD2021, HRSC2016, and UCAS-AOD. To ensure a fair comparison, we do not use the preprocessing module BCDM for all algorithms. Meanwhile, to evaluate the performance of our 3WM-AugNet more comprehensively, we also select several excellent algorithms that use our proposed BCDM in the preprocessing. This step eliminates the influence of the preprocessing module BCDM on the experimental results and ensures a fair evaluation of the performance of our 3WM-AugNet.

1) Results on FGSD2021: As shown in Table VII, we can see that, when all algorithms did not use our proposed preprocessing module BCDM, our 3WM-AugNet improved the mAP from 80.19% to 89.45% compared to the baseline S^2A-Net [28]. Our 3WM-AugNet outperforms two-stage algorithms in terms of performance, and this advantage is even more pronounced compared to one-stage anchor-free methods. In particular, the anchor-free method GF-CSL [34] achieves 88.49% mAP at a speed of 40.32 FPS, while our 3WM-AugNet outperforms GF-CSL by 0.96% mAP, with the highest accuracy of 89.45% mAP and a speed of only slightly lower than GF-CSL. This is attributed to MFAM enhancing the interaction between low- and high-level features, preserving critical high-level features, and significantly improving the detection performance for ships of different sizes through adaptive feature fusion.

In addition, Table VII shows that, compared to other state-of-the-art methods using the same preprocessing module BCDM, our 3WM-AugNet achieved the highest detection accuracy, reaching 91.53% mAP at 26.49 FPS. Compared with SCRDet++ [19], RSDet [23], GF-CSL, and S^2A-Net, our 3WM-AugNet improves the detection accuracy by 1.49%, 15.78%, 0.93%, and 8.45%, respectively. Meanwhile, our 3WM-AugNet achieves the best detection accuracy in Was, Tar, Aus, Whi, Tic, Bur, Per, Lew, Mer, Ind, Sub, and Oth detection categories. Furthermore, our 3WM-AugNet achieves 94.21% mAP when using multiscale training and testing, which is 1.14% higher than GF-CSL and obtains the best performance. The reason is that BCDM performs effective image deblurring on RS images. It can accurately identify and process blurry images while avoiding unnecessary processing of clear images, thereby improving the quality and clarity of the images. This further contributes to enhancing the detection performance of ships of different sizes in RS images.

Given that our 3WM-AugNet incorporates the preprocessing module BCDM, we further compare its computation and number of parameters with several excellent methods to highlight the superiority of our 3WM-AugNet. The experimental results are shown in Table VIII. According to Table VIII, it can be observed that our 3WM-AugNet utilizes the BCDM preprocessing module, which introduces only a few parameters and computations, while significantly improving the accuracy of ship detection.

2) Results on HRSC2016: We evaluate our 3WM-AugNet on PASCAL VOC2007 and PASCAL VOC2012 metrics. Under the VOC2007 metric, we evaluate R^2CNN [15], ROI Transformer [16], Gliding Vert [18], Oriented R-CNN [26], DEA-Net [33], CSL [20], R^3Det [22], RSDet [23], R^3Det-DCL [21], BBAVectors [24], Oriented RepPoint [30], CHPDet [27], GF-CSL [34], and S^2A-Net [28] methods. Under the VOC2012 index, we evaluate R^2CNN, Oriented R-CNN, SCRDet++ [19], CSL, R^3Det, R^3Det-DCL, Oriented RepPoint, GF-CSL, and S^2A-Net methods. The experimental results are shown in Table IX. When all algorithms do not use our proposed preprocessing module BCDM, our 3WM-AugNet achieves 90.60% mAP and 96.94% mAP under the VOC2007 and VOC2012 indicators, demonstrating the effectiveness of the network design of 3WM-AugNet. This achievement is attributed to our integration of the multigranularity concept into the global and local coordinate attention mechanisms. Moreover, compared with several state-of-the-art methods using the same preprocessing module BCDM,

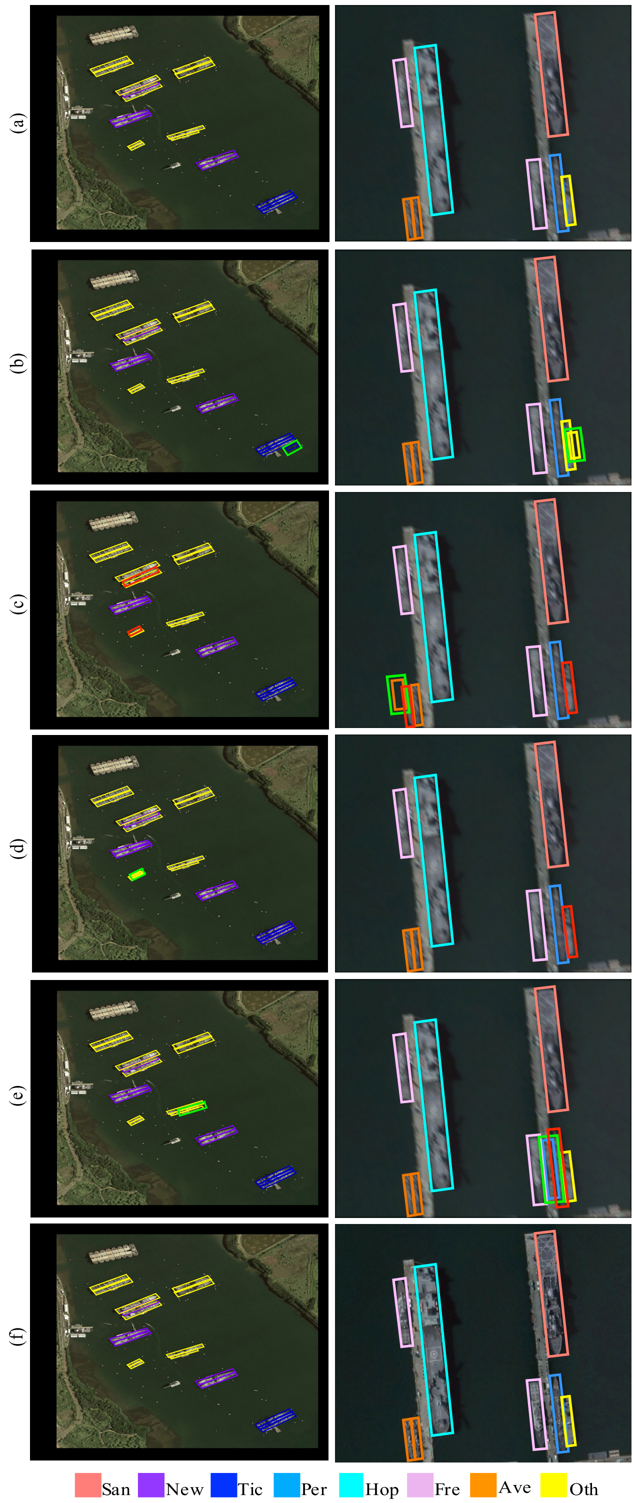


Fig. 10. Comparison of detection visualization results of different methods on FGSD2021. Each row represents the detection results of (a) ground truth, (b) SCRDet++, (c) RSDet, (d) GF-CSL, (e) S^2A-Net, and (f) proposed 3WM-AugNet. Different colored rotating boxes represent different types of ship targets. Red and green boxes indicate missed and false detections, respectively.

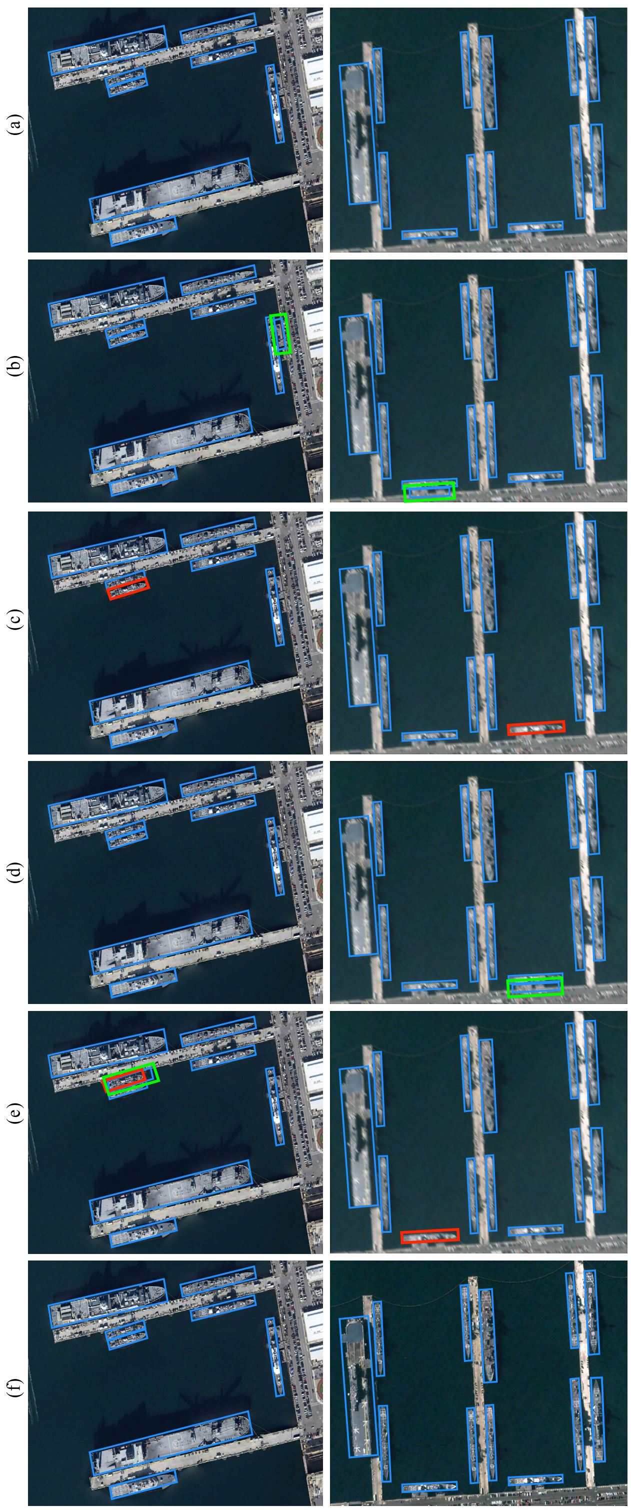


Fig. 11. Comparison of detection visualization results of different methods on HRSC2016. Each row represents the detection results of (a) ground truth, (b) DEA-Net, (c) R^3Det-DCL, (d) GF-CSL, (e) S^2A-Net, and (f) proposed 3WM-AugNet. Light blue boxes represent detected ship targets. Red and green boxes indicate missed and false detections, respectively.

Fig. 12. Comparison of detection visualization results of different methods on UCAS-AOD. Each row represents the detection results of (a) ground truth, (b) DEA-Net, (c) Oriented RepPoint, (d) GF-CSL, (e) S^2A-Net, and (f) proposed 3WM-AugNet. Green and purple boxes represent detected cars and planes, respectively. Red and green boxes indicate missed and false detections, respectively.

3WM-AugNet achieves the best performance on the VOC2007 metric, achieving 90.69% mAP, demonstrating its superior performance. However, our 3WM-AugNet fails to achieve state-of-the-art performance under the PASCAL VOC2012 metric. Since our 3WM-AugNet mainly focuses on image blur and ship targets of different sizes, it does not fully consider other complex and diverse scene factors, such as occlusion and

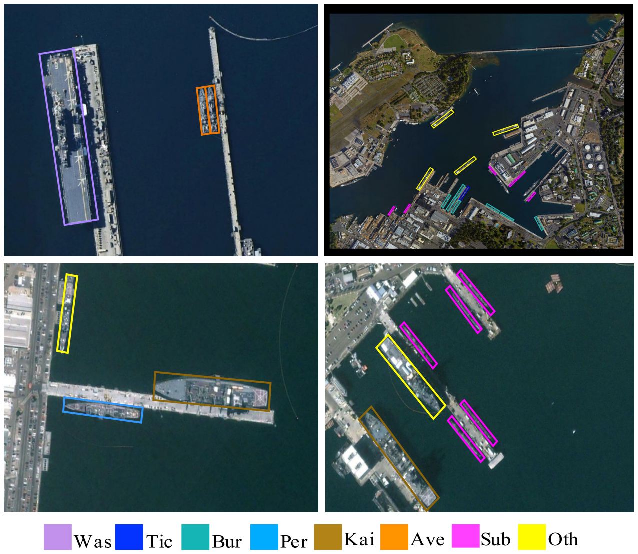


Fig. 13. Some examples of detection results of the proposed 3WM-AugNet on FGSD2021. Different color rotating boxes represent different types of ship targets.

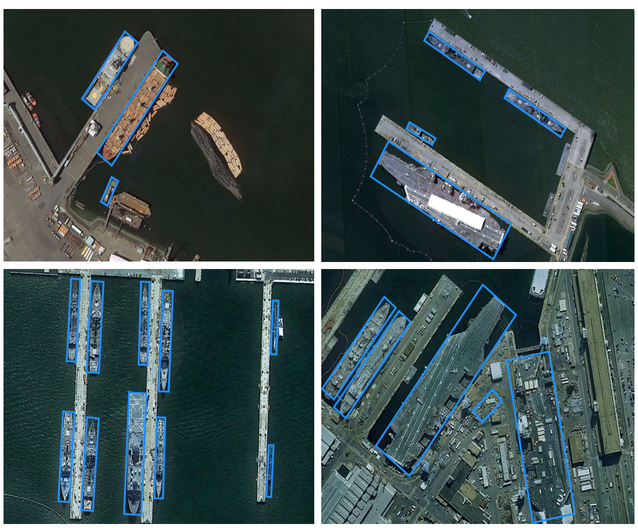


Fig. 14. Some examples of detection results of the proposed 3WM-AugNet on HRSC2016. Light blue boxes represent detected ship targets.

illumination changes, resulting in poor performance under the PASCAL VOC2012 metric.

3) Results on UCAS-AOD: To better evaluate the generalization performance and practical value of the proposed 3WM-AugNet, we also choose an RS dataset, UCAS-AOD, containing high-resolution images of airplanes and cars for detection. In addition, since the image scenes in the UCAS-AOD are similar to those in ship detection, this experiment can comprehensively evaluate the performance of the algorithm in RS ship detection.

We choose PASCAL VOC2007 and PASCAL VOC2012 as evaluation metrics to comprehensively evaluate the performance of different methods. Table X shows the detection results of different methods on UCAS-AOD. Remarkably, our 3WM-AugNet demonstrates outstanding performance on UCAS-AOD even without using the preprocessing module BCDM. It achieves state-of-the-art performance in PASCAL VOC2007 and PASCAL VOC2012 evaluation metrics, achieving 90.38% mAP and 97.12% mAP, respectively. In addition, compared with several excellent and popular methods employing the same preprocessing module BCDM, our 3WM-AugNet achieves state-of-the-art performance on both metrics. Specifically, its mAP on the two indicators of PASCAL VOC2007 and PASCAL VOC2012 reached 90.50% and 97.21%, respectively. This further verifies the outstanding performance of 3WM-AugNet in terms of generalization ability. The reason is that our 3WM-AugNet uses image preprocessing operations and makes full use of the multigranularity attention mechanism when dealing with image blur and ships of varying sizes in the UCAS-AOD dataset, resulting in significant performance improvements.

Fig. 15. Some examples of detection results of the proposed 3WM-AugNet on UCAS-AOD. Green and purple boxes represent detected cars and planes, respectively.

F. Visualizing Results and Insight

We visualize the detection results on FGSD2021, HRSC2016, and UCAS-AOD, as shown in Figs. 10, 11, and 12, respectively. Furthermore, we also give some examples of the proposed 3WM-AugNet to further verify its accuracy and feasibility, as shown in Figs. 13, 14, and 15, respectively.

1) Visualization Results of FGSD2021: To facilitate the comparison of RS ship detection performance among different algorithms, we visualize the results of several excellent algorithms and compare their missed and false detection rates to highlight the strengths of our proposed 3WM-AugNet. Fig. 10 shows the visual comparison of detection results on FGSD2021 for ground truth, SCRDet++ [19], RSDet [23], GF-CSL [34], S^2A-Net [28], and 3WM-AugNet.

In Fig. 10, the first column displays clear images, and all compared detectors exhibit a certain degree of missed or false detection. Specifically, RSDet has the highest missed detection rate, missing one Oth, and one New. GF-CSL and S^2A-Net both have a false detection of one Oth. SCRDet++ has a false detection of one Tic. However, our 3WM-AugNet can accurately detect each type of ship with low missed detection and false detection rates. The second column displays blurry images. Due to motion blur, the compared detectors exhibit a certain degree of missed and false detections for ships of various sizes, especially for small and medium-sized ships, such as Ave, Oth, and Per. SCRDet++ has one false detection of Oth. RSDet misses one Oth and one Ave while falsely detecting one Ave. GF-CSL misses one Oth. S^2A-Net misses one Per and falsely detects one Per. Compared to other detectors, our 3WM-AugNet demonstrates significantly better performance and can accurately detect ships of varying sizes, even under blur interference.

Fig. 13 shows some detection results of 3WM-AugNet on FGSD2021. The first row displays clear images, while the second displays blurry images caused by motion blur interference. It should be noted that the first and second rows show the detection results of FGSD2021 by 3WM-AugNet added to the preprocessing module BCDM. The results indicate that 3WM-AugNet can effectively detect ship targets of different sizes in scenes with motion blur interference, further demonstrating the high accuracy and feasibility of 3WM-AugNet.

2) Visualization Results of HRSC2016: The visual comparison results of ground truth, DEA-Net [33], R^3Det-DCL [21], GF-CSL [34], S^2A-Net [28], and 3WM-AugNet on HRSC2016 are shown in Fig. 11. The first column displays clear images, where R^3Det-DCL and S^2A-Net miss one ship each, while DEA-Net and S^2A-Net falsely detect one ship each. Both GF-CSL and 3WM-AugNet accurately detect all ship targets. The second column displays blurry images, where DEA-Net and GF-CSL falsely detect one ship each, while R^3Det-DCL and S^2A-Net miss one ship each. In contrast, our 3WM-AugNet can effectively detect ship targets of different sizes in blurry scenes. In addition, in Fig. 14, we present some detection results of 3WM-AugNet on HRSC2016, showcasing its superiority in detecting ship targets of different sizes in clear and blurry scenes.

3) Visualization Results of UCAS-AOD: To comprehensively evaluate the performance and generalization ability of 3WM-AugNet, we present the detection results compared with other algorithms in Fig. 12 and visualize some examples of detections in Fig. 15. This approach can not only comprehensively evaluate the performance of the model in different datasets and scenarios but also demonstrates its generalization ability and applicability.

In Fig. 12, the first column shows clear images, and the second shows blurry images. All detectors have some degree of missed detection and false detection. In the case of blurry images, missed and false detection is more severe. DEA-Net [33] has the highest missed and false detection rates, while Oriented RepPoint [30], GF-CSL [34], and S^2A-Net [28] also have some degree of missed and false detection. In contrast, our 3WM-AugNet has extremely low missed and false detection rates, and can correctly detect all cars and planes in clear and blurry scenes. In addition, Fig. 15 shows some detection results of 3WM-AugNet in different scenes of UCAS-AOD, including clear scenes (first row) and blurry scenes (second row). These results demonstrate the robustness and generalization ability of our model, which can perform well in different scenarios and datasets.

V. Conclusion

This article proposes a novel one-stage RS ship detection method 3WM-AugNet. To achieve the goal of only deblurring blurry images without reducing the clarity of clear images, we design BCDM based on 3WD, resulting in high-quality images. In addition, to enhance the interactions of low- and high-level features, we incorporate multigranularity feature learning in coordinate attention. Specifically, an ACA is designed to reduce the loss of high-level features. Meanwhile, BF^2L is added to enhance the influence of bottom-level features on the overall features. Extensive ablation experiments show that the proposed 3WM-AugNet not only effectively improves the baseline performance but also achieves state-of-the-art performance on FGSD2021, HRSC2016, and UCAS-AOD. However, despite BCDM and MFAM significantly improving model performance by implementing data augmentation and enhancing multigranularity feature representation, they also introduce additional parameters, thereby reducing the detection speed of the model to some extent. In future work, we plan to explore more advanced blur classification methods, aiming to eliminate complex preprocessing procedures and construct a more elegant, concise, and efficient end-to-end RS ship detection network, to better balance the performance and speed of the model.

References

[1] Z. Cui, J. Leng, Y. Liu, T. Zhang, P. Quan, and W. Zhao, "SKNet: Detecting rotated ships as keypoints in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8826–8840, Oct. 2021.

[2] Y. Yu, X. Yang, J. Li, and X. Gao, "A cascade rotated anchor-aided detector for ship detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5600514.

[3] Y. Zhuang, Y. Liu, T. Zhang, and H. Chen, "Contour modeling arbitrary-oriented ship detection from very high-resolution optical remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[4] K. Liu, H. Zhang, J. Xia, J. Gao, and Y. Liu, "A coarse-to-fine network for ship detection in optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 780–784, Jan. 2019.

[5] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery," *Remote Sens.*, vol. 11, no. 5, pp. 531–544, 2019.

[6] Y. Yao, "Three-way decisions with probabilistic rough sets," *Inf. Sci.*, vol. 180, no. 3, pp. 341–353, Feb. 2010.

[7] J. Zhou, Z. Lai, D. Miao, C. Gao, and X. Yue, "Multigranulation rough-fuzzy clustering based on shadowed sets," *Inf. Sci.*, vol. 507, pp. 553–573, Jan. 2020.

[8] S. Hu, D. Miao, and W. Pedrycz, "Multi granularity based label propagation with active learning for semi-supervised classification," *Expert Syst. Appl.*, vol. 192, Apr. 2022, Art. no. 116276.

[9] F. Xu, J. Liu, C. Dong, and X. Wang, "Ship detection in optical remote sensing images based on wavelet transform and multi-level false alarm identification," *Remote Sens.*, vol. 9, no. 10, pp. 985–1003, 2017.

[10] H. He, Y. Lin, F. Chen, H.-M. Tai, and Z. Yin, "Inshore ship detection in remote sensing images via weighted pose voting," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3091–3107, Jun. 2017.

[11] W. Rao, L. Gao, Y. Qu, X. Sun, B. Zhang, and J. Chanussot, "Siamese transformer network for hyperspectral image target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5526419.

[12] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS^3LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504218.

[13] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500117.

[14] X. Yang, X. Zhang, N. Wang, and X. Gao, "A robust one-stage detector for multiscale ship detection with complex background in massive SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5217712.

[15] Y. Jiang et al., "R^2CNN: Rotational region CNN for orientation robust scene text detection," 2017, *arXiv:1706.09579*.

[16] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning RoI transformer for oriented object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2849–2858.

[17] X. Yang et al., "SCRDet: Towards more robust detection for small, cluttered and rotated objects," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8232–8241.

[18] Y. Xu et al., "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1452–1459, Apr. 2021.

[19] X. Yang, J. Yan, W. Liao, X. Yang, J. Tang, and T. He, "SCRDet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2384–2399, Feb. 2023.

[20] X. Yang and J. Yan, "Arbitrary-oriented object detection with circular smooth label," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 677–694.

[21] X. Yang, L. Hou, Y. Zhou, W. Wang, and J. Yan, "Dense label encoding for boundary discontinuity free rotation detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15819–15829.

[22] X. Yang, J. Yan, Z. Feng, and T. He, "R^3Det: Refined single-stage detector with feature refinement for rotating object," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 4, pp. 3163–3171.

[23] W. Qian, X. Yang, S. Peng, J. Yan, and Y. Guo, "Learning modulated loss for rotated object detection," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 3, pp. 2458–2466.

[24] J. Yi, P. Wu, B. Liu, Q. Huang, H. Qu, and D. Metaxas, "Oriented object detection in aerial images with box boundary-aware vectors," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Jul. 2021, pp. 2150–2159.

[25] Q. Ming, Z. Zhou, L. Miao, H. Zhang, and L. Li, "Dynamic anchor learning for arbitrary-oriented object detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 3, pp. 2355–2363.

[26] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented R-CNN for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3520–3529.

[27] F. Zhang, X. Wang, S. Zhou, Y. Wang, and Y. Hou, "Arbitrary-oriented ship detection through center-head point extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5612414.

[28] J. Han, J. Ding, J. Li, and G.-S. Xia, "Align deep features for oriented object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5602511.

[29] T. Zhang et al., "FFN: Fountain fusion net for arbitrary-oriented object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5609913.

[30] W. Li, Y. Chen, K. Hu, and J. Zhu, "Oriented reppoints for aerial object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1829–1838.

[31] C. Zhang, B. Xiong, X. Li, and G. Kuang, "TCD: Task-collaborated detector for oriented objects in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4700714.

[32] Y. Li, Q. Hou, Z. Zheng, M.-M. Cheng, J. Yang, and X. Li, "Large selective kernel network for remote sensing object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Mar. 2023, pp. 1–16.

[33] D. Liang et al., "Anchor retouching via model interaction for robust object detection in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5619213.

[34] J. Wang, F. Li, and H. Bi, "Gaussian focal loss: Learning distribution polarized angle prediction for rotated object detection in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4707013.

[35] 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.

[36] Y. Liu et al., "Motion consistency-based correspondence growing for remote sensing image matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[37] Y. Liu et al., "Robust feature matching via advanced neighborhood topology consensus," *Neurocomputing*, vol. 421, pp. 273–284, Jan. 2021.

[38] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 70–79, 2007.

[39] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors, supplementary material," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2009, pp. 1033–1041.

[40] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M.-H. Yang, "Image deblurring via enhanced low-rank prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3426–3437, Jul. 2016.

[41] L. Xu, S. Zheng, and J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1107–1114.

[42] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5971–5979.

[43] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14816–14826.

[44] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 8174–8182.

[45] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Jun. 2019, pp. 8878–8887.

[46] D. Park, D. U. Kang, J. Kim, and S. Y. Chun, "Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 327–343.

[47] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Jun. 2021, pp. 4641–4650.

[48] S.-W. Ji et al., "XYDeblur: Divide and conquer for single image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17421–17430.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–14.

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

[52] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "NAS-FPN: Learning scalable feature pyramid architecture for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7029–7038.

[53] G. Zhao, W. Ge, and Y. Yu, "GraphFPN: Graph feature pyramid network for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2743–2752.

[54] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 8759–8768.

[55] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.

[56] S. Huang, Z. Lu, R. Cheng, and C. He, "FaPN: Feature-aligned pyramid network for dense image prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 844–853.

[57] Z. Jin, D. Yu, L. Song, Z. Yuan, and L. Yu, "You should look at all objects," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 332–349.

[58] Y. Zhou, Q. Ye, Q. Qiu, and J. Jiao, "Oriented response networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 519–528.

[59] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

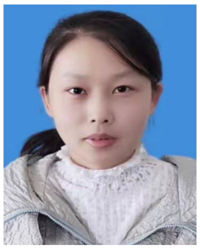
[60] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.

[61] Z. Liu, L. Yuan, L. Weng, and Y. Yang, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *Proc. 6th Int. Conf. Pattern Recognit. Appl. Methods*, 2017, pp. 324–331.

[62] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, "Orientation robust object detection in aerial images using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3735–3739.

[63] K. Chen et al., "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*.

[64] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.



Li Ying received the M.S. degree in applied mathematics from the Anhui University of Science and Technology, Huainan, China, in 2019. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology, Tongji University, Shanghai, China.

Her research interests include computer vision, deep learning, and object detection, focusing on ship detection.

Duoqian Miao is currently a Professor and a Ph.D. Tutor with the College of Electronics and Information Engineering, Tongji University, Shanghai, China. He has published more than 200 papers in IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Pattern Recognition*, *Information Sciences*, and *Knowledge-Based Systems*. His interests include machine learning, data mining, big data analysis, granular computing, artificial intelligence, and text image processing.

Dr. Miao's representative awards include the Second Prize of Wuwenjun AI Science and Technology in 2018, the First Prize of Natural Science of Chongqing in 2010, the First Prize of Technical Invention of Shanghai in 2009, and the First Prize of the Ministry of Education Science and Technology Progress Award in 2007. He serves as the Honorary Chair of the CAAI Granular Computing Knowledge Discovery Technical Committee, the Vice-Director for MOE Key Laboratory of Embedded System & Service Computing, and the Vice-President of the Shanghai Association for Artificial Intelligence. He serves as an Associate Editor for *International Journal of Approximate Reasoning* and *Information Sciences* and an Editor of the *Journal of Computer Research and Development* (in Chinese).

Zhifei Zhang (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent systems from Tongji University, Shanghai, China, in 2014.

From 2014 to 2015, he was a Post-Doctoral Fellow with the University of Montreal, Montreal, QC, Canada. He is currently an Associate Professor with Tongji University. He has authored over 30 scientific papers in international journals and conferences. His major research interests include pattern recognition and big data analysis.

Dr. Zhang won the Second Prize of the Wu Wenjun AI Science and Technology Award in 2018 and the Second Prize of the Shanghai Science and Technology Progress Award in 2022. He serves as the Secretary-General of the Computer Vision Committee, Shanghai Computer Federation.