Received: 15 June 2023 | Revised: 6 September 2023 | Accepted: 27 September 2023

CAAI Transactions on Intelligence Technology

DOI: 10.1049/cit2.12310

ORIGINAL RESEARCH

Multi-granularity feature enhancement network for maritime ship detection

Li Ying[1,2] | Duoqian Miao[1] | Zhifei Zhang[1,2] | Hongyun Zhang[1] | Witold Pedrycz[3,4]

[1]Department of Computer Science and Technology, Tongji University, Shanghai, China

[2]Project Management Office of China National Scientific Seafloor Observatory, Tongji University, Shanghai, China

[3]Department of Electrical and Computer Engineering, Alberta, Edmonton, Canada

[4]System Research Institute, Polish Academy of Sciences, Warsaw, Poland

Correspondence

Duoqian Miao and Zhifei Zhang.
Email: dqmiao@tongji.edu.cn and zhifeizhang@tongji.edu.cn

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2022YFB3104700; National Natural Science Foundation of China, Grant/Award Numbers: 62376198, 61906137, 62076040, 62076182, 62163016, 62006172; The China National Scientific Sea-floor Observatory, The Natural Science Foundation of Shanghai, Grant/Award Number: 22ZR1466700; The Jiangxi Provincial Natural Science Fund, Grant/Award Number: 20212ACB202001

Abstract

Due to the characteristics of high resolution and rich texture information, visible light images are widely used for maritime ship detection. However, these images are susceptible to sea fog and ships of different sizes, which can result in missed detections and false alarms, ultimately resulting in lower detection accuracy. To address these issues, a novel multi-granularity feature enhancement network, MFENet, which includes a three-way dehazing module (3WDM) and a multi-granularity feature enhancement module (MFEM) is proposed. The 3WDM eliminates sea fog interference by using an image clarity automatic classification algorithm based on three-way decisions and FFA-Net to obtain clear image samples. Additionally, the MFEM improves the accuracy of detecting ships of different sizes by utilising an improved super-resolution reconstruction convolutional neural network to enhance the resolution and semantic representation capability of the feature maps from YOLOv7. Experimental results demonstrate that MFENet surpasses the other 15 competing models in terms of the mean Average Precision metric on two benchmark datasets, achieving 96.28% on the McShips dataset and 97.71% on the SeaShips dataset.

KEYWORDS

object classification, object recognition, rough sets, rough set theory

1 | INTRODUCTION

With the rapid development of the economy, China's demand for inland river and ocean shipping is increasing. How to ensure the safety of shipping has become one of the important research pursuits in the field of modern shipping research. Ship detection can not only significantly reduce the probability of collision accidents during navigation, ensure the safety of navigation and transportation, but also help detect instances of illegal entry and illegal operations in time. At present, Tongji University is taking the lead in coordinating the construction of the major national scientific and technological infrastructure of the Seafloor Observatory Network. Intensive fishing activities in the East China Sea have posed a significant threat to the safe operation of the observation network, necessitating real-time monitoring of surface ships.

In the field of maritime ship detection against ocean backgrounds, various detection methods have been proposed in the literature. Four types of images have been explored for this purpose: (1) satellite remote sensing imagery, which has low resolution and cannot correctly identify illegal activities in real time [1]; (2) synthetic aperture radar images, which lack rich spectral information on objects and are not convenient for subsequent object segmentation or tracking [2–4]; (3) infrared

imaging, which suffers from low resolution and is easily affected by inhomogeneous noise, leading to reduced detection efficiency [5, 6]; and (4) visible light images, which are widely used due to their high resolution and rich texture information [7]. Therefore, the use of visible light images for ship detection has become one of the important research directions of our research on maritime ship detection. However, the issues of sea fog and diverse ship sizes will seriously affect the accuracy of ship detection in visible light images. Therefore, how to achieve accurate ship detection in visible light images has become a challenging and timely research topic.

Ship detection in visible light images is a typical object detection task that can be accomplished using traditional image processing or deep learning methods. Traditional ship detection algorithms rely mainly on manually designed feature extractors to solve image tasks in specific scenarios, with poor generalisation ability and robustness. In contrast, deep learning-based ship detection algorithms do not require manual feature design, can adapt to detection tasks in various complex environments, have strong generalisation and robustness, and have fast detection speed. The deep learning-based methods can be classified into two categories: (1) two-stage algorithms, including Faster R-CNN [8], GWFEF-Net [9], and MIDN [10]. (2) single-stage algorithms, such as YOLOv5 [11], HyperLi-Net [12], CBNet [13], and FCOS [14].

Due to the influence of sea fog on the clarity of visible light ship images, image preprocessing is required before conducting ship image detection, wherein blurry images are selected through image clarity classification. Subsequently, deep learning methods are used to detect all preprocessed images. Nevertheless, all existing dehazing algorithms apply direct dehazing processing to all visible light images, including those without sea fog. This practice may result in excessively blurry clear images, reducing clarity. Therefore, using the three-way decisions (3WD) [15] for image clarity classification can lead to better dehazing processing and further improve the accuracy of ship detection. The 3WD is a classical granular computing model that can be used for image clarity classification. The model reflects the way humans think to solve problems. First, the acceptance and rejection domains are determined. Then, the uncertainty domain is studied. Finally, objects within the uncertainty domain are transformed into the acceptance and rejection domains, thereby reducing their uncertainty [16, 17]. Therefore, using the 3WD can achieve better image clarity classification.

Meanwhile, when using deep learning methods to detect a large amount of image data, single-granularity features can only extract image feature information at specific angles. Therefore, it becomes necessary to consider multi-granularity features, fully utilise the interrelationships between different granularity, and achieve multi-granularity information fusion to obtain better feature representations [18–20], which is beneficial to improving the accuracy of ship detection.

This paper proposes a network called MFENet, which effectively improves the accuracy of multi-sized ship detection under sea fog interference. First, an automatic classification algorithm for image clarity based on three-way decisions is designed to obtain blurry images with sea fog and clear images without sea fog. The FFA-Net [21] is then introduced to process the blurry images, eliminating the influence of sea fog on the clarity of visible images. This operation is conducive to subsequent ship detection. Next, a multi-granularity feature enhancement module (MFEM) is designed to perform super-resolution reconstruction on the three feature maps extracted by YOLOv7 [22] and enhance the semantic information of the feature maps. This refinement aims to reduce missed detections and false alarms, further elevating the accuracy of ship detection at sea. In short, our main contributions are as follows:

1) A 3WDM is proposed to reduce the interference of sea fog on ship detection in visible light images. 3WDM can efficiently classify blurred images with sea fog and perform dehazing processing to avoid excessive blurring of clear images without fog. This further enhances the quality of data samples and greatly promotes subsequent feature extraction.
2) An MFEM is designed to enhance the feature representation capability of multi-sized ship detection. MFEM employs the improved super-resolution reconstruction convolutional neural network (SRCNN) to conduct super-resolution reconstruction on the three feature maps extracted by the Head of YOLOv7, aiming to enhance the details and clarity of these feature maps. This strategy enriches the feature representation of multi-sized ship detection.
3) To verify the effectiveness of the proposed MFENet, we conduct extensive experiments on the McShips and SeaShips datasets. The mean Average Precision (mAP) can reach 96.28% and 97.71%, respectively, achieving state-of-the-art detection performance.

The remainder of this paper is organised as follows. Section 2 provides a review of related work. Section 3 presents a detailed description of the proposed MFENet. Section 4 discusses experimental results. Section 5 draws some conclusions and potential future work.

2 | RELATED WORK

This subsection mainly introduces the ship detection method, the image dehazing method, and the network structure of SRCNN.

2.1 | Ship detection method

Early maritime ship detection methods such as HOG [23], SIFT [24], and LBP [25] typically use handcrafted features combined with traditional classifiers. However, these methods are not only slow in detection speed but also poor in detection effects. In recent years, technologies based on convolutional neural networks (CNNs) [26–28] gradually replaced traditional methods and have become a research hotspot. Some CNN-

based maritime ship detection methods use single-stage detectors, such as Deformable DETR [29], CFP [30], and PDNet [31], which can achieve fast and accurate detection. However, these methods have limitations in detecting small targets. In addition, some methods use two-stage detectors, such as Quad-FPN [32], BUAA-PAL-OICR [33], and SCSD [34], which are more effective for small objects, but the detection speed is slower.

To solve the above problems, some improved methods for maritime ship detection have been proposed. These methods incorporate attention mechanisms, feature pyramid network (FPN), multi-granularity feature fusion, and other techniques to enhance detection accuracy and efficiency. These approaches are continuously emerging, introducing novel ideas and methodologies to the field of marine ship detection research and application. Cui et al. [35] proposed DAPN, which achieves efficient and accurate detection of multi-sized ships by employing a multi-granularity feature pyramid and an attention mechanism. Zhang et al. [36] designed MDCN, which can share similar feature representations among domains with different granularities, thus enabling more generalised ship detection. Li et al. [37] designed HSF-Net, which employs a multi-granularity deep feature embedding method to capture image features of different granularities, enhancing ship detection. Zhang et al. [38] proposed BL-Net, which combines synthetic data and real data, and introduces an attention mechanism to solve the data imbalance problem, thereby improving the performance of ship detection. Although these methods can enhance the detection performance of ships of different sizes to some extent, they do not fully consider the sea fog problem existing in ship detection. Given the unique characteristics of visible light ship data, including factors such as sea fog and multi-sized ship targets, this paper integrates 3WDM and MFEM into YOLOv7. This integration aims to enhance detection performance while maintaining high-speed detection, achieving a better balance between speed and accuracy.

2.2 | Image dehazing method

Currently, traditional image processing dehazing algorithms [39–41] mainly rely on enhancing contrast to improve the visual effect of images or modelling haze images based on atmospheric scattering laws and image degradation reasons to achieve dehazing restoration. However, these algorithms are computationally complex and are not completely effective for dehazing different hazy scenes. In recent years, CNN has been rapidly developed in various fields, benefiting from its powerful learning ability, and has been used to process hazy images. CNN algorithms combine the characteristics of the haze itself and the reasons for dehazing, making them have better performance in dehazing. Li et al. [42] designed AOD-Net, which appropriately transformed the atmospheric scattering model formula and learnt its related parameters through a neural network. Zhang et al. [43] proposed DCPDN, which can automatically learn image dehazing and adopts a pyramid-shaped dense connection structure to improve the effectiveness and speed of dehazing. Shao et al. [44] proposed a domain adaptation method based on generative adversarial networks, which improves dehazing performance by learning the mapping from a labelled source domain to an unlabelled target domain. Liu et al. proposed GridDehazeNet [45], which utilises an attention mechanism to effectively remove haze from images and has strong robustness and generalisation performance in complex scenes. Qin et al. designed FFA-Net [21], which removes haze from single images through feature fusion and attention mechanism and has high dehazing quality and real-time performance.

The above dehazing algorithm usually performs dehazing processing on all images, which may cause some images without fog or uncertain whether there is fog to become blurry, which is not conducive to the extraction of image features in subsequent target detection. Therefore, in this study, we design a 3WDM to alleviate the problem of excessive image dehazing.

2.3 | The structure of SRCNN

The SRCNN [46] is a popular method for solving the problem of image super-resolution. SRCNN mainly uses a three-layer CNN to fit the non-linear mapping between low-resolution and high-resolution images and obtain the network output, which is the reconstructed high-resolution image. The structure of SRCNN is shown in Figure 1, where the three convolutional operations are used for feature block extraction and representation, non-linear mapping, and reconstruction, respectively. The first convolutional layer uses kernels of size $f_1 \times f_1$, with n_1 kernels and n_1 feature maps as outputs. The second convolutional layer uses kernels of size $f_2 \times f_2$, with n_2 kernels and n_2 feature maps as output. The third convolutional layer uses kernels of size $f_3 \times f_3$, with n_3 kernels and n_3 feature maps as outputs. The final output feature map is the reconstructed high-resolution image. In SRCNN, the values of f_1, f_2, f_3, n_1, n_2, and n_3 are set to 9, 1, 5, 64, 32, and 3, respectively.

3 | PROPOSED METHOD

This section presents the details of the proposed method. First, a brief review of the baseline YOLOv7 is presented. Then, the overall architecture of the proposed MFENet (as shown in Figure 2) is introduced. Finally, the three-way dehazing module (3WDM) and the multi-granularity feature enhancement model (MFEM) in the proposed MFENet are explained in detail.

3.1 | YOLOv7 as baseline

Compared with two-stage detectors, YOLOv7 is a real-time object detection model based on a deep neural network, which quickly identifies and locates objects. The Backbone uses ELANet-l to extract image features, which are composed of ELAN and MP-1 structures. By feeding the features

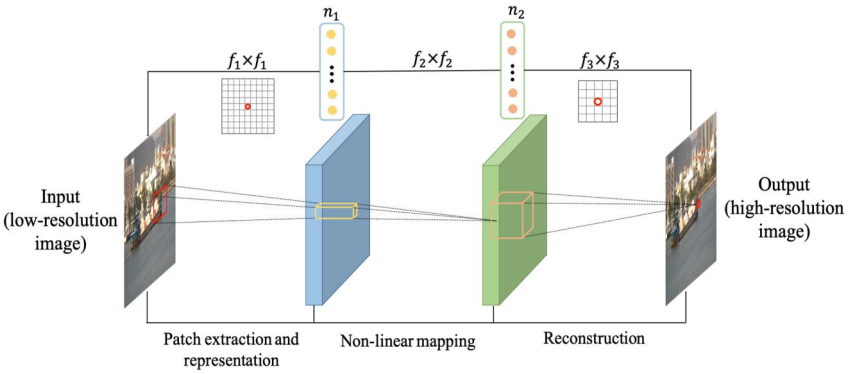

FIGURE 1 Architecture of SRCNN. SRCNN consists of feature block extraction and representation, non-linear mapping, and reconstruction. SRCNN, super-resolution reconstruction convolutional neural network.

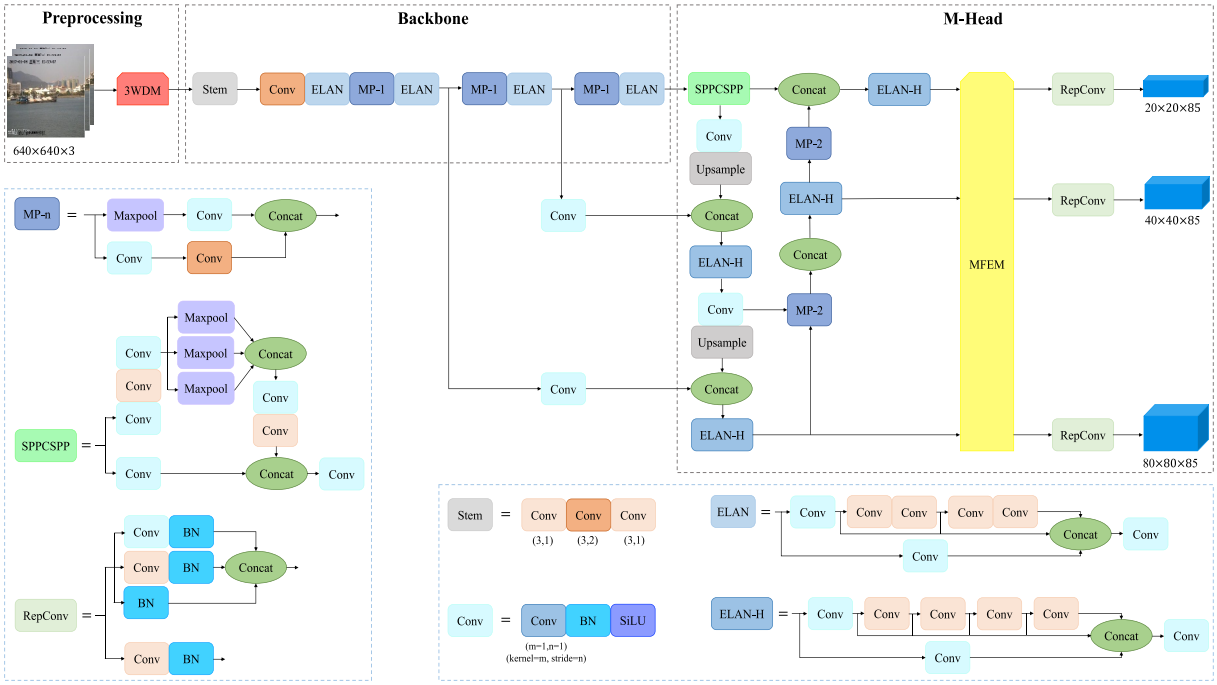

FIGURE 2 Architecture of the proposed MFENet. Compared with the baseline YOLOv7, MFENet proposes a 3WDM, which aims to eliminate the interference of sea fog on ship detection in visible light images. Simultaneously, it also designs an MFEM to improve the resolution and details of the feature map, thereby enhancing the feature representation of ships of different sizes. Conv, BN, SiLU, Maxpool, and Concat refer to convolution, batch normalisation, SiLU activation function, max pooling, and tensor concatenation, respectively. MFEM, multi-granularity feature enhancement model; 3WDM, three-way dehazing module.

extracted from the Backbone into the Head, YOLOv7 generates predictions at three different scales (20 × 20, 40 × 40, and 80 × 80), capturing objects of various sizes and preserving spatial information at different levels within the model.

The Stem layer consists of three convolutional layers, and the Conv layer consists of convolution, batch normalisation, and SiLU activation functions. The ELAN layer is used to adjust the length of the gradient path to force the network to learn more features. The difference between ELAN and ELAN-H is the number of splices. MP-n is mainly responsible for spatial downsampling, generating features with n times the input feature channels while halving the spatial resolution. The SPPCSPC module utilises maximal pooling to alleviate computational load, enhance speed and accuracy, catering to the demands of multi-resolution images. The RepConv layer is mainly used for structural reparameterisation, which helps deploy and accelerate the model.

3.2 | Proposed MFENet

YOLOv7 is a single-stage object detector. To make it more suitable for maritime ship target detection, this paper selected YOLOv7 as the baseline and proposed MFENet (as shown in Figure 2), which combines three-way dehazing modules (3WDM) and a MFEM to improve YOLOv7. The proposed

MFENet consists of Preprocessing, Backbone, and M-Head. Preprocessing is mainly to insert our 3WDM into the Input of YOLOv7 to eliminate the influence of sea fog on visible light images. The Backbone remains unchanged from YOLOv7. M-Head represents adding an MFEM after the Head of YOLOv7, which further enhances the model's ability to express the target features of ships of different sizes, thereby obtaining accurate detection results.

3.3 | Three-way dehazing module

Visible light images have broad application prospects in maritime ship detection. However, they are often greatly affected by sea fog, resulting in varying degrees of attenuation in contrast, colour fidelity, and other aspects of the image, making the image blurry, which further affects the detection accuracy and leads to seriously missed detections and false alarms. To solve this problem, this paper proposes a three-way dehazing module (3WDM) to defog from visible light ship images and obtain high-quality visible light image samples for subsequent training (as shown in Figure 3). First, the idea of three-way decisions is integrated into the K-means algorithm to design an automatic classification algorithm for visible light image clarity to obtain clear images without sea fog and blurry images with sea fog. Second, the FFA-Net [21] is used to defog blurry images. Finally, the obtained clear images are fed into the deep neural network for training.

In the process of designing the automatic classification algorithm for visible light image clarity, the image is mainly classified according to different feature values. Since the selection of features directly affects the timeliness of the algorithm, we select the image entropy and gradient as features and input them into the K-means algorithm based on three-way decisions for classification. Finally, blurry images with sea fog and clear images without sea fog are obtained.

The pseudocode of our proposed 3WD-based automatic classification algorithm for visible light image clarity is shown in Algorithm 1. The algorithm primarily classifies image clarity based on different feature values. Since the selection of features directly affects the real-time performance of the algorithm, we choose the entropy function [47] and Laplacian gradient function [48] of the image as features to determine the clarity of the images. The steps of Algorithm 1 are as follows. First, for all images $I = (i_1, i_2, \ldots, i_n)$, we calculate the entropy function value $e(i_j)$ and the Laplacian gradient function value $l(i_j)$ for each image i_j $(j = 1, 2, \ldots, n)$, that is, the sample sets $E = (e(i_1), e(i_2), \ldots, e(i_n))$ and $L = (l(i_1), l(i_2), \ldots, l(i_n))$.

The entropy function of image i_j can be expressed as follows:

$$e(i_j) = -\sum_{a=0}^{N-1} p_a ln(p_a), \tag{1}$$

where N is the number of greyscale levels in the image $(N = 256)$, and p_a represents the probability of each greyscale level. The larger $e(i_j)$ is, the clearer the image becomes, and vice versa. Meanwhile, the Laplacian gradient function of image i_j can be written as follows:

$$\begin{aligned} l(i_j) &= \sqrt{g_x^2(s,t) + g_y^2(s,t)}, \\ g_x(s,t) &= g(s,t) * A, \\ g_y(s,t) &= g(s,t) * A, \\ A &= \frac{1}{6}\begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}, \end{aligned} \tag{2}$$

where g_x and g_y are the convolution of the Laplacian horizontal convolution kernel A and the vertical convolution kernel A at the pixel point (s, t), respectively. The larger $l(i_j)$ is, the clearer the image becomes, and vice versa. Then, we use the $K-$means clustering algorithm and the three-way decisions theory to cluster E and L separately. The specific steps are as follows:

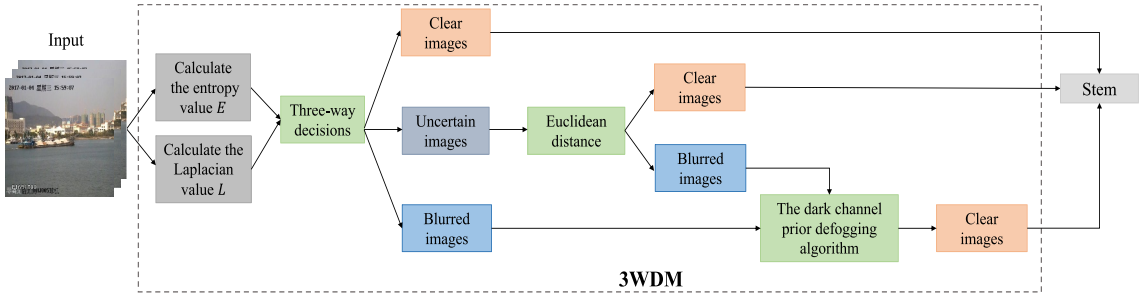

FIGURE 3 The structure of 3WDM. First, the 3WDM divides all input images into three categories based on the three-way decisions, including clear images without sea fog, uncertain images, and blurry images with sea fog. Second, the uncertain images are further divided by calculating the Euclidean distance. Then, the FFA-Net is used to defog the blurry images. Finally, the obtained clear images are inputted into a deep neural network for training. 3WDM, three-way dehazing module.

Step 1: Randomly select k samples from E as the initial mean vector $\{\mu_1, \mu_2, \ldots, \mu_k\}(1 \le m \le k)$.
Step 2: Let $C_m = \phi(1 \le m \le k)$ and calculate the distance d_{jm} between the sample e_j and each mean vector μ_m. d_{jm} can be expressed as follows:

$$d_{jm} = \left\| e(i_j) - \mu_m \right\|_2 \ (1 \le m \le k,\ j = 1, 2, \ldots, n). \tag{3}$$

Step 3: Determine the cluster label λ_j of $e(i_j)$ according to the nearest mean vector, which can be expressed as follows:

$$\lambda_j = \arg\ \min_{m \in \{1,2,\ldots,k\}} d_{jm}. \tag{4}$$

Step 4: Assign the calculated samples $e(i_j)$ to the corresponding cluster $C_{\lambda_j} = C_{\lambda_j} \bigcup \{e(i_j)\}$, $e(i_j)$ and i_j are one-to-one correspondence.
Step 5: Calculate the new mean vector μ_i', which can be expressed as follows:

$$\mu_m' = \frac{1}{|C_m|} \sum_{x \in C_m} x \ (1 \le m \le k). \tag{5}$$

Step 6: Repeat steps 2 and 5 until the current mean vector μ_m remains unchanged.
Step 7: Use E to divide all images I into blurry images C_1^b and clear images C_2^c. Similarly, use L to divide all images I into blurry images C_3^b and clear images C_4^c.

Next, based on the three-way decisions theory, we divide all images into blurry images $C_b = C_1^b \cap C_3^b$, clear images $C_c = C_2^c \cap C_4^c$, and uncertain images $C_u = I - (C_b \cup C_c)$. Finally, we calculate the threshold α belonging to the blurry image, and the threshold β belonging to the clear image is as follows:

$$\begin{aligned} \alpha_{C_b}^1 &= \min |l(i_j) - e(i_j)| (i_j \in C_b), \\ \alpha_{C_b}^2 &= \max |l(i_j) - e(i_j)| (i_j \in C_b), \\ \alpha &= \left\{ \alpha_{C_b}^1, \alpha_{C_b}^2 \right\}. \end{aligned} \tag{6}$$

$$\begin{aligned} \beta_{C_c}^1 &= \min |l(i_j) - e(i_j)| (i_j \in C_c), \\ \beta_{C_c}^2 &= \max |l(i_j) - e(i_j)| (i_j \in C_c), \\ \beta &= \left\{ \beta_{C_c}^1, \beta_{C_c}^2 \right\}, \end{aligned} \tag{7}$$

where $l(i_j)$ is the Laplace gradient function value of image i_j. $e(i_j)$ is the entropy function value of image i_j. And we calculate the Euclidean distance between the uncertain images C_u and α and β to determine the category of the uncertain images C_u as follows:

$$\begin{aligned} \gamma_{C_u}^1 &= \min |l(i_j) - e(i_j)| (i_j \in C_u), \\ \gamma_{C_u}^2 &= \max |l(i_j) - e(i_j)| (i_j \in C_u), \\ B &= \sqrt{\left(\gamma_{C_u}^1 - \alpha_{C_b}^1\right)^2 + \left(\gamma_{C_u}^2 - \alpha_{C_b}^2\right)^2}, \\ D &= \sqrt{\left(\gamma_{C_u}^1 - \beta_{C_c}^1\right)^2 + \left(\gamma_{C_u}^2 - \beta_{C_c}^2\right)^2}, \end{aligned} \tag{8}$$

where B represents the distance between the uncertain image and the blurry image in terms of the values of the Laplacian gradient function and the entropy function. D represents the distance between the uncertain image and the clear image in terms of the values of the Laplacian gradient function and the entropy function. If $B > D$, image i_j ($i_j \in C_u$) is assigned to the cluster of the clear image, and vice versa. Therefore, C_u is divided into blurry images C_5^b and clear images C_6^c, resulting in the final clear images without sea fog $I_c = C_2^c \cap C_4^c \cup C_6^c$ and blurry images with sea fog $I_b = C_1^b \cap C_3^b \cup C_5^b$.

Algorithm 1 Automatic screening of image clarity based on 3WD.

Require: All images $I = (i_1, i_2, \ldots, i_n)$, cluster k
Ensure: Clear images I_c, blurry images I_b
For all images I, $e(i_j)$ $(j = 1, 2, \ldots, n)$ and $l(i_j)$ are obtained by Equations (1) and (2). Note the sample set $E = (e(i_1), e(i_2), \ldots, e(i_n))$, $L = (l(i_1), l(i_2), \ldots, l(i_n))$
Select k samples randomly from E as the initial mean vector $\{\mu_1, \mu_2, \ldots, \mu_k\}(1 \le m \le k)$
Repeat
Let $C_m = \phi(1 \le m \le k)$
for $j = 1, 2, \ldots, n$ **do**
    Assign sample $e(i_j)$ into the corresponding cluster $C_{\lambda_j} = C_{\lambda_j} \bigcup \{e(i_j)\}$ by Equations (3) and (4), $e(i_j)$ and i_j are one-to-one correspondence. Update the mean vector μ_m' by Equation (5).
end for
Until the stop condition is met, the current μ_m remains unchanged.
The E is divided into blurry images C_1^b and clear images C_2^c. Similarly, the L is divided into C_3^b and C_4^c
Combined with the three-way decisions thought, get blurry images $C_b = C_1^b \cap C_3^b$, clear images $C_c = C_2^c \cap C_4^c$, and uncertain images $C_u = I - (C_b \cup C_c)$
The uncertain images C_u is divided into clear images C_6^c and blurry images C_5^b by Equations (6)–(8).
Return $I_c = C_2^c \cap C_4^c \cup C_6^c$, $I_b = C_1^b \cap C_3^b \cup C_5^b$

3.4 | Multi-granularity feature enhancement module

Aiming at the characteristics of various sizes of ships in visible light images, this paper designs an MFEM (as shown in Figure 4). This module aims to improve the quality of the feature maps Y_1, Y_2, and Y_3 outputted by the Head part of YOLOv7. It utilises information from multi-granularity layers to effectively accommodate various sizes of ship features in visible light images. Specifically, we introduce an improved SRCNN (as shown in Figure 4) for the feature maps Y_1, Y_2, and Y_3 of sizes 20 × 20, 40 × 40, and 80 × 80 output by the Head part and perform feature enhancement processing on them. The improved SRCNN reconstructs and enhances the detailed information of feature maps by learning the mapping relationship of feature maps. In this way, we can obtain richer and more accurate feature representations from different granularity levels to better capture the tiny features of ships of different sizes. This multi-granularity enhancement strategy fully utilises the information from different granularity levels, improving the quality and details of the feature maps. It enables better adaptation to the detection requirements of ships of different sizes and enhances the overall performance of ship detection.

Compared with SRCNN, the improved SRCNN maintains the same receptive field size while reducing the number of parameters, which benefits the model by adding MFEM to obtain the best ship detection accuracy at a faster speed. The improved SRCNN mainly consists of the following three convolutional layers.

The first convolutional layer is to extract and represent the features of the low-resolution feature map Y_i, and its operation can be expressed as F_1:

$$F_1(Y_i) = \max(0,\ W_1 * Y_i + B_1)\ (i\ =\ 1,\ 2,\ 3) \quad (9)$$

where F_1 represents the mapping relationship of the feature extraction layer, and W_1 and B_1 represent the filter and bias, respectively. W_1 contains 64 convolution kernels, and each of which is a 5 × 5 kernel with a dilation rate of 2 [49]. B_1 is a 64-dimensional vector, and each element is associated with a filter. After the first layer, a 64-dimensional feature map is extracted for each low-resolution feature map.

The second convolutional layer remaps a 64-dimensional feature map into a 32-dimensional feature map. The operation of the second layer can be expressed as F_2.

$$F_2(Y_i) = \max(0,\ W_2 * F_1(Y_i) + B_2)\ (i\ =\ 1,\ 2,\ 3) \quad (10)$$

where F_2 represents the mapping relationship of the non-linear mapping layer, and W_2 contains 32 convolution kernels, each of which is a 1 × 1 kernel. B_2 is a 32-dimensional vector. After the second non-linear mapping layer, the mapping process from low-resolution features to high-resolution features is completed.

The third convolutional layer recombines the high-resolution features obtained from the second layer, and its operation can be expressed as F.

$$F(Y_i) =\ W_3 * F_2(Y_i) + B_3\ (i\ =\ 1,\ 2,\ 3) \quad (11)$$

where F represents the mapping relationship of the reconstruction layer, W_3 contains 3 convolution kernels, each of which is a 5 × 5 kernel, and B_3 is a 3-dimensional vector.

4 | EXPERIMENTS

4.1 | Datasets

We primarily choose two benchmark datasets, McShips and SeaShips, for marine ship detection. Table 1 shows the ship category distribution for each dataset. Next, we will introduce these two datasets.

McShips [50] is a challenging and multi-class dataset provided by Northwestern Polytechnical University, which includes military ships in addition to the ship types found on the SeaShips dataset. The dataset contains 14,709 images (9000 images are publicly available) with characteristics such as changing weather conditions, multiple scale changes, and cluttered backgrounds. The dataset is split into training and

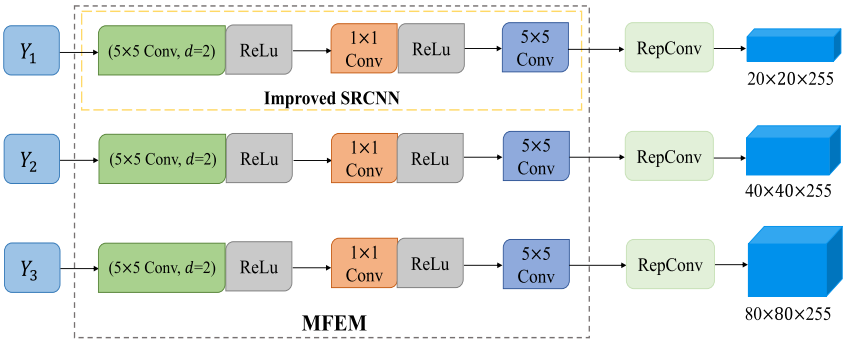

FIGURE 4 The structure of MFEM. MFEM includes three branches, each consisting of an improved SRCNN. The improved SRCNN contains three convolutional layers for feature extraction, representation, non-linear mapping, and reconstruction. The main purpose of MFEM is to improve the accuracy of the model in detecting ships of different sizes in visible light images. MFEM, multi-granularity feature enhancement module; SRCNN, super-resolution reconstruction convolutional neural network.

TABLE 1 Number of images for each ship category on different datasets.

Dataset	Ship category	Images	Percentage	Objects	Percentage
McShips	Civilian ship	4129	0.4919	5382	0.4750
	Warship	4265	0.5081	5949	0.5250
	Total	8394	1	11,331	1
SeaShips	Container ship	2084	0.2537	2199	0.2385
	Passenger ship	455	0.0554	474	0.0514
	Ore carrier	898	0.1093	901	0.0977
	General cargo ship	1426	0.1736	1505	0.1632
	Bulk cargo carrier	1811	0.2205	1952	0.2117
	Fishing boat	1539	0.1874	2190	0.2375
	Total	8213	1	9221	1

testing sets with an 8:2 ratio, and each image is cropped to a size of 640 × 640.

SeaShips [51] is the first publicly available ship detection dataset provided by Wuhan University. The dataset includes 31,455 images (7000 images are publicly available) with a resolution of 1080 × 1920 pixels, captured by 156 surveillance cameras installed in Hengqin Island, Zhuhai, China. It contains six different categories of ships with varying sizes, including bulk carriers, general cargo ships, container ships, fishing boats, passenger ships, and ore carriers, and has characteristics such as variations in ship scale, complex background interference, and changes in lighting conditions. The dataset is divided into training and test sets in an 8:2 ratio, with each image resized to 640 × 640.

4.2 | Evaluation metrics

We primarily use average precision (AP), mAP, and frames per second (FPS) as evaluation metrics to assess the performance of different maritime 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 labelled as true positive (TP) if the IoU between the two boxes exceeds a threshold. Otherwise, it is labelled as false positive (FP). A ground truth box is labelled as false negative (FN) if it has no corresponding detections. AP and mAP are obtained by calculating precision $P = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$ and recall rate $R = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$. They can be expressed as $\mathrm{AP} = \int_0^1 P(R)dR$ and $\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i$, where AP_i represents an AP value for each class i, A denotes the total number of classes. Meanwhile, we consider model parameters (Params) and floating point operation counts (FLOPs) to evaluate the complexity of the proposed method.

4.3 | Experimental settings

The experiments are run on 2 T V100 GPUs with 32 GB memory, and we use CUDA11.4, CUDNN8.0.4, and Pytorch1.9.0. We set k to two in Algorithm 1. ELANet-l serves as the backbone network. We utilise the SGD optimiser with an initial learning rate of 1×10^{-3} and a batch size of 8. The momentum and weight decay are set to 0.9 and 0.0001, respectively. The McShips dataset is trained for 60 epochs, while the SeaShips dataset is trained for 50 epochs.

4.4 | Parametric analysis

4.4.1 | Comparing the effects of different dehazing algorithms on 3WDM

In Section 3.3, we propose a 3WDM and study the impact of different dehazing algorithms on the detection performance of 3WDM. We conduct a comparison of FFA-Net with other dehazing algorithms to verify the effects of FFA-Net on 3WDM. We keep other settings unchanged and only replace FFA-Net with AOD-Net or DCPDN, as shown in Table 2. We find that 3WDM using FFA-Net outperforms the baseline by 1.09% and 1.02% in terms of mAP on the McShip and SeaShips datasets, respectively. In contrast, the performance of AOD-Net and DCPDN models is relatively weak. This indicates that FFA-Net is very effective in improving the detection performance of the model.

4.4.2 | Comparing the impact of different anchor boxes on YOLOv7

To improve the accuracy of ship detection, it is necessary to redesign the anchor box. Because the original anchor box of YOLOv7 is mainly suitable for general object detection and cannot meet the needs of ship detection. We use the K-means algorithm to redesign the anchor box on the McShips and SeaShips datasets, which are listed in Tables 3 and 4. Figure 5 compares the redesigned anchor box with the original YOLOv7 anchor box in two datasets. Furthermore, Figure 6 shows the distribution of ground truth width and height for ships in two datasets. By comparing Figures 5 and 6, we

TABLE 2 Comparison of experimental results in 3WDM using different dehazing algorithms.

	McShips		SeaShips	
Methods	**mAP (%)**	**FPS**	**mAP (%)**	**FPS**
Baseline	93.26	**60.85**	93.92	**67.57**
AOD-Net	93.84	53.02	94.02	58.63
DCPDN	94.01	54.34	94.56	59.24
FFA-Net	**94.28**	55.17	**95.01**	60.12

Note: Bold highlights best results.

TABLE 3 Redesigned anchor boxes on McShips.

The granularity of feature map	Anchor box 1	Anchor box 2	Anchor box 3
Fine-grained	(25, 17)	(45, 53)	(69, 27)
Medium-grained	(100, 122)	(169, 60)	(214, 230)
Coarse-grained	(391, 106)	(425, 309)	(441, 183)

TABLE 4 Redesigned anchor boxes on SeaShips.

The granularity of feature map	Anchor box 1	Anchor box 2	Anchor box 3
Fine-grained	(30, 14)	(61, 25)	(86, 42)
Medium-grained	(114, 27)	(135, 72)	(146, 46)
Coarse-grained	(211, 97)	(242, 57)	(338, 105)

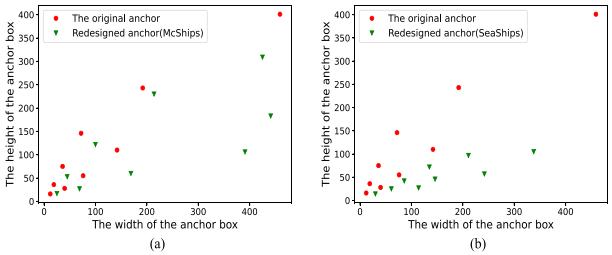

FIGURE 5 Comparison of the redesigned anchor box and the original YOLOv7 anchor box in two datasets. (a) McShips. (b) SeaShips.

clearly observe that there are significant differences between the two anchor boxes. Simultaneously, we also conduct a series of experiments on the original and the redesigned anchor box on the two datasets, and the results are shown in Table 5. Experiments show that the redesigned anchor box can significantly improve the accuracy of the YOLOv7 algorithm in ship detection, increasing the detection accuracy of McShips and SeaShips by 1.86% and 1.75%, respectively. Therefore, it is necessary to redesign the anchor box to improve the accuracy of ship detection.

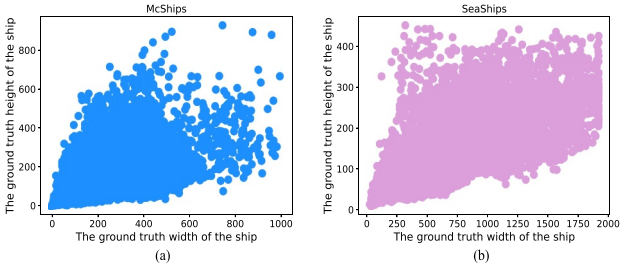

FIGURE 6 Distribution of ground truth width and height for ships in two datasets. (a) McShips. (b) SeaShips.

TABLE 5 The detection effect of different anchor boxes on different datasets.

	The original anchor boxes		Redesigned anchor boxes	
Detection algorithm	**mAP (%)**	**FPS**	**mAP (%)**	**FPS**
YOLOv7(McShips)	91.40	60.85	**93.26**	60.85
YOLOv7(SeaShips)	92.17	67.57	**93.92**	67.57

Note: Bold highlights best results.

4.4.3 | Comparing the impact of training epochs on different datasets

To investigate the effects of training epochs on the McShips and SeaShips datasets, Figure 7 presents the accuracy-epoch and loss-epoch curves on the McShips and SeaShips datasets during training. The McShips dataset quickly converges after 60 training epochs and exhibits minor variations in the subsequent training. In contrast, the SeaShips dataset requires fewer training epochs to achieve sufficient convergence and demonstrates a more stable trend within 50 training epochs. The McShips and SeaShips datasets are trained for 60 and 50 epochs, respectively.

4.5 | Ablation study

To investigate the impact of different modules on MFENet, this subsection conducts a series of experiments on the McShips and SeaShips datasets to validate the proposed MFENet.

4.5.1 | YOLOv7 as baseline

YOLOv7 is a single-stage object detection network with the most balanced speed and accuracy, especially for the detection of small targets. As shown in Tables 6 and 7, YOLOv7 achieves 93.26% mAP and 93.92% mAP on the McShips and SeaShips datasets, respectively. These results demonstrate the good competitive performance of our baseline.

4.5.2 | Performance of 3WDM

The proposed MFENet uses 3WDM to dehaze images containing sea fog in visible light images. To verify the effectiveness of 3WDM, we test Baseline and Baseline + 3WDM on McShips and SeaShips, as shown in Tables 6 and 7. In Table 6, compared with Baseline, the addition of 3WDM increases the AP of civilian ships and warships by 1.31% and 0.72%, respectively, and the mAP reaches 94.28%. In Table 7, the AP on each category is higher than the Baseline, reaching 95.01% mAP. The experimental results verify that the designed 3WDM significantly improves the detection accuracy of ships at sea. Because using 3WDM can effectively classify blurry images with sea fog and clear images without sea fog and defog the blurry images, it avoids the influence of sea fog on the clarity of visible light ship images, reduces missed and false detections in ship detection, and achieves efficient localisation and classification of ships.

4.5.3 | Performance of MFEM

To evaluate the effectiveness of MFEM, we conduct experiments on the McShips and SeaShips datasets and compare the results with Baseline and Baseline + 3WDM, as shown in Tables 6 and 7. In Table 6, compared with Baseline, Baseline + MFEM increases the AP of each class by 2.19% and 1.87%, respectively. Meanwhile, Baseline + 3WDM + MFEM not only achieves 96.28% mAP but also increases the AP of civilian ships and warships to 93.85% and 98.71%, respectively. In Table 7, we can see that Baseline + MFEM increases the AP of container ship, passenger ship, ore carrier, general cargo ship, bulk cargo carrier, and fishing boat by 2.54%, 2.69%, 2.92%, 2.10%, 4.20%, and 1.93%, respectively, reaching 96.65% mAP. Moreover, Baseline + 3WDM + MFEM improves the mAP by 2.60% compared to Baseline + 3WDM. This demonstrates that MFEM can effectively improve the detection accuracy of ships of different sizes by improving the resolution of feature maps of different granularity layers, thereby reducing missed detections and false alarms in visible light ship detection.

In addition, it is worth noting that, compared to the Baseline, the number of parameters and FLOPs of MFENet (Baseline + 3WDM + MFEM) on the McShips dataset only increase by 4.61 M and 9.59 G. Similarly, on the SeaShips dataset, MFENet only adds 4.61 M parameters and 11.64 G FLOPs. Therefore, our MFENet significantly improves detection performance without significantly increasing the number of parameters and computing overhead.

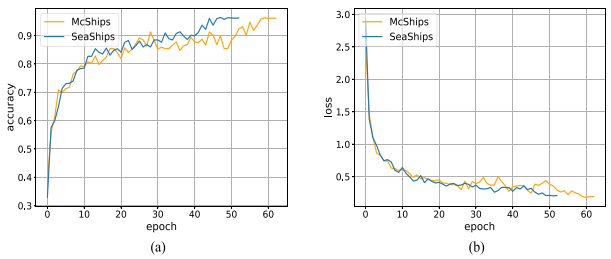

FIGURE 7 The accuracy-epoch and loss-epoch curves on the McShips and SeaShips datasets during training. (a) Accuracy-epoch. (b) Loss-epoch.

4.6 | Comparison with state-of-the-art

This subsection compares our MFENet with other state-of-the-art methods on the two datasets, that is, McShips and SeaShips. To ensure a fair comparison, we do not use the preprocessing module 3WDM for all algorithms. Meanwhile, to evaluate the performance of our MFENet more comprehensively, we also select several excellent algorithms that use our proposed 3WDM in the preprocessing. This step eliminates the influence of the preprocessing module 3WDM on the experimental results and ensures a fair evaluation of the performance of our MFENet.

4.6.1 | Results on McShips

We evaluate MFENet on the mAP metric. Under the mAP metric, we evaluate Faster R-CNN [8], FCOS [14],YOLOv5 [11], Deformable DETR [29], CBNet [13], PDNet [31], BUAA-PAL-OICR [33], CFP [30], HSF-Net [37], DAPN [35], Quad-FPN [32], BL-Net [38], SCSD [34], MDCN [36] and YOLOv7 [22] methods. The experimental results of MFENet on McShips are shown in Table 8. When all algorithms do not use our proposed preprocessing module 3WDM, our MFENet achieves 95.29% mAP, demonstrating the effectiveness of the network design of MFENet. This achievement benefits from incorporating the multi-granularity concept into SRCNN and

TABLE 6 Performance of different module combination strategies on McShips.

			AP (%)					
Baseline	**3WDM**	**MFEM**	**Civilian ship**	**Warship**	**mAP (%)**	**Params (M)**	**FLOPs (G)**	**FPS**
✓			90.82	95.70	93.26	**36.90**	**104.70**	**60.85**
✓	✓		92.13	96.42	94.28	41.36	113.53	55.17
✓		✓	93.01	97.57	95.29	37.05	105.46	56.46
✓	✓	✓	**93.85**	**98.71**	**96.28**	41.51	114.29	50.78

Note: Bold highlights best results.

TABLE 7 Performance of different module combination strategies on SeaShips.

Baseline	3WDM	MFEM	AP(%) Container ship	Passenger ship	Ore carrier	General cargo ship	Bulk cargo carrier	Fishing boat	mAP (%)	Params (M)	FLOPs (G)	FPS
✓			96.29	92.02	93.13	95.78	93.23	93.09	93.92	**36.90**	**95.81**	**67.57**
✓	✓		96.77	92.82	94.53	95.84	96.40	93.67	95.01	41.36	105.94	60.12
✓		✓	98.83	94.71	96.05	97.88	97.43	95.02	96.65	37.05	97.32	62.75
✓	✓	✓	**99.03**	**95.47**	**97.98**	**98.93**	**98.55**	**96.27**	**97.71**	41.51	107.45	56.43

Note: Bold highlights best results.

TABLE 8 Comparisons with state-of-the-art methods on McShips.

Methods	Backbone	AP(%) Civilian ship	Warship	mAP (%)	FPS
Object detection					
Faster R-CNN [8]	VGG16	70.19	83.23	76.71	11.46
FCOS [14]	VGG16	86.50	90.02	88.26	15.79
YOLOv5 [11]	CSPDarknet53	90.52	95.02	92.77	48.21
Deformable DETR [29]	ResNet50	89.01	92.43	90.72	12.13
CBNet [13]	ResNet50	91.27	96.07	93.67	23.10
PDNet [31]	ResNet50	90.77	94.12	92.45	15.74
BUAA-PAL-OICR [33]	VGG-16	91.05	92.57	91.81	17.13
CFP [30]	CSPDarknet53	91.58	96.52	94.05	30.17
CFP [30]*	CSPDarknet53	92.31	97.81	95.06	25.49
Ship detection					
HSF-Net [37]	VGG16	85.16	89.08	87.12	11.25
DAPN [35]	ResNet101	88.14	91.01	89.58	10.34
Quad-FPN [32]	ResNet50	90.19	93.25	91.72	11.88
BL-Net [38]	ResNet101	92.09	95.64	93.87	25.08
SCSD [34]	ResNet-101	91.55	94.36	92.96	9.46
MDCN [36]	ResNet-50	91.73	96.08	93.91	13.01
MDCN [36]*	ResNet-50	92.44	97.45	94.95	7.32
YOLOv7 [22]	ELANet-l	90.82	95.70	93.26	**60.85**
YOLOv7 [22]*	ELANet-l	92.13	96.42	94.28	55.17
MFENet (ours)	ELANet-l	93.01	97.57	95.29	56.46
MFENet (ours)*	ELANet-l	**93.85**	**98.71**	**96.28**	50.78

Note: Bold highlights best results.

using the improved SRCNN to enhance the resolution of different granularity features extracted by YOLOv7. Moreover, compared with several state-of-the-art methods using the same preprocessing module 3WDM, MFENet achieves the best performance, achieving 96.28% mAP. MDCN has the best detection performance on small and medium-sized ships, achieving 92.44% mAP at a speed of 7.32 FPS, while MFENet achieves 93.85% mAP at 50.78 FPS, far superior to MDCN. CFP performs best for large ships, achieving 97.81% mAP, while MFENet outperforms YOLOv7 by 0.9%. Therefore, MFENet can effectively detect ships of different sizes in visible light images, and its detection performance is far superior to other excellent algorithms.

4.6.2 | Results on SeaShips

We evaluate our MFENet on the SeaShips and compared it with other state-of-the-art methods. The results are presented in Table 9. As shown in Table 9, we can see that when all

TABLE 9 Comparisons with state-of-the-art methods on SeaShips.

Methods	Backbone	AP(%)						mAP (%)	FPS
		Container ship	Passenger ship	Ore carrier	General cargo ship	Bulk cargo carrier	Fishing boat		
Object detection									
Faster R-CNN [8]	VGG16	85.71	88.01	89.72	91.05	87.85	87.56	88.32	13.34
FCOS [14]	VGG16	95.41	91.01	90.03	91.04	91.41	92.34	91.87	20.34
YOLOv5 [11]	CSPDarknet53	96.11	91.08	93.34	94.77	93.61	91.53	93.41	56.05
Deformable DETR [29]	ResNet50	90.34	87.12	84.79	84.96	83.33	86.76	86.22	15.03
CBNet [13]	ResNet50	96.01	93.05	94.23	95.67	93.74	93.07	94.30	29.02
PDNet [31]	ResNet50	89.04	92.25	94.21	93.04	93.22	92.38	92.36	22.36
BUAA-PAL-OICR [33]	VGG-16	93.04	91.27	93.26	94.11	92.54	92.79	92.84	24.17
CFP [30]	CSPDarknet53	96.56	93.75	94.12	97.75	95.71	94.86	95.46	40.17
CFP [30]*	CSPDarknet53	98.51	94.92	96.07	97.92	96.63	95.53	96.60	35.75
Ship detection									
HSF-Net [37]	VGG16	90.34	87.12	84.79	84.96	83.33	86.76	86.22	12.54
DAPN [35]	ResNet101	85.73	89.22	89.04	90.45	88.33	87.65	88.40	11.20
Quad-FPN [32]	ResNet50	90.47	92.64	93.76	94.59	91.14	90.61	92.20	12.76
BL-Net [38]	ResNet101	94.70	91.01	93.38	94.67	93.41	92.74	93.82	30.77
SCSD [34]	ResNet-101	90.47	92.64	93.76	94.59	91.14	90.61	92.20	10.31
MDCN [36]	ResNet-50	95.56	92.45	94.12	93.51	93.31	95.43	94.06	15.44
MDCN [36]*	ResNet-50	97.64	93.01	95.60	96.03	94.22	95.27	95.30	9.70
YOLOv7 [22]	ELANet-l	96.29	92.02	93.13	95.78	93.23	93.09	93.92	**67.57**
YOLOv7 [22]*	ELANet-l	96.77	92.82	94.53	95.84	96.40	93.67	95.01	60.12
MFENet (ours)	ELANet-l	98.83	94.71	96.05	97.88	97.43	95.02	96.65	62.75
MFENet (ours)*	ELANet-l	**99.03**	**95.47**	**97.98**	**98.93**	**98.55**	**96.27**	**97.71**	56.43

Note: Bold highlights best results.

algorithms did not use our proposed preprocessing module 3WDM, our MFENet improves the mAP from 93.92% to 96.65% compared to the baseline YOLOv7 [22]. Our MFENet outperforms object detection algorithms in terms of performance, and this advantage is even more pronounced compared to ship detection methods. In particular, the ship detection method MDCN [36] achieves 94.06% mAP at a speed of 15.44 FPS, while our MFENet outperforms MDCN by 2.59% mAP, with the highest accuracy of 96.65% mAP and high speed than MDCN. In addition, Table 9 shows that compared to other state-of-the-art methods using the same preprocessing module 3WDM, our MFENet achieved the highest detection accuracy, reaching 97.71% mAP at 56.43 FPS. Compared with CFP, MDCN, and YOLOv7, our MFENet improves the detection accuracy by 1.11%, 2.41%, and 3.79%, respectively. Meanwhile, our MFENet achieves the best detection accuracy in the container ship, ore carrier, bulk cargo carrier, and bulk cargo carrier detection categories. 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 enhances the detection performance of ships of different sizes in visible light images.

4.7 | Visualising results and insight

4.7.1 | Visualisation results of different module combinations on McShips and SeaShips

To further elaborate on the combination effects of 3WDM and MFEM, we select two visible light images from McShips and SeaShips, respectively, and visualise their detection results. The visualisation results are shown in Figures 8 and 9. The first column is clear images without sea fog, and the second is blurry images with sea fog. There are two missed detections and one false alarm in Figure 8b, while there is one missed detection in

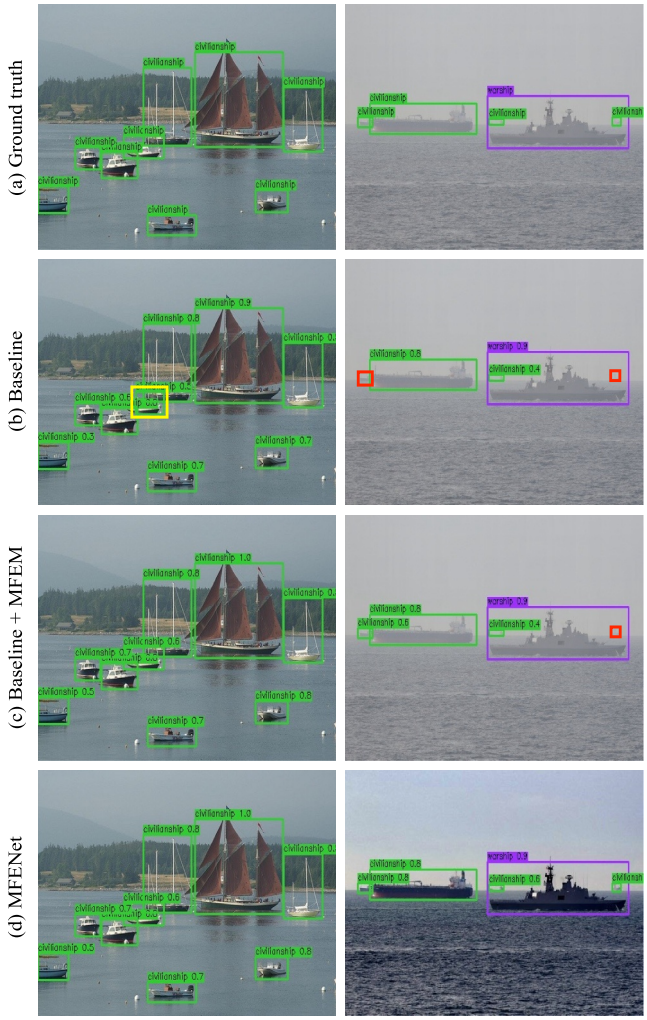

FIGURE 8 Comparison of detection visualisation results of different module combinations on McShips. Each row represents the detection results of ground truth, Baseline, Baseline + MFEM, and the proposed MFENet (Baseline + 3WDM + MFEM). Different coloured boxes represent different types of ship targets. Red and yellow boxes indicate missed detections and false alarms, respectively.

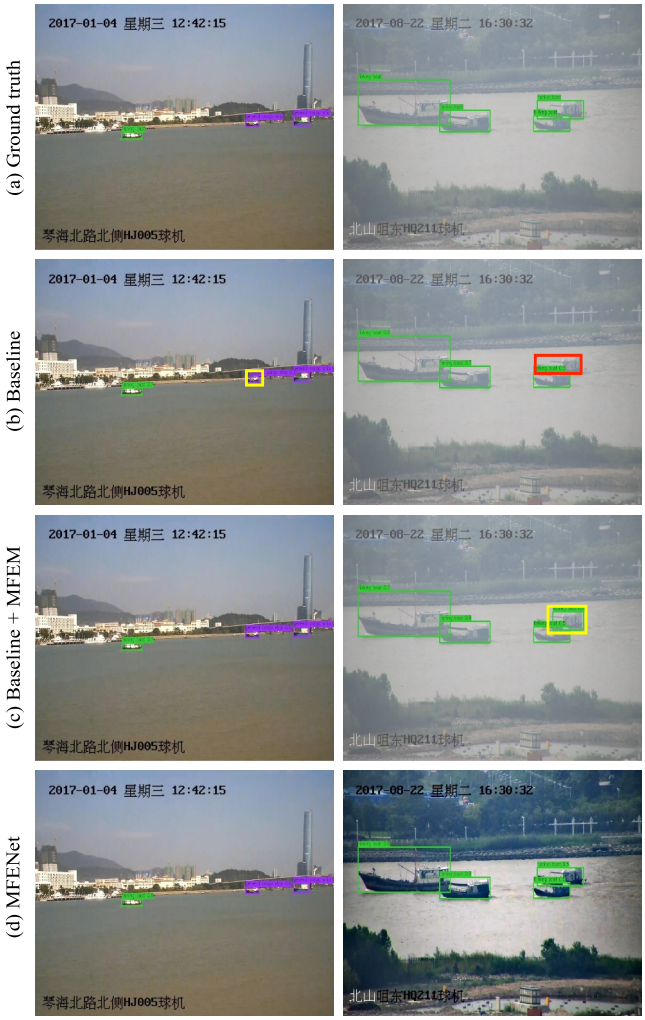

FIGURE 9 Comparison of detection visualisation results of different module combinations on SeaShip. Each row represents the detection results of ground truth, Baseline, Baseline + MFEM, and the proposed MFENet (Baseline + 3WDM + MFEM). Different coloured boxes represent different types of ship targets. Red and yellow boxes indicate missed detections and false alarms, respectively.

Figure 8c. Meanwhile, it is worth noting that all ships in Figure 8d are detected correctly. In addition, Figure 9b shows one missed detection and one false alarm. Figure 9c shows one false alarm. In contrast, every ship in Figure 9d is correctly detected. Therefore, MFENet (Baseline + 3WDM + MFEM) uses the fusion of three-way decisions and multi-granularity ideas to not only correctly distinguish different-sized ship targets in clear images without sea fog but also improve the detection accuracy of different-sized ship targets by improving the clarity of blurry images with sea fog.

4.7.2 | Visualisation results of different methods on McShips

To facilitate the comparison of ship detection performance among different algorithms, we show the visualisation results of several excellent algorithms and visually compare their missed detection and false alarm rates. These comparisons aim to highlight the unique strengths of our proposed MFENet compared to other methods. Figure 10 shows the visual comparison of detection results on McShips for ground truth, CFP [30], MDCN [36], YOLOv7 [22], and MFENet. In Figure 10, the first column is clear images without sea fog, and all compared detectors exhibit a certain degree of missed detection or false alarm. Specifically, YOLOv7 has the highest missed detection rate, missing one civilian ship. Meanwhile, CFP has the highest false alarm rate and a false alarm of one civilian ship. However, our MFENet can accurately detect each type of ship with low missed detection and false alarm rates. The second column is blurry images with sea fog. Due to the impact of sea fog and ship sizes, it is difficult to effectively detect all ships. Other detection methods also have varying degrees of missed detections and false alarms, with YOLOv7 having the most serious issues. Compared to other methods, MFENet exhibits significantly lower rates of false alarms and

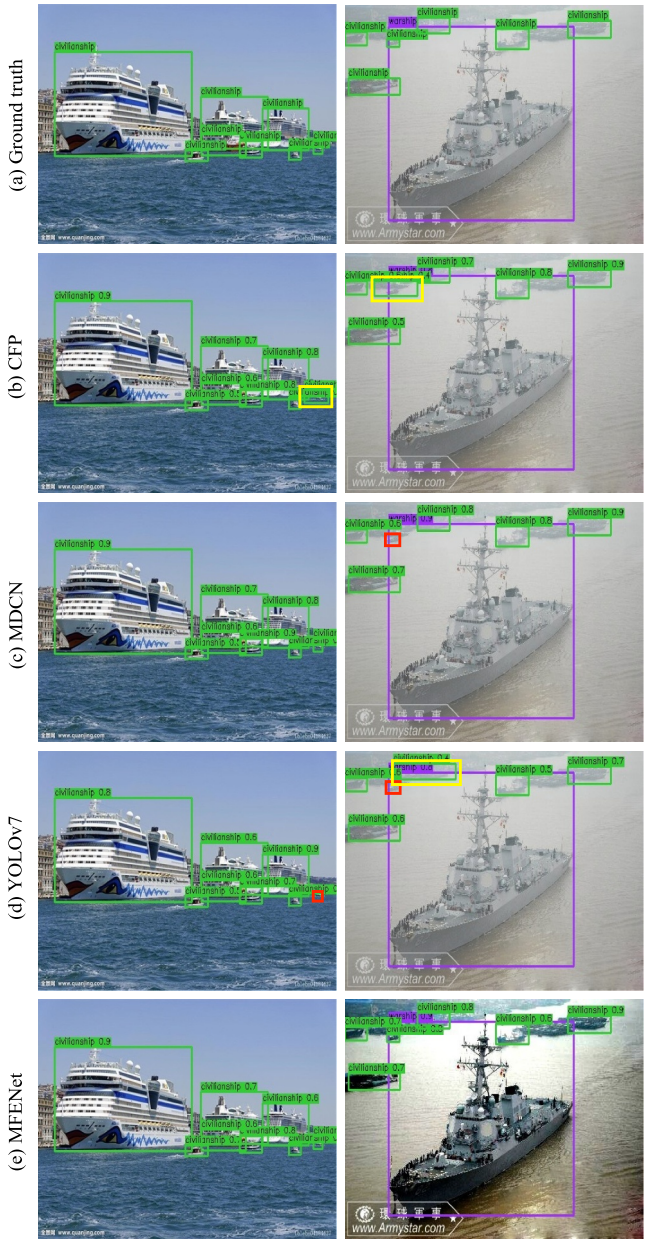

FIGURE 10 Comparison of detection visualisation results of different methods on McShips. Each row represents the detection results of ground truth, CFP, MDCN, YOLOv7, and the proposed MFENet. Different coloured boxes represent different types of ship targets. Red and yellow boxes indicate missed detections and false alarms, respectively.

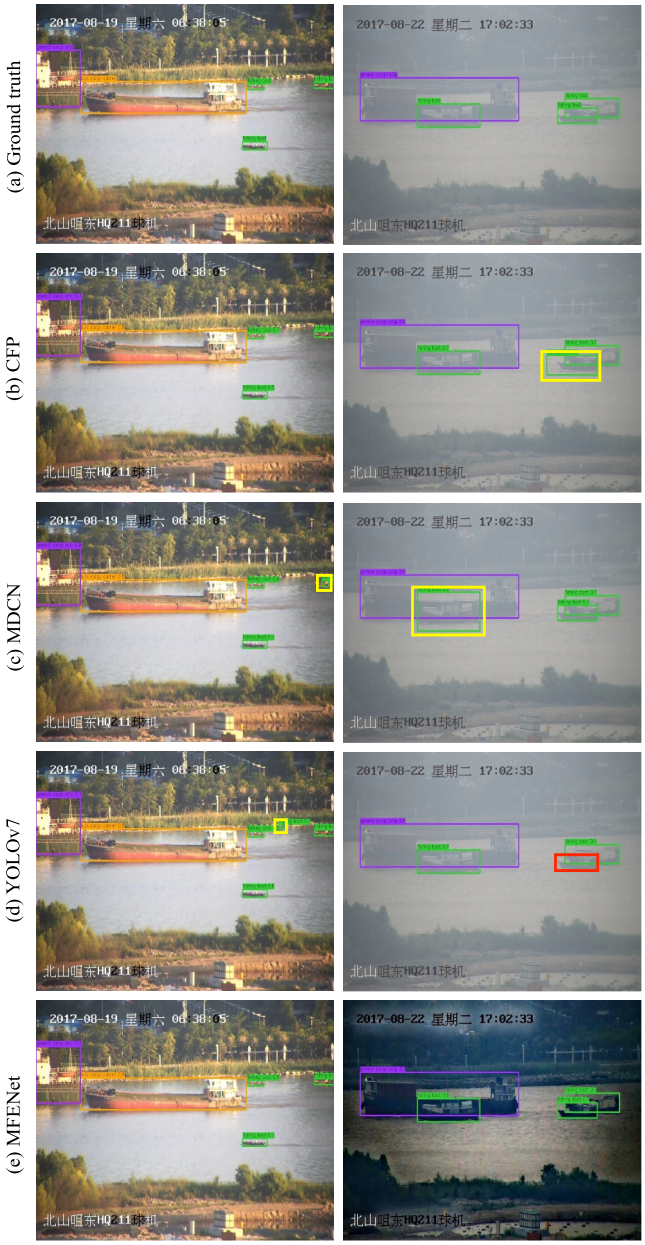

FIGURE 11 Comparison of detection visualisation results of different methods on SeaShips. Each row represents the detection results of ground truth, CFP, MDCN, YOLOv7, and the proposed MFENet. Different coloured boxes represent different types of ship targets. Red and yellow boxes indicate missed detections and false alarms, respectively.

missed detections. Additionally, its prediction accuracy for ships of various sizes in sea fog scenes is notably superior to that of other detection methods, highlighting its effectiveness in detecting ships of different sizes.

4.7.3 | Visualisation results of different methods on SeaShips

The visual comparison results of ground truth, CFP [30], MDCN [36], YOLOv7 [22], and MFENet on SeaShips are shown in Figure 11. The first column is clear images without sea fog, where MDCN and YOLOv7 falsely detect one fishing boat each. Both CFP and MFENet accurately detect all ship targets. The second column is blurry images with sea fog, where CFP and MDCN falsely detect one fishing boat each, while YOLOv7 misses one fishing boat. In contrast, our MFENet can effectively detect ship targets of different sizes in sea fog scenes. MFENet has the best detection performance, with very few missed detections and false alarms, indicating that MFENet is more suitable for detecting multi-sized ship targets in sea fog scenes.

TABLE 10 Detection results of different SAR ship datasets.

	Dataset			
	SSDD		LS-SSDD-v1.0	
Method	mAP	FPS	mAP	FPS
MFENet (ours)	94.58	47.28	74.36	56.14

4.8 | Comparison with SAR ship dataset

To evaluate the generalisation ability of our MFENet on the SSDD [52] and LS-SSDD-v1.0 [53] SAR ship detection, we conduct a series of experiments, and the results are shown in Table 10. Table 10 indicates that our MFENet achieves higher accuracy and FPS on both SSDD and LS-SSDD-v1.0, demonstrating its adaptability to SAR ships. This shows that MFENet not only performs well for the visible light ship but also exhibits good accuracy for the SAR ship.

5 | CONCLUSION

This paper proposes a novel single-stage network, MFENet, for ship detection in visible light images. We design 3WDM based on 3WD, which is used for dehazing foggy images and obtaining high-quality training samples. In addition, we also integrate the idea of multi-granularity into SRCNN and enhance the resolution of feature maps of different scales in YOLOv7 through the improved SRCNN, thereby improving the detection accuracy of YOLOv7 for ships of different sizes. Extensive ablation experiments show that the proposed MFENet can effectively improve the baseline performance and achieve the best performance on McShips and SeaShips datasets. Although 3WDM and MFEM significantly improve model performance by implementing data augmentation and enhancing multi-granularity feature representation. However, when ships tilt or rotate at different angles within the image, the use of rectangular bounding boxes fails to accurately capture the true shape of the ships, leading to a decrease in detection accuracy. In the future, to further improve the accuracy of visible light image ship detection, we will investigate the use of rotation rectangular boxes with angle information to detect ship targets in visible light images without reducing detection speed.

ACKNOWLEDGEMENT

The work is supported by the National Key Research and Development Programme (Gran No. 2022YFB3104700), the National Natural Science Foundation of China (Grant Nos. 62376198, 61906137, 62076040, 62076182, 62163016, 62006172), the Jiangxi 'Double Thousand Plan', the Jiangxi Provincial Natural Science Fund (No. 20212ACB202001), the China National Scientific Sea-floor Observatory, the Natural Science Foundation of Shanghai (Grant No. 22ZR1466700), and the Interdisciplinary Project in Ocean Research of Tongji University.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ['SeaShips'] at https://github.com/jiaming-wang/SeaShips.

ORCID

Li Ying https://orcid.org/0000-0003-4442-7367
Hongyun Zhang https://orcid.org/0000-0001-9781-5078

REFERENCES

1. Iwin, T., Sasikala, J., Sujitha, J.D.: Ship detection and recognition for offshore and inshore applications: a survey. IJIUS 7(4), 177–188 (2019). https://doi.org/10.1108/ijius-04-2019-0027
2. Wu, Z., et al.: Enhanced spatial feature learning for weakly supervised object detection. IEEE Transact. Neural Networks Learn. Syst., 1–12 (2022)
3. Wu, Z., et al.: Deep object detection with example attribute based prediction modulation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2020–2024. IEEE (2022)
4. Zhang, T., Zhang, X.: High-speed ship detection in SAR images based on a grid convolutional neural network. Rem. Sens. 11(10), 1–24 (2019). https://doi.org/10.3390/rs11101206
5. Zhou, M., et al.: Multi-resolution networks for ship detection in infrared remote sensing images. Infrared Phys. Technol. 92, 183–189 (2018). https://doi.org/10.1016/j.infrared.2018.05.025
6. Anupriya, K., Sasilatha, T.: Ship intrusion detection system-a review of the state of the art. In: Proceedings of the International Conference on Soft Computing Systems, pp. 147–154. Springer Singapore (2018)
7. Wu, F., et al.: Inshore ship detection based on convolutional neural network in optical satellite images. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, pp. 4005–4015 (2018)
8. Ren, S., et al.: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–1149 (2017). https://doi.org/10.1109/tpami.2016.2577031
9. Xu, X., et al.: A group-wise feature enhancement-and-fusion network with dual-polarization feature enrichment for SAR ship detection. Rem. Sens. 14(20), 1–24 (2022). https://doi.org/10.3390/rs14205276
10. Wu, Z., et al.: Multiple instance detection networks with adaptive instance refinement. IEEE Trans. Multimed. 25, 267–279 (2023). https://doi.org/10.1109/tmm.2021.3125130
11. Wang, C.-Y., Bochkovskiy, A. YOLOv5 release v6.1. https://github.com/ultralytics/yolov5/releases/]tag/v6.1 (2022)
12. Zhang, T., et al.: A hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery. ISPRS J. Photogrammetry Remote Sens. 167, 123–153 (2020). https://doi.org/10.1016/j.isprsjprs.2020.05.016
13. Liu, Y., et al.: CBNet: a novel composite backbone network architecture for object detection. IEEE Trans. Image Process. 31, 6893–6906 (2022). https://doi.org/10.1109/tip.2022.3216771
14. Tian, Z., et al.: FCOS: fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9627–9636 (2019)
15. Ying, L., Miao, D., Zhang, Z.: 3WM-AugNet: a feature augmentation network for remote sensing ship detection based on three-way decisions and multigranularity. IEEE Trans. Geosci. Rem. Sens. 61, 1–19 (2023). https://doi.org/10.1109/tgrs.2023.3313603
16. Miao, D., et al.: From human intelligence to machine implementation model: Theories and applications based on granular computing. CAAI Transactions on Intelligent Systems 6, 743–757 (2016)

17. Zhang, Y., et al.: Three-way enhanced convolutional neural networks for sentence-level sentiment classification. Inf. Sci. 477, 55–64 (2019). https://doi.org/10.1016/j.ins.2018.10.030
18. Zhao, C., et al.: Maximal granularity structure and generalized multi-view discriminant analysis for person re-identification. Pattern Recogn. 79, 79–96 (2018). https://doi.org/10.1016/j.patcog.2018.01.033
19. Xiao, J., et al.: Learning discriminative representation with global and fine-grained features for cross-view gait recognition. CAAI Transactions on Intelligent Systems 7(2), 187–199 (2022). https://doi.org/10.1049/cit2.12051
20. Ying, L., et al.: Multi-granularity-aware network for sar ship detection in complex backgrounds. Geosci. Rem. Sens. Lett. IEEE 8, 1–5 (2024). https://doi.org/10.1109/lgrs.2024.3352633
21. Qin, X., et al.: FFA-Net: feature fusion attention network for single image dehazing. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 11908–11915 (2020)
22. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7464–7475 (2023)
23. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893. Ieee (2005)
24. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60(2), 91–110 (2004). https://doi.org/10.1023/b:visi.0000029664.99615.94
25. Öztürk, Ş., Bayram, A.: Comparison of HOG, MSER, SIFT, FAST, LBP and CANNY features for cell detection in histopathological images. Helix 8(3), 3321–3325 (2018). https://doi.org/10.29042/2018-3321-3325
26. Dai, L., et al.: Ao2-detr: Arbitrary-oriented object detection transformer. IEEE Trans. Circ. Syst. Video Technol. 33(5), 2342–2356 (2023). https://doi.org/10.1109/tcsvt.2022.3222906
27. Xu, X., Zhang, X., Zhang, T.: Lite-YOLOv5: a lightweight deep learning detector for on-board ship detection in large-scene Sentinel-1 SAR images. Rem. Sens. 14(4), 1–27 (2022). https://doi.org/10.3390/rs14041018
28. Zhang, T., Zhang, X.: A mask attention interaction and scale enhancement network for SAR ship instance segmentation. Geosci. Rem. Sens. Lett. IEEE 19, 1–5 (2022). https://doi.org/10.1109/lgrs.2022.3189961
29. Zhu, X., et al.: Deformable DETR: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159 (2020)
30. Quan, Y., et al.: Centralized feature pyramid for object detection. IEEE Trans. Image Process. 32, 4341–4354 (2023). https://doi.org/10.1109/tip.2023.3297408
31. Yang, L., et al.: PDNet: towards better one-stage object detection with prediction decoupling. IEEE Trans. Image Process. 31, 5121–5133 (2022). https://doi.org/10.1109/tip.2022.3193223
32. Zhang, T., Zhang, X., Ke, X.: Quad-FPN: a novel quad feature pyramid network for SAR ship detection. Rem. Sens. 13(14), 1–30 (2021). https://doi.org/10.3390/rs13142771
33. Wu, Z., et al.: Selecting high-quality proposals for weakly supervised object detection with bottom-up aggregated attention and phase-aware loss. IEEE Trans. Image Process. 32, 682–693 (2023). https://doi.org/10.1109/tip.2022.3231744
34. Heng, X., et al.: Dual Teacher: a semisupervised cotraining framework for cross-domain ship detection. IEEE Trans. Geosci. Rem. Sens. 61, 1–12 (2023). https://doi.org/10.1109/tgrs.2023.3287863
35. Cui, Z., et al.: Dense attention pyramid networks for multi-scale ship detection in SAR images. IEEE Trans. Geosci. Rem. Sens. 57(11), 1–15 (2019). https://doi.org/10.1109/tgrs.2019.2923988
36. Zhang, X., et al.: A universal ship detection method with domain-invariant representations. IEEE Trans. Geosci. Rem. Sens. 60, 1–11 (2022). https://doi.org/10.1109/tgrs.2022.3200957
37. Li, Q., et al.: Multiscale deep feature embedding for ship detection in optical remote sensing imagery. IEEE Trans. Geosci. Rem. Sens. 56(12), 1–15 (2018)
38. Zhang, T., et al.: Balance learning for ship detection from synthetic aperture radar remote sensing imagery. ISPRS J. Photogrammetry Remote Sens. 182, 190–207 (2021). https://doi.org/10.1016/j.isprsjprs.2021.10.010
39. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. 33(12), 2341–2353 (2010)
40. Raanan, F., et al.: Dehazing using color-lines. ACM Trans. Graph. 34(13), 1–14 (2015). https://doi.org/10.1145/2699648
41. Berman, D., Treibitz, T., Avidan, S.: Single image dehazing using haze-lines. IEEE Trans. Pattern Anal. Mach. Intell. 42(3), 720–734 (2018). https://doi.org/10.1109/tpami.2018.2882478
42. Li, B., et al.: All-in-one dehazing network. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4770–4778 (2017)
43. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3194–3203 (2018)
44. Shao, Y., et al.: Domain adaptation for image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2808–2817 (2020)
45. Liu, X., et al.: GridDehazeNet: attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7314–7323 (2019)
46. Dong, C., et al.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38(2), 295–307 (2015). https://doi.org/10.1109/tpami.2015.2439281
47. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOB - Mob. Comput. Commun. Rev. 5(1), 3–55 (2001). https://doi.org/10.1145/584091.584093
48. Yao, Y., et al.: Evaluation of sharpness measures and search algorithms for the auto-focusing of high-magnification images. In: Proceedings of the IEEE Conference on Visual Information Processing XV, pp. 132–143 (2006)
49. Yu, F., Koltun, V., Funkhouser, T.: Dilated residual networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 472–480 (2017)
50. Zheng, Y., Zhang, S.: McShips: a large-scale ship dataset for detection and fine-grained categorization in the wild. In: Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 1–6 (2020)
51. Shao, Z., et al.: SeaShips: a large-scale precisely annotated dataset for ship detection. IEEE Trans. Multimed. 20(10), 2593–2604 (2018). https://doi.org/10.1109/tmm.2018.2865686
52. Zhang, T., et al.: SAR ship detection dataset (SSDD): official release and comprehensive data analysis. Rem. Sens. 13(18), 1–41 (2021). https://doi.org/10.3390/rs13183690
53. Zhang, T., et al.: LS-SSDD-v1.0: a deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images. Rem. Sens. 12(18), 1–37 (2020). https://doi.org/10.3390/rs12182997

How to cite this article: Ying, L., et al.: Multi-granularity feature enhancement network for maritime ship detection. CAAI Trans. Intell. Technol. 9(3), 649–664 (2024). https://doi.org/10.1049/cit2.12310