Preprints.org

Article

MrisNet: Robust Ship Instance Segmentation in Challenging Marine Radar Environments

Feng Ma , Zhe Kang , Chen Chen * , Jie Sun , Jizhu Deng

Article

MrisNet: Robust Ship Instance Segmentation in Challenging Marine Radar Environments

Feng Ma [1,2,3]**, Zhe Kang** [1,2,3]**, Chen Chen** [4,*]**, Jie Sun** [5] **and Jizhu Deng** [6]

[1] State Key Laboratory of Maritime Technology and Safety, Wuhan University of Technology, Wuhan 430063, China; martin7wind@whut.edu.cn(F.M.); kz258852@whut.edu.cn(Z.K.)

[2] National Engineering Research Center for Water Transport Safety, Wuhan University of Technology, Wuhan 430063, China

[3] Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China

[4] School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

[5] Nanjing Smart Water Transportation Technology Co., Ltd, Nanjing 210028, China

[6] Nanjing Port Co., Ltd, Nanjing 210011, China

* Correspondence: chenchen0120@wit.edu.cn

Abstract: In high-traffic harbor waters, marine radar frequently encounters signal interference stemming from various obstructive elements, thereby presenting formidable obstacles in the precise identification of ships. To achieve precise pixel-level ship identification in the complex environments, a customized neural network-based ship segmentation algorithm named MrisNet is proposed. MrisNet employs a lightweight and efficient FasterYOLO network to extract features from radar images at different levels, capturing fine-grained edge information and deep semantic features of ship pixels. To address the limitation of deep features in the backbone network lacking detailed shape and structured information, an adaptive attention mechanism is introduced after the FasterYOLO network to enhance crucial ship features. To fully utilize the multi-dimensional feature outputs, MrisNet incorporates a Transformer structure to reconstruct the PANet feature fusion network, allowing for the fusion of contextual information and capturing more essential ship information and semantic correlations. In the prediction stage, MrisNet optimizes the target position loss using the EIoU function, enabling the algorithm to adapt to ship position deviations and size variations, thereby improving segmentation accuracy and convergence speed. Experimental results demonstrate that MrisNet achieves high recall and precision rates of 94.8% and 95.2%, respectively, in ship instance segmentation, outperforming various YOLO and other single-stage algorithms. Moreover, MrisNet has a model parameter size of 13.8M and real-time computational cost of 23.5G, demonstrating notable advantages in terms of convolutional efficiency. In conclusion, MrisNet accurately segments ships with different spot features and under diverse environmental conditions in marine radar images. It exhibits outstanding performance, particularly in extreme scenarios and challenging interference conditions, showcasing robustness and applicability.

Keywords: ship segmentation; radar image; lightweight convolution; adaptive attention mechanism; loss function

1. Introduction

Real-time monitoring of busy waterways is a critically important task in maritime management, with positive implications for ensuring ship navigation safety, and improving port operational efficiency. Marine radar is widely applied in various domains such as ship collision avoidance, weather forecasting, and marine resource monitoring [1,2]. Similarly, shore-based deployed marine radar plays a significant role as it enables continuous monitoring of ships in expansive water areas, even under adverse weather and low visibility conditions. Compared to detection technologies like Automatic Identification System (AIS) and Very High-Frequency (VHF), marine radar does not rely on real-time information responses from ships, thereby greatly enhancing the speed of obtaining navigation information.

In general, the images presented by marine radar consists of a two-dimensional dataset composed of a series of spots. These spots represent the intensity and positional information of the radar signal's reflection and echo from objects such as ships, islands, and air masses in space. By analyzing and processing these spots, radar images with different resolutions and clarity can be generated. Additionally, by overlaying and calculating a series of radar images taken at different time intervals, the positions and trajectories of ships can be simulated, resulting in ship spots with varying lengths of trails.

Traditional ship recognition methods for marine radar images mainly involve techniques such as filtering and pattern recognition [3]. These methods demonstrate appropriate suppression capabilities when dealing with relatively simple interference. However, they are relatively inapplicable in tasks that involve low pixel features, slow movement speeds, and strong interference in ship recognition. In recent years, deep learning techniques have rapidly advanced in the field of object recognition, particularly with the widespread application of Convolutional Neural Networks (CNNs). CNN-based methods have shown more competitive performance in complex scenarios, such as equipment defect detection, autonomous driving, and aerospace applications. Compared to traditional methods, these approaches are capable of extracting deep semantic information from images and achieving more accurate object localization.

Both traditional recognition methods and CNN-based object detection algorithms can only provide rough positional information of the targets. As a derivative method in the field of computer vision, instance segmentation offers higher precision in pixel-level target localization and can effectively distinguish multiple instances. For marine radar images, instance segmentation methods also provide richer ship motion information, such as heading and speed. Therefore, utilizing these methods to process marine radar images can provide more accurate and comprehensive ship identification and tracking information for maritime authorities.

Compared to ship instance segmentation in natural images, performing such tasks in radar images faces more challenges. Firstly, there are only small numbers of spots represent actual moving ships. This significantly affects the accurate classification and localization of ship targets, while interferences such as waves, clouds, rain, clutter, etc., further increase the difficulty of interference removal. Secondly, both long-tail and short-tail ships in radar images are considered small or even tiny objects. Particularly, in cases of dense sailing or crossing navigation, distinguishing between ship spots can be challenging. Lastly, due to the extensive use of embedded devices in radar systems, traditional fractal algorithms and micro-Doppler techniques have been commonly employed for radar signal processing for several decades, despite their relatively limited adaptability in many cases. In contrast, deep learning-based recognition methods generally exhibit preferable performance, but they require more computational resources, and only a fraction of mature algorithms can be directly applied to embedded devices.

In response to the characteristics of marine radar images, this paper proposes a customized CNN-based instance segmentation algorithm called MrisNet. Compared to previous research, this method exhibits significant differences in several aspects.

[1] We enhance the feature network to extract crucial ship features by employing more efficient convolutional modules.

[2] A convolutional enhancement method that incorporates channel correlations is introduced to further enhance the generalization ability of the feature network.

[3] An attention mechanism with contextual awareness is utilized to enhance the multi-scale feature fusion structure, enriching the representation of convolutional features at different levels and effectively integrating micro-level and global-level ship features.

[4] The positioning loss estimation of predicted box is optimized to improve the precision of ship localization and enhance the segmentation performance for dense ship scenarios.

[5] To evaluate the effectiveness of various algorithms for ship segmentation in radar images, a high-quality dataset called RadarSeg, consisting of 1280 radar images, is constructed.

The remaining sections of this paper are organized as follows: Chapter 2 provides a brief overview of relevant research on instance segmentation and detection of ships in different imaging

scenarios. In Chapter 3, a customized CNN-based ship segmentation algorithm is proposed. Chapter 4 presents a comparative analysis of experimental results using various algorithms in marine radar images. Finally, Chapter 5 summarizes the main contributions of the proposed method and discusses future directions for development.

2. Related works

Currently, ship recognition in natural scenes and remote sensing images has been widely explored and developed. Early traditional methods primarily employed techniques such as clustering analysis and filtering for ship recognition [4]. However, these methods have obvious limitations and relatively poor adaptability, making it challenging to achieve satisfactory results in different scenarios. Meanwhile, the rapid development of deep learning and neural network-based technologies has brought significant advancements in tasks like instance segmentation. However, in the field of ship recognition in marine radar images, the application of deep learning-related methods remains relatively limited, presenting vast opportunities for further development.

In this section, we first provided an overview of ship recognition techniques in marine radar images, including neural network-based approaches for object detection and instance segmentation. It is worth noting that object detection and instance segmentation share commonalities in feature extraction, object localization, and post-processing, thus the design principles of object detection algorithms can offer valuable insights for the design of instance segmentation algorithms. Additionally, we reviewed and summarized ship recognition techniques and relevant optimization methods in other scenarios.

2.1. Ship identification methods under radar and other scenarios

In recent years, significant advancements have been made in object detection and instance segmentation using CNNs. In 2014, the R-CNN algorithm demonstrated a significant advantage on the PASCAL VOC dataset, gradually establishing the dominance of deep learning-based algorithms in the field of object recognition. As is widely known that single-stage algorithms have achieved a good balance between computational speed and recognition accuracy, allowing for end-to-end training and widespread adoption in various recognition tasks. However, in certain tasks, single-stage algorithms may exhibit slightly lower accuracy compared to two-stage algorithms such as Faster R-CNN [5] and Mask R-CNN [6]. Therefore, the selection of benchmark algorithms and network architectures requires careful consideration and decision-making based on specific application scenarios and requirements.

In the research on ship recognition in marine radar images, significant progress has been made by scholars who have incorporated deep learning methods into the field of object detection. By designing effective algorithm structure and utilizing techniques such as clutter suppression and feature enhancement, even single-stage algorithms can achieve satisfactory recognition results. Chen et al. [7] proposed a ship recognition algorithm based on a dual-channel convolutional neural network (DCCNN) and a false alarm-controllable classifier (FACC) to suppress clutter and accurately extract ship features in images. Furthermore, some studies have achieved high levels of ship recognition accuracy and effective interference suppression using two-stage recognition algorithms. For instance, Chen et al. [8] made several improvements to Faster R-CNN in multiple aspects, including optimizing the backbone network, sample data balancing, and scale normalization, aiming to enhance the algorithm's accuracy and robustness. Additionally, differential neural architecture search, label reassignment, and various types of feature pyramid structures have been widely applied in algorithm design, yielding promising recognition results [9–11]. In conclusion, deep neural networks have proven effective in extracting ship features. However, current object detection techniques still have limitations in suppressing complex interference, and there is room for improvement in recognizing dense small-sized ships.

Currently, research on ship segmentation under marine radar images is still relatively limited. It is well-known that ships in synthetic aperture radar (SAR) images exhibit small scales, limited effective pixels, and significant background interference, which are similar to the challenges faced in

marine radar images. In the context of SAR images, there has been extensive research on ship segmentation, and we can draw upon the technical methods from these studies to provide insights for the design of ship segmentation algorithms in marine radar images.

In ship segmentation under complex background conditions of SAR images, a well-structured feature extraction network plays a crucial role in enhancing the model's recognition capability. Zhang et al. [12] proposed methods such as multi-resolution feature extraction networks, and enhanced feature pyramid networks to ensure more adaptable performance in complex scenarios. Moreover, various attention mechanisms can significantly enhance the extraction of ship features and contextual information. Zhao et al. [13] proposed a ship segmentation method based on collaborative attention mechanism, which improved the recognition performance of multi-task branches. Typically, instance segmentation networks employ horizontal bounding boxes to fit the objects, which may include redundant information. Moreover, in dense ship scenarios, it becomes challenging to accurately differentiate between individual targets. Yang et al. [14] proposed a novel ship segmentation network that utilizes rotated bounding boxes as the segmentation foundation, enclosing the targets along the direction of the ships. Besides, they designed a dual feature alignment module to capture representative features of the ships. Instance segmentation of small-scale or multi-scale ships has been a focal topic in SAR images. Shao et al. [15] proposed a multi-scale ship segmentation method specifically designed for SAR images. Specifically, they achieved more precise target regression by re-designing the input unit, backbone network, and ROI module of the Mask R-CNN.

The above-mentioned methods provide valuable references for the design of instance segmentation algorithms in marine radar images, particularly in terms of feature network design, attention mechanism application, and multi-scale segmentation.

In other domains, research on ship segmentation primarily focuses on natural images. Researchers have employed various techniques and methods for algorithm design. Some studies have utilized single-stage algorithms, incorporating efficient backbones, feature fusion structures, and prediction networks to achieve high-precision ship segmentation [16]. Others have adopted two-stage algorithm designs, involving the extraction of target candidate regions followed by classification and fine-grained segmentation to enhance accuracy and robustness [17]. These diverse algorithm design strategies cater to the demands and challenges of various domains.

2.2. Optimization method for ship identification research

The ship recognition in radar images can be enhanced through the utilization of auxiliary techniques. For instance, employing multi-modal fusion techniques and clutter suppression techniques can result in clearer and higher-resolution radar images, thereby providing more accurate ship information. Guo et al. [18] proposed a method that utilizes deep learning techniques to identify targets in marine radar images and achieves consistent fusion of electronic chart and marine radar images by treating these targets as reference points. Mao et al. [19] introduced a marine radar imaging framework based on non-uniform imaging theory, which combines techniques such as beam recursive anti-interference, non-uniform sampling models, and dimensionality reduction iterations for improved computational efficiency and higher-quality imaging results. Zhang et al. [20] employed generative adversarial networks to remove noise in radar images and utilized image registration methods to eliminate imaging discrepancies between radar images and chart data, supporting efficient ship recognition through parallel computation of feature data and image fusion.

Currently, there are several techniques available to improve the ship recognition models. Among them, constructing datasets that include rich ship features is an effective method for enhancing recognition performance [21]. In addition, the ship detection can be increased by integrating feature augmentation and key point extraction into the recognition algorithms [22]. Furthermore, through the design of feature enhancement modules and small target attention mechanisms, substantial improvements in the recognition of small-scale ships might be made [23,24]. In certain scenarios, it is necessary to consider the computational limitations of embedded devices, thus requiring the design of efficient lightweight algorithms. Yin et al. [25] addressed this challenge by introducing depthwise separable convolutions and point-wise group convolutions, resulting in a

lightweight feature extraction network that balances the accuracy and inference speed of ship recognition.
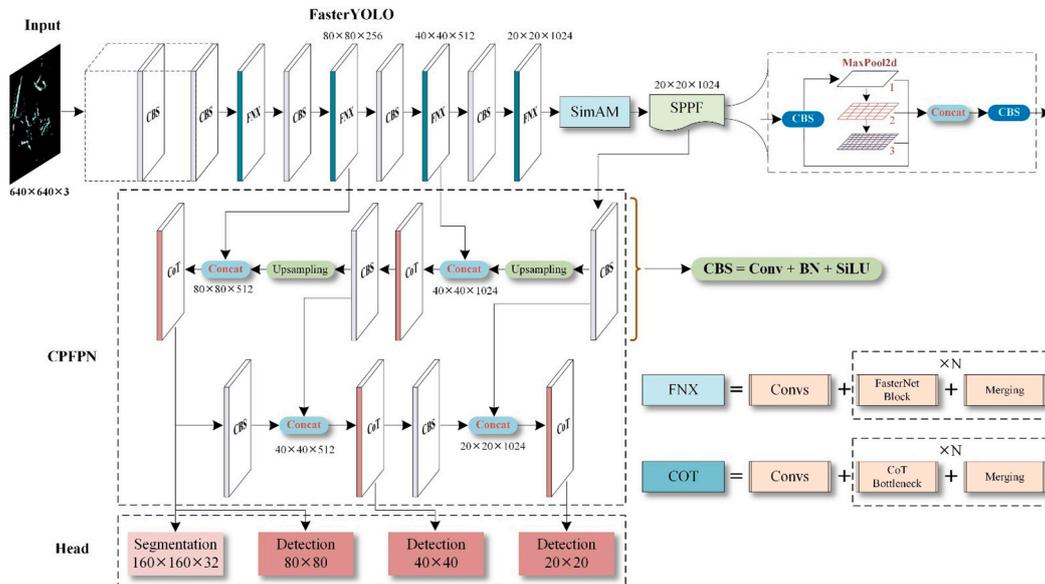

Figure 1. Overall structure of the proposed algorithm.

Based on the above analysis, it can be concluded that leveraging deep learning techniques to design feature networks and prediction structures significantly enhances the effectiveness of ship recognition. Furthermore, compared to traditional methods, deep learning approaches exhibit prominent advantages. Given the presence of numerous small-scale targets in marine radar images, this study proposes a ship segmentation algorithm that combines deep convolutional networks, feature attention mechanisms, and multi-scale feature fusion structures. By integrating different hierarchical features from the images, this algorithm effectively suppresses interference and enables accurate localization and precise segmentation of different types of ships.

3. A Proposed Method

The overall framework of the ship segmentation model for marine radar images proposed in this paper is illustrated in Figure 1. The model consists of components such as the feature extraction network, feature fusion network, and prediction structure. To be specific, the feature network adopts a novel convolutional structure that combines efficient convolutional units and adaptive attention mechanisms. The network reduces redundant computations and improves the extraction of spatial features of the targets, facilitating the model's understanding of the hierarchical pixel and semantic information of ships. The feature fusion network employs a convolutional method based on the Transformer structure to achieve the fusion of ship features at different scales. It captures global and local dependencies in the feature sequences, thereby avoiding feature loss and degradation. Finally, in the prediction structure of the proposed algorithm, the EIoU loss [26] is introduced to optimize the loss calculation for predicting bounding box positions. This method improves the convergence speed and accuracy of the predicted boxes, enabling the model to learn feature representations with better generalization ability.

3.1. Feature extraction network

As mentioned earlier, marine radar images exhibit unique characteristics such as noise and low resolution, making them significantly different from natural images. Classic feature networks like ResNet-50, ResNet-101, and SENet are typically designed for general natural datasets and may not adapt well to the distinctive image features of radar images, resulting in suboptimal feature extraction performance. Furthermore, these networks often contain a large number of redundant parameters,

which can lead to overfitting issues when applied to ship segmentation in marine radar images. Therefore, employing lightweight customized feature networks may be more suitable for such tasks.

Furthermore, due to the generally small scale of ships in marine radar images, the feature network should possess the ability to perceive small objects. Moreover, to capture different signature features of ships, the feature network should employ suitable convolutional structures and downsampling ratios to obtain feature representations of targets at different scales, thereby enhancing the accuracy of ship segmentation. Based on the aforementioned analysis, the overall framework of the proposed feature extraction network is illustrated in Figure 2. This network is an adaptive neural network that effectively extracts ship features at various levels and achieves more prominent performance.
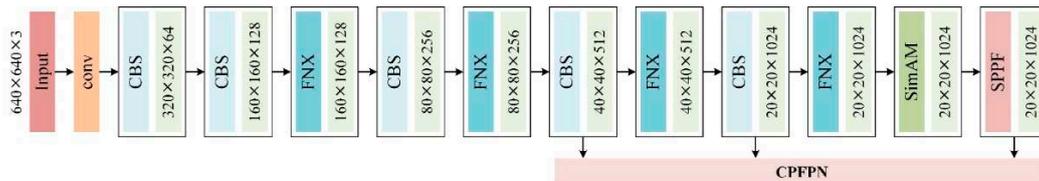

Figure 2. Overall architecture of the feature network.

3.1.1. YOLOv5(S) network

In this research, we selected the YOLOv5(S) [27] as the baseline feature network. Despite its relatively smaller convolutional scale, YOLOv5(S) is still able to perform effectively. The network is illustrated in Figure 3, YOLOv5(S) utilizes a series of downsampling layers to progressively extract image features, adapting to variations in target scale. It employs different sizes of convolutional kernels and pooling layers to design receptive fields at different levels, capturing multi-scale features of the targets while maintaining a lower number of convolutional parameters and computational complexity. This contributes to improving the accuracy and robustness of ship segmentation, particularly enhancing the recognition of small-scale ships in radar images. It is important to note that while YOLOv5(S) has fewer convolutional layers and parameters, as well as lower computational complexity, this lightweight design makes it more suitable for instance segmentation tasks in resource-constrained environments.

3.1.2. The main architecture of FasterYOLO network

In marine radar images, image sequences often contain ship regions of various sizes and shapes, with the majority of ships being relatively small in scale. Therefore, it is crucial to capture key ship features with high sensitivity. Additionally, the feature network should possess noise resistance and robustness. To address these requirements, valid
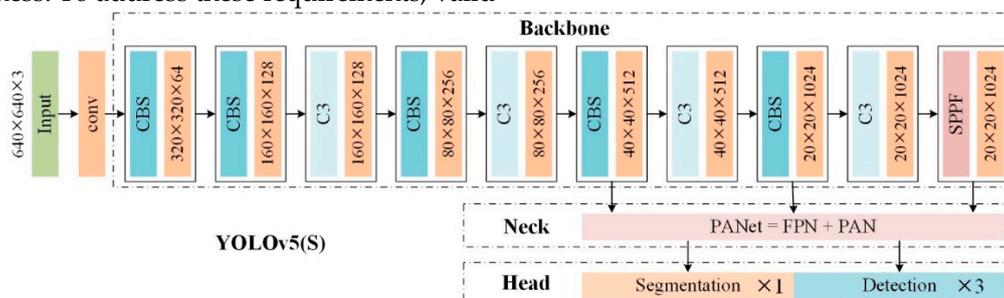

Figure 3. Network structure of instance segmentation of YOLOv5(S).

convolutional computation units can be employed to enhance network performance. This paper introduces a simple and fast convolutional method, combining standard convolutions and pointwise convolutions, to design a more efficient feature network.

With the advancement of CNNs, classical computation methods represented by standard convolutions often suffer from redundant calculations, resulting in inefficient increases in model parameters and computational costs. Moreover, a significant amount of ineffective convolutional computations can impact the extraction of crucial features, especially in ship recognition under radar images where targets are small in scale and feature information is limited. Excessive convolutional calculations can lead to overfitting issues. Research has revealed that convolutional feature maps exhibit high similarity across different channels, and standard convolution unavoidably duplicates the extraction of image features in a per-channel computation. Therefore, simplifying the convolutional computation to reduce feature information not only ensures network performance but also greatly enhances computational speed and efficiency. Specifically, this paper proposes a network structure called FasterYOLO, which incorporates the FasterNeXt modules [28] after standard convolutions, as illustrated in Figure 4. This improvement aims to enhance the model's ability to extract key ship features and reduce the risk of overfitting.
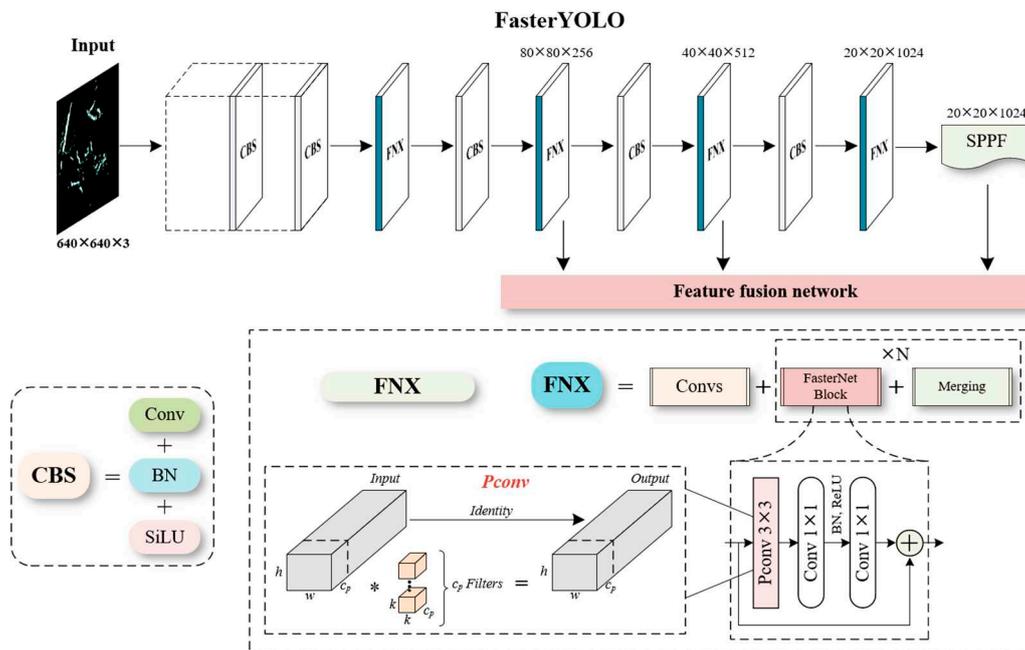

Figure 4. Network structure of FasterYOLO.

Compared to conventional convolutional networks, integrating the FasterNeXt modules in the feature network significantly reduces the computational cost of convolutions. The FasterNeXt employs PConv convolutions to optimize per-channel convolutional computations. In practical terms, PConv convolution is a convolutional structure with multiple branches and spatial feature extraction capabilities. It utilizes partial convolution operations, which means it performs conventional convolutions only on a subset of channels in the input feature maps, thereby reducing the computational workload. Additionally, PConv incorporates a parallel branch structure that performs convolutional operations on different spatial positions of the input feature maps, capturing more spatial feature information.

Figure 5 illustrates a brief comparison of the computational processes between PConv convolution, standard convolution, and depthwise convolution. Depthwise convolution, a classical convolutional variant widely used in various neural networks, has shown significant effectiveness in MobileNet series. For an input $I \in R^{c \times h \times w}$, We apply c filters $w \in R^{k \times k}$ to compute the output $O \in R^{c \times h \times w}$. In this case, the computational cost of standard convolution is $h \times w \times k^2 \times c^2$, while depthwise convolution only incurs a cost of $h \times w \times k^2 \times c$. It can be observed that depthwise convolution is effective in reducing computational cost. However, it directly leads to a decrease in recognition accuracy and therefore cannot directly replace standard convolution. On the other hand, PConv convolution only applies regular convolution to a subset of input channels, resulting in a

computational cost of $h \times w \times k^2 \times c_p^2$. When r is 1/4, the computational cost of PConv is merely 1/16 of that of standard convolution. Consequently, PConv convolution greatly simplifies the computational process of standard convolution while maintaining identical input and output dimensions. Without altering the network hierarchy, standard convolution can be directly replaced with PConv convolution.
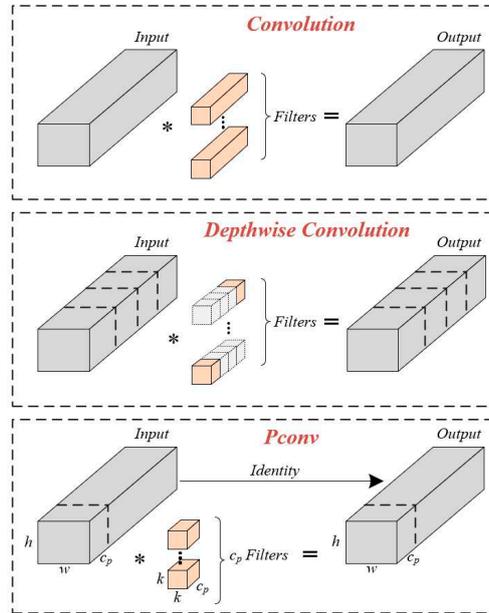

Figure 5. Comparison of several types of convolution methods.

3.1.3. Feature enhancement mechanism based on SimAM attention

For the feature network, it is crucial to fully leverage the contextual information surrounding ships in radar images to enhance the accuracy of ship segmentation. To fulfill this requirement, attention mechanisms can be introduced or convolutional structures with global contextual awareness can be utilized. By capturing the correlations between ships and their surrounding environments, the network can better comprehend the shape, contour, and semantic information of ships, thereby enhancing the effectiveness of ship segmentation.

In this research, the SimAM attention module [29], which is constructed based on principles derived from visual neuroscience, is appended after the FasterYOLO network. According to the theory of visual neuroscience, neurons carrying more information are more salient compared to their neighboring neurons when processing visual tasks, and thus, they should be assigned higher weights. In ship segmentation, it is equally important to enhance the neurons in the convolutional network that are responsible for extracting crucial ship features. SimAM captures both spatial and channel attention simultaneously and possesses spatial inhibition capabilities that are translation-invariant. Unlike methods such as SENet and CBAM that focus on designing attention mechanisms through pooling and fully connected layers, SimAM evaluates the importance of each feature based on an energy function derived from neuroscientific principles. It offers better interpretability and does not require the introduction of additional learnable parameters. Consequently, SimAM effectively extracts and enhances salient information of ships in marine radar images. Specifically, SimAM evaluates each neuron in the network by defining an energy function based on linear separability, as shown in Equations (1)-(4).

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (1)$$

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i \quad (2)$$

$$\widehat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i - \widehat{\mu})^2 \quad (3)$$

$$\widetilde{X} = sigmoid(1/e_t^*) \odot X \quad (4)$$

Among them, t is the target neuron, x is the adjacent neuron, M is the number of neurons, and λ is a hyperparameter. A lower energy value of e_t^* indicates higher discriminability between the neuron and its neighboring neurons, implying a higher level of importance for that neuron. In Equation (4), we weight the importance of neurons using $1/e_t^*$. The introduction of SimAM enables the feature network to comprehensively assess feature weights, thereby enhancing the representation of crucial ship information, and reducing reliance on prior information regarding significant variations in the shape of the targets.

3.2. Feature fusion network

In marine radar images, the majority of ships exhibit relatively small scales. Therefore, developing an efficient convolutional architecture to capture fine and effective feature representations of ships becomes crucial. In convolutional networks, shallow convolutional features possess higher resolution, providing more detailed spatial information that aids in the recognition of object edge structures. Moreover, shallow features are more robust to image noise and lighting variations. However, shallow features exhibit limited adaptability to transformations such as translation, scaling, and rotation in images. In contrast, deep convolutional features have lower resolution but contain stronger multi-scale and semantic information, which helps filter out the effects of noise and lighting variations in images. However, deep features have relatively weaker capability in extracting fine-grained details from images. Hence, this paper proposes a more efficient feature fusion network that fully integrates feature information extracted from different scales and depths of receptive fields in the images.

Currently, object detection or instance segmentation algorithms based on deep neural networks commonly employ feature pyramid structures to address the challenge of scale variation. Among them, Feature Pyramid Network (FPN) serves as the most widely used feature structure, delivering more adaptive results in both single-stage and two-stage algorithms. FPN realizes the fusion of features at different scales through a top-down feature propagation path. However, high-level features need to undergo multiple intermediate-scale convolutions and be fused with features at these scales before merging with low-level features. In this process, the semantic information of high-level features may be lost or degraded. In contrast, dual-path fusion structures such as PANet compensate for the shortcomings of FPN in preserving high-level features but also introduce the opposite problem, where the detailed information of low-level features may be degraded during fusion. To address this issue, this paper introduces an attention mechanism that integrates rich feature information into the PANet network. This mechanism adaptively learns the importance of different regions in the feature map and weights the fusion of features at different scales, resulting in more comprehensive feature representations. Based on these theoretical considerations, we refer to the proposed feature fusion network as CPFPN, and its overall structure is illustrated in Figure 6.
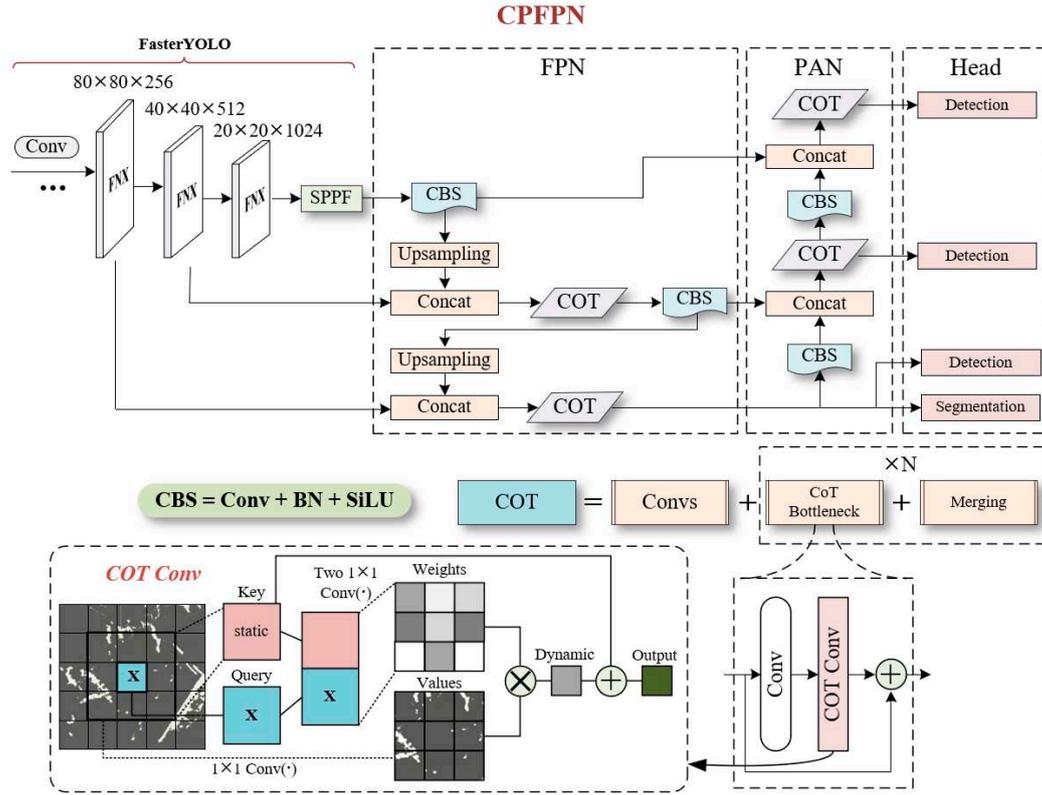
Figure 6. Network structure of CPPFN.

Specifically, in this research, we employ an attention mechanism based on CoT (Contextual Transformer) [30] to extract crucial features of ships in radar images. This mechanism fully utilizes the contextual information of the input data and enhances the expressive power of key features by learning a dynamic attention matrix. In comparison to traditional self-attention mechanisms, the CoT provides a more comprehensive treatment of contextual information. Traditional self-attention mechanisms only interact information in the spatial domain and independently learn correlation information, thereby overlooking rich contextual information among adjacent features and limiting the self-attention learning capability of feature maps. In terms of specific implementation, the CoT module integrates two types of contextual information about the image. Firstly, the input data is encoded through convolutional operations to capture contextual information and generate static contextual representations. Next, dynamic contextual representations are obtained through concatenation and consecutive convolutions. Finally, the static and dynamic contextual representations are fused to form the final output.

In the context of marine radar images, the CoT module effectively captures global and local contextual features of targets by encoding the input feature maps with contextual information. By leveraging contextual information, the network is able to extract more diverse and meaningful features, thereby obtaining additional information about the background surrounding the ships. Furthermore, CoT adaptively allocates attention weights to different features, enabling the model to flexibly acquire the most relevant and salient ship characteristics from the feature maps. Moreover, CoT exhibits the capability to directly substitute the standard 3×3 convolutional structure, facilitating its seamless integration into other classical convolutional networks.

The convolutional process of the CoT module, as depicted in Figure 7, operates on the input feature map $X \in R^{H \times W \times C}$, with keys ($K = X$), queries ($Q = X$), and values ($V = XW_v$) associated with the convolutional feature map. Unlike traditional self-attention methods, the CoT module employs $k \times k$ groups of convolutional operations to extract contextual information. Through this process, the resulting $K^1 \in R^{H \times W \times C}$ is further utilized as the static contextual representation of the input X.

Next, the previously obtained K^1 is concatenated with Q and passed through two consecutive convolutional operations to compute the attention matrix A.

$$A = [K^1, Q]W_\theta W_\delta \quad (5)$$

Based on the given attention matrix A, the enhanced feature K^2 is generated and computed according to the following formula. By leveraging the enhanced feature K^2, it becomes capable of capturing dynamic contextual representations regarding the input X. Ultimately, the fusion of these two contextual representations is achieved through a classical attention mechanism, resulting in the final output.

$$K^2 = V \odot A \quad (6)$$
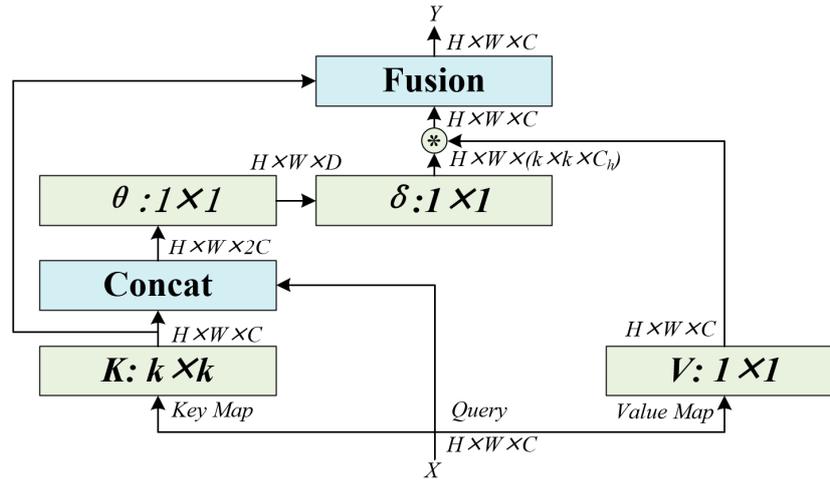
Figure 7. Convolution fusion process of COT module.

3.3. Ship prediction module

As is well-known, backpropagation is a crucial step in training neural networks, as it computes the gradients of the loss function with respect to the network parameters, enabling parameter updates and optimization. In the ship segmentation algorithm presented in this paper, the overall loss is comprised of the summation of position loss, classification loss, and confidence loss. Among them, the position loss plays a key role in guiding the network to learn and adjust the positions of the predicted boxes, thereby improving the accuracy and precision of ship localization. Therefore, we introduce a novel Intersection over Union (IoU) evaluation method, referred to as EIoU, as the computation standard for the target's position loss, aiming to further enhance the localization accuracy of ship bounding box predictions. The loss function takes into account multiple factors such as IoU loss, distance loss, and width-height loss, fully considering the relationship between the predicted box and the ground truth box. This leads to a more stable gradient during the algorithm training. Moreover, the EIoU loss helps the model better understand the spatial distribution and relative positions of the targets, which is crucial for achieving high-precision ship segmentation. Specifically, the EIoU loss is defined as follows:

$$Loss_{EIoU} = 1 - (IoU - \frac{\rho^2(b^{gt}, b)}{c^2} - \frac{\rho^2(h^{gt}, h)}{c_h^2} - \frac{\rho^2(w^{gt}, w)}{c_w^2}) \quad (7)$$

$$IoU = \frac{S^{pr} \bigcap S^{gt}}{S^{pr} \bigcup S^{gt}} \quad (8)$$

Wherein, c represents the Euclidean distance of the minimum bounding rectangle diagonals between the predicted box and ground truth box; $\rho(b^{gt}, b)$ denotes the Euclidean distance between the center points of the two boxes(i.e. ground truth box and predicted box); $\rho(h^{gt}, h)$ represents the difference in length between the two boxes; $\rho(w^{gt}, w)$ represents the difference in width between the two boxes; c_h and c_w are the height and width of the minimum bounding rectangle, respectively;

S^{gt} represents the area of the ground truth box; S^{pr} represents the area of the predicted box, and other key indicators are illustrated as shown in Figure 8.
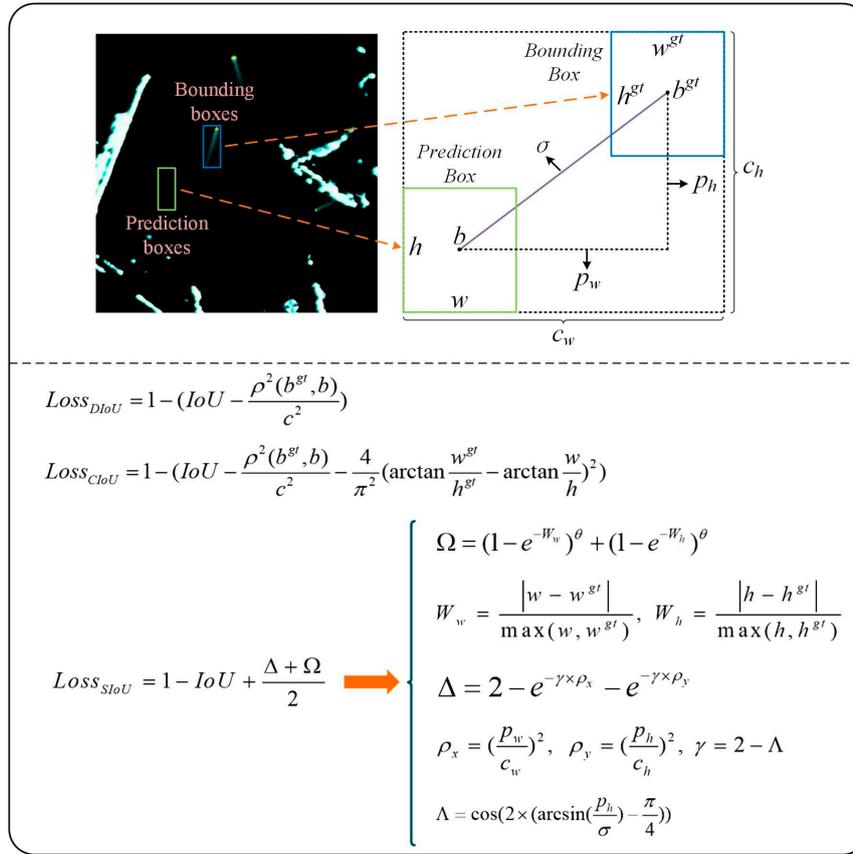

Figure 8. Key indicators of EIoU and calculation comparison of different loss functions.

In this section, we compared the EIoU loss with the classical DIoU, CIoU [31], and SIoU [32] and demonstrated their computational processes (refer to Figure 8). Specifically, unlike the IoU metric that solely focuses on the overlapping region, DIoU considers both the distance and the area overlap between the ground truth box and predicted box, thereby enhancing the regression stability of the predicted boxes. CIoU not only takes into account the positional information and localization errors of the predicted box but also introduces shape information to more accurately measure the target accuracy. However, when the aspect ratios of the two boxes are small, there may be issues of gradient explosion or vanishing. Building upon CIoU, SIoU further considers the angle of the regression vector between the two boxes by defining an angle penalty vector, which encourages the predicted box to quickly converge to the nearest horizontal or vertical axis. However, SIoU defines multiple different IoU thresholds, which present challenges in the practical training and evaluation of the model. In comparison, the EIoU function exhibits stronger interpretability in capturing the variations of loss during the regression of the predicted box. This leads to more accurate measurement of position, shape, distance, and width-height losses for small-scale ship predicted boxes in radar images, thereby facilitating the training and convergence of the algorithm.

4. A Case Study

4.1. Dataset

A high-quality dataset can significantly improve the empirical outcomes of CNN-based algorithms. Typically, the performance of the model is heavily reliant on the quality and quantity of the training data. A good dataset should encompass multiple scenarios and situations. This research conducted pre-processing on the real data obtained from the JMA5300 marine radar deployed at

Zhoushan Port in Zhejiang, resulting in the creation of a high-quality marine radar image dataset named RadarSeg. The dataset comprises 1280 images. As shown in Figure 9, ships in the images are mainly classified into two categories, i.e., long-wake ships and short-wake ships. To be specific, long-wake ships have distinct features that are easy to extract, while the pixel features of short-wake ships are similar to those of interferences such as reefs, which can interfere with the instance segmentation of ships. Moreover, the RadarSeg dataset covers complex background environments, including different weather conditions, harbor environments, and imaging conditions. Additionally, the dataset also takes into account factors such as the variations in ship heading, and traffic flow. Besides, it also emphasizes an augmentation in the number of images depicting ships of small-scale and miniature sizes. Considering that different types of ships exhibit similar spot features in marine radar images, all ships in the dataset are uniformly labeled as "boat" category.

In the RadarSeg dataset, each image has been annotated with ship type labels and tightly fitting bounding boxes along the target edges. These annotations are saved in JSON format files and have been preprocessed to meet the requirements of the YOLO algorithms. We divided all the images into training, validation, and testing sets in an 8:1:1 ratio. Various algorithms were trained on the training and validation sets, and evaluation metrics and algorithm performance were assessed on the testing set. Additionally, cluster analysis revealed that the number of pixels occupied by ships accounts for approximately 0.035% of the entire image. Thus, in marine radar images, ships are predominantly small-scale or even miniature targets.

4.2. Training optimization methods

This research optimized the training process of the proposed algorithm by improving aspects such as learning rate decay, loss calculation, and data augmentation. Normally, traditional neural network-based algorithms usually employ a series of fixed learning rates for training, which may result in significant learning rate decay at different training stages, leading to unstable changes in model momentum and negatively affecting the algorithm's training effectiveness. To address this issue, the Adam optimizer was introduced to optimize the learning rate, which adaptively adjusts the learning rate without the need for manual tuning, based on the magnitude of parameter gradients. Additionally, label smoothing was introduced to enhance the generalization performance of ship segmentation model. This approach reduces the model's reliance on noise or uncertainty information in the training data. Furthermore, data augmentation methods, including random cropping and horizontal flipping, were employed to increase the diversity of training data and improve the model's robustness and generalization abilities.

4.3. Experimental environment and training results

The experimental setup relied on the Ubuntu 20.04 operating system and utilized the NVIDIA RTX3090 graphics card with an effective memory size of 24GB. The CUDA version used was 11.1.0, PyTorch version was 1.9.1, and the Python environment was 3.8. We compared the performance of different algorithms in the ship segmentation, including the YOLO series and its improved variants, classical two-stage algorithms, hierarchical algorithm based on the Transformer, and MrisNet. All algorithms were evaluated using the same dataset of ship images. During algorithm training, the input image size was set to 640x640 pixels, the momentum was set to 0.9, the batch size was set to 16, and the training was conducted for 800 epochs with an initial learning rate of 10^{-4}.
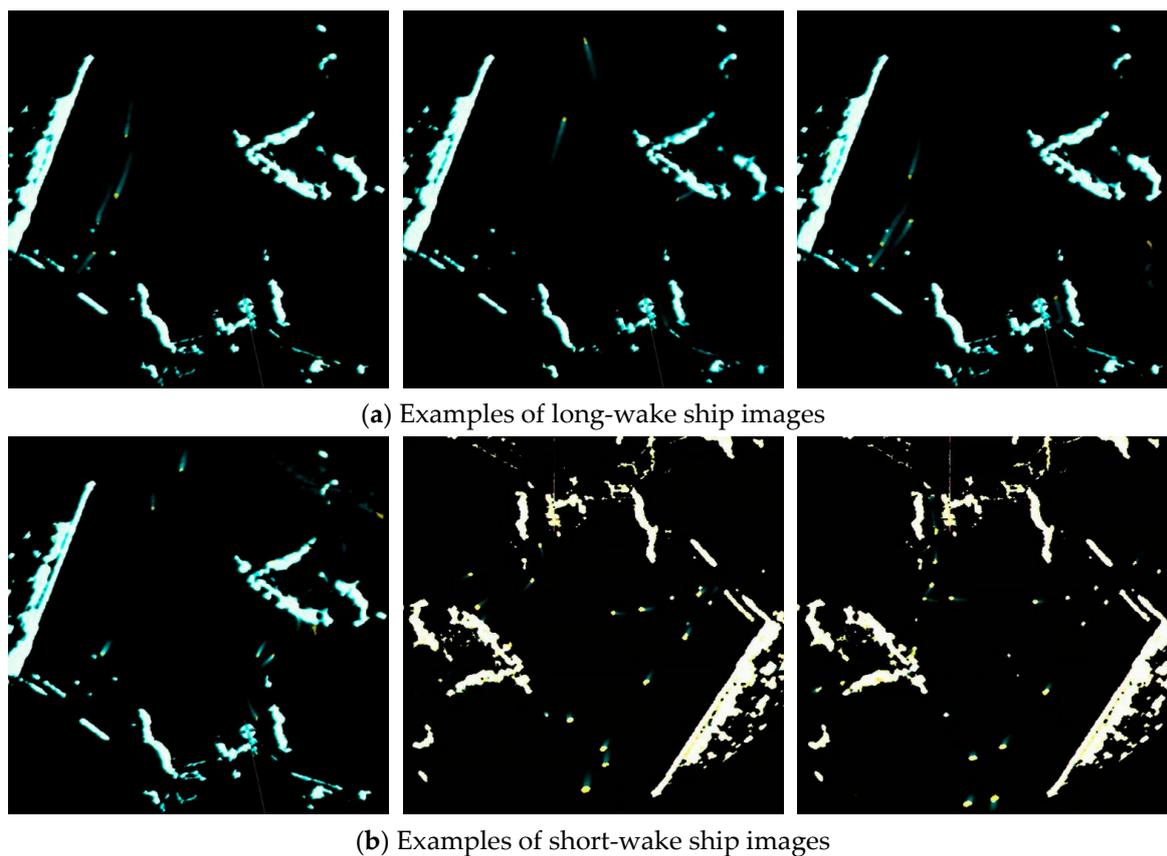
(**a**) Examples of long-wake ship images

(**b**) Examples of short-wake ship images

Figure 9. Marine radar images.

Under the aforementioned settings, the training of various algorithms and related experiments were carried out. Among them, YOLOv5(S), YOLOv8(S) [33], and MrisNet achieved convergence of the loss values on the training set to 14.91×10^{-3}, 15.09×10^{-3}, and 14.52×10^{-3}, respectively. The experimental results indicate that, when evaluated using the same metrics, MrisNet exhibits lower convergence loss compared to the standard YOLO algorithms. This illustrates that MrisNet is capable of more accurately identifying ship pixel-level features in radar images. Furthermore, through the analysis of Figure 10, it can be concluded that under different threshold conditions, MrisNet has achieved favorable experimental results across various evaluation metrics, further substantiating its ability to minimize misjudgments and omissions in ship segmentation while accurately locating and identifying targets.
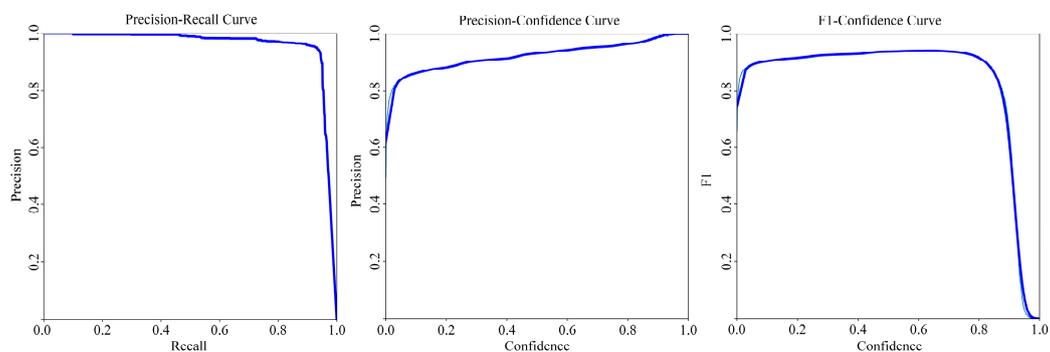

Figure 10. Various verification results of MrisNet.

4.4. Comparisons and Discussions

This research selected Recall and Precision as metrics to measure the effectiveness of ship segmentation and object detection. Additionally, we evaluated the convolution parameter count and computational cost using parameters (PARAMs) and floating-point operations (FLOPs). Furthermore, five sets of experiments were designed to assess the practical performance of MrisNet. Firstly, comparative experiments were designed to assess the actual performance of various algorithms across different evaluation metrics, thereby validating the effectiveness of the MrisNet. Moreover, ablation experiments were conducted to analyze the individual improvements in MrisNet and verify the specific effects of different methods. Finally, the adaptability of the MrisNet for ship segmentation in marine radar images was examined through the identification of different categories of radar images.

4.4.1. Experimental analysis of different algorithms

On the constructed RadarSeg dataset, a comparison was conducted between several commonly used standard algorithms and the proposed MrisNet. All the algorithms were trained using the same hyperparameters and tested on the same dataset. Moreover, each algorithm underwent cross-testing, and the Recall, and Precision were averaged over three experimental trials. As shown in Table 1, for the ship segmentation, our proposed method achieved a recall rate of 94.8% and a precision of 95.2% on the testing images. Compared to other algorithms, our method demonstrated satisfactory experimental results. This is attributed to the MrisNet's enhanced capability to accurately identify densely navigated and small-scale ships in radar images. Furthermore, MrisNet demonstrated remarkable performance in the object detection, achieving a recall rate of 98% and precision of 98.6%. This highlights the beneficial ship localization capability of our proposed method. Additionally, MrisNet showcased desirable performance in terms of parameter count and real-time computational cost, with values of 13.8M and 23.5G, respectively. Compared to the standard YOLOv7 [34], MrisNet achieved a reduction of 61.77% and 83.44% in model parameters and computational cost, respectively. These results indicate that MrisNet is more suitable for deployment on edge computing devices.

The Swin Transformer [35], a novel recognition algorithm based on the evolution of the Transformer architecture, achieves a recall of 94% and precision of 93.7% in ship segmentation. Compared to MrisNet, it experiences a decrease of 0.8% in recall and 1.5% in precision. The analysis suggests that due to the limited pixel information of ships in the images and the presence of significant background noise, the fixed-size image patches used in Swin Transformer may struggle to effectively capture and represent targets of different scales. As a result, the algorithm illustrates relatively underwhelming performance in recognizing different types of ships and fails to extract detailed ship features accurately, making it difficult to distinguish ships from interfering objects.

To evaluate the practical performance of MrisNet, multiple single-stage, two-stage, and contour-based instance segmentation algorithms were employed for comparison. The experimental results demonstrate that MrisNet outperforms single-stage algorithms such as YOLACT [36] and SOLOv2 [37] in terms of core evaluation metrics. This indicates that MrisNet, through its adaptive design in network architecture, loss function, and feature enhancement, enables more precise extraction of ship features from radar images, resulting in accurate localization and segmentation of ship instances. Due to the introduction of an additional segmentation branch outside the object detection framework, Mask R-CNN achieves relatively robust precision and segmentation quality. Experimental data reveals that for ship segmentation, this algorithm surpasses 93% in both accuracy and recall rate, which is very close to the experimental results of MrisNet. However, due to its high parameter count and computational demands, it exceeds the computational capacity of most maritime devices, making it unsuitable for direct deployment on such devices. Additionally, Deepsnake [38], combining deep learning with the Snake algorithm in active contour models, enables real-time instance segmentation in various common scenarios. However, in the marine radar images, Deepsnake achieves a segmentation precision of only 91.7%, significantly lower than the proposed method. The analysis suggests that the fine and small ship contours in radar images do not provide

sufficient information for accurate learning of vertex offsets, thereby affecting the segmentation accuracy and recall of targets.

Table 1. Specific experimental results of various algorithms.

Different types	Algorithms	P^{box}	R^{box}	P^{mask}	R^{mask}	PARAMs/(M)	GFLOPs
Single-stage algorithms	SOLOv2	0.897	0.871	0.882	0.855	61.3	232.6
	YOLACT	0.901	0.86	0.875	0.826	53.72	240.2
Two-stage algorithm	Mask R-CNN	0.97	0.975	0.938	0.942	62.74	244.8
Swin-Transformer	Swin-Transformer(T)	0.979	0.976	0.937	0.94	88	745
Deepsnake	Deepsnake	0.939	0.94	0.917	0.919	16.37	25.94
Standard YOLO series	YOLOv5(N)	0.957	0.961	0.913	0.908	1.8	6.7
	YOLOv5(S)	0.962	0.965	0.932	0.933	7.1	25.7
	YOLOv5(M)	0.966	0.968	0.923	0.924	20.65	69.8
	YOLOv5(L)	0.965	0.971	0.924	0.925	45.27	146.4
	YOLOv5(X)	0.967	0.971	0.925	0.924	84.2	264
	YOLOv7	0.968	0.965	0.919	0.917	36.1	141.9
	YOLOv8(S)	0.961	0.909	0.917	0.851	11.23	42.4
	YOLOv8(M)	0.967	0.956	0.923	0.914	25.96	110
	YOLOv8(L)	0.958	0.973	0.92	0.925	43.79	220.1
	YOLOv8(X)	0.969	0.964	0.925	0.922	68.4	343.7
Proposed algorithms	Mris_APFN	0.966	0.965	0.92	0.923	10.3	20.6
	MrisNet	0.986	0.98	0.952	0.948	13.8	23.5

Compared to the standard YOLOv5, YOLOv7, and YOLOv8 series, MrisNet exhibits relatively robust segmentation accuracy and recall rate, surpassing various YOLO algorithms, including the relatively high-performing YOLOv5(S). Moreover, MrisNet exhibits significantly lower parameter count and computational costs compared to many deep-layer algorithms in the YOLO series. Additionally, in terms of recall rate, MrisNet increases 9.7% and 1.5% compared to the lightweight YOLOv8(S) and YOLOv5(S), respectively. This indicates that MrisNet experiences fewer instances of ship loss in marine radar images and shows improved efficiency throughout the experiments. Furthermore, the proposed method demonstrates noteworthy capability in suppressing false positives, accurately identifying interferences that resemble ship features, such as coastal objects, reefs, and clouds. This effectively reduces misidentification rates and enhances the precision of ship segmentation in various scenarios.

To investigate the influence of different levels of image features on ship segmentation, we combined a progressive feature pyramid with the MrisNet. This network integrated multi-scale features through a layer-by-layer concatenation approach, preserving deep-level semantic information more effectively compared to standard PANet. In the process of algorithm improvement, the feature network and prediction structure of MrisNet were retained, and the original PANet was replaced with the AsymptoticFPN [39] to construct the Mris_APFN algorithm for comparison. The experimental results demonstrate that Mris_APFN achieved a recall rate of 92.3% and a precision of 92%, but it falls short of MrisNet in all evaluation metrics. This indicates that Mris_APFN suffers from more instances of missing ships and fails to extract features relevant to ship instances effectively. Further analysis suggests that deep-level features tend to focus more on abstract semantic information while being relatively weaker in extracting contour details. Therefore, for ship segmentation in radar images, it is necessary to appropriately restrict the influence weight of deep-level features to better express salient features of the targets.

4.4.2. Ablation Experiments

To further validate the practical performance of each improvement method in MrisNet, a comprehensive decomposition analysis was conducted based on the RadarSeg dataset to analyze

their impact on ship segmentation. The main experimental process involved step-by-step application of various improvement methods on the standard YOLOv5(S), followed by testing their respective performance metrics. The specific results of the ablation experiments for MrisNet are presented in Table 2.

(1) Analysis of the FasterYOLO network. By replacing the feature network with FasterYOLO, experimental results reveal that the improved feature network increased the ship segmentation precision by 1.0%, with an improvement in recall as well. This indicates that the algorithm's ability to suppress false targets has been enhanced, reducing the misidentification rate and decreasing the probability of target omissions. Furthermore, through multiple experiments, it has been observed that FasterYOLO could accelerate the convergence speed of the algorithm, leading to relatively rapid and stable convergence of the loss values based on the training and validation sets.

(2) Analysis of SimAM. In this paper, a SimAM attention mechanism was appended after the FasterYOLO feature network, enabling the algorithm to weight the convolutional feature maps based on the principle of similarity. This allows the model to focus more on the relevant feature representations associated with the target's spatial location. As a result, the model exhibits enhanced robustness when faced with challenges such as low illumination and occlusion in challenging scenarios. Based on the experimental results, the SimAM module enhances the salient features of ships, resulting in improvements in both precision and recall rates of the ship segmentation model.

(3) Analysis of the CPFPN network. As mentioned earlier, applying attention mechanisms based on Transformer structures to feature fusion network enables better modeling of global contextual information in images. According to Table 2, it is evident that compared to the standard PANet network, CPFPN improves the recall and precision of ship segmentation by 0.5% and 0.3% respectively, leading to a relatively clear improvement in the overall ship recognition capability. However, this improvement comes at the cost of a moderate increase in model parameters. In theory, due to the presence of various interfering factors in radar images, the complexity and diversity of objects such as islands and reefs make it challenging to distinguish ships from the background. The CPFPN network, by adaptively attending to target features and effectively filtering out noise or redundant objects, plays a crucial role in the instance segmentation of small-scale ships.

(4) Analysis of EIoU loss. In neural network-based algorithms, the position loss function aids the model in better accommodating variations in the position and scale of the targets. Experimental results indicate that the EIoU loss improves ship segmentation precision by approximately 0.6%. To further validate the effectiveness of EIoU, the SIoU and DIoU functions were also compared in this section. Theoretically, these two types of loss functions are also capable of accelerating model convergence. It can be observed that the model employing the aforementioned comparative functions achieved somewhat favorable results, yet lower than EIoU in terms of evaluation metrics. Analysis reveals that the position loss function, which takes into account spatial and shape factors, exhibits relatively better performance in ship segmentation under radar images.

Table 2. Ablation experiments of MrisNet.

| Methods	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
YOLOv5(S)	★	★	★	★	★	★	★
+FasterYOLO		★	★	★	★	★	★
+SimAM			★	★	★	★	★
+CPFPN				★	★	★	★
+SIoU					★		
+DIoU						★	
+EIoU							★
R^{mask}	0.933	0.937	0.94	0.945	0.944	0.94	0.948
P^{mask}	0.932	0.942	0.943	0.946	0.941	0.938	0.952

4.4.3. Comparisons in radar images

We presented in Figure 11 the ship segmentation results in marine radar images under different scenarios using MrisNet, with a specific focus on evaluating its performance on ships with various trail features. From Figure 11a, it can be observed that MrisNet accurately segments ships with long trails in different water areas. This demonstrates the algorithm's ability to recognize ships with prominent features, as these targets maintain a certain consistency in position and shape across consecutive frames. In radar images, it is noticeable that longer ship trails often exhibit trajectory interruptions or significant curvature, posing a significant challenge to the algorithm's adaptability. However, experimental results show that MrisNet effectively avoids the degradation of segmentation precision caused by these two scenarios. Figure 11b reveals that MrisNet achieves satisfactory segmentation results for short-trail ships under different backgrounds, without missing or misidentifying small-scale or even tiny-scale targets. This indicates the algorithm's relatively positive capability in extracting fine-grained object features and mitigating the influence of terrain, sea conditions, and clutter interference. Figure 11c demonstrates that MrisNet accurately segments dense ships in radar images, and exhibits good recognition capabilities for complex scenarios such as head-on or crossing trajectories. Analysis suggests that the adoption of adaptive attention mechanisms and efficient convolutional computations in MrisNet enables it to capture salient ship features even in scenarios with dense small targets.
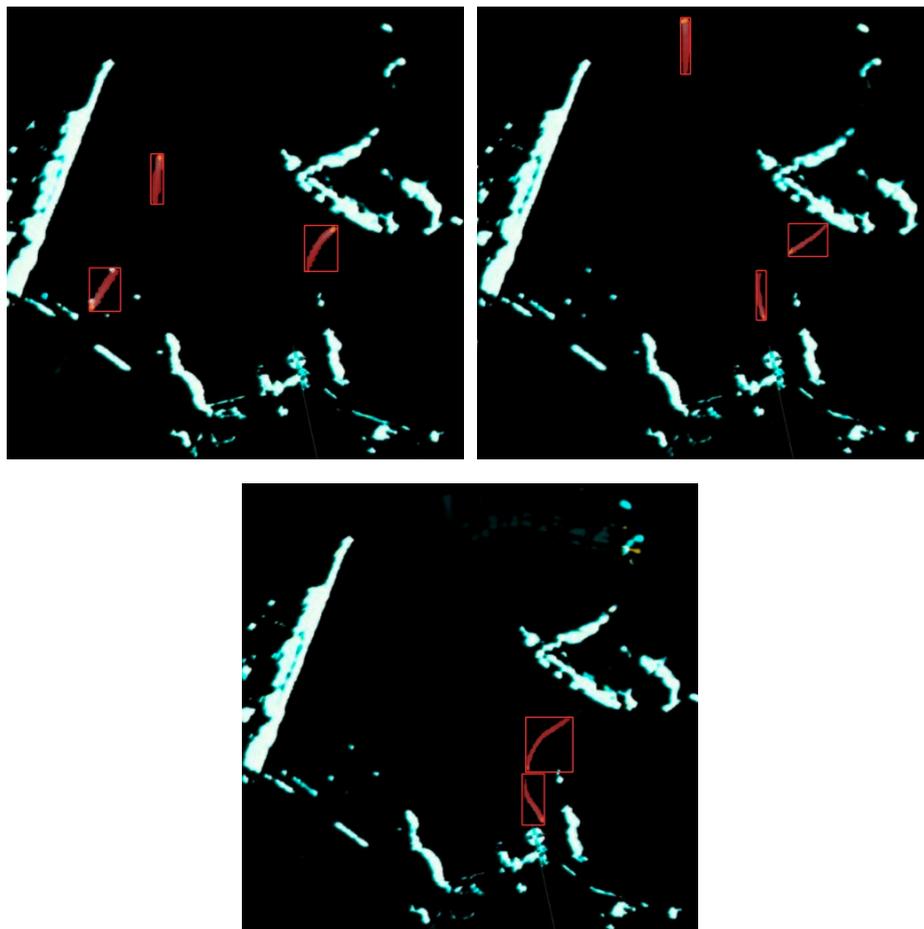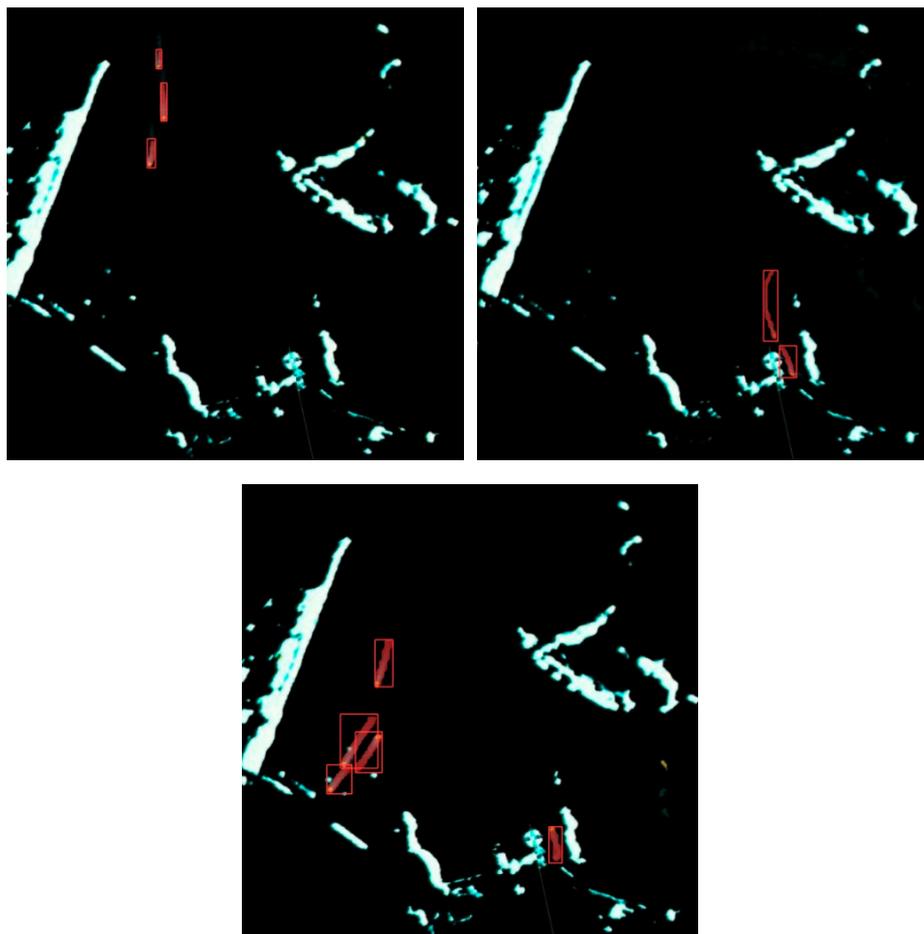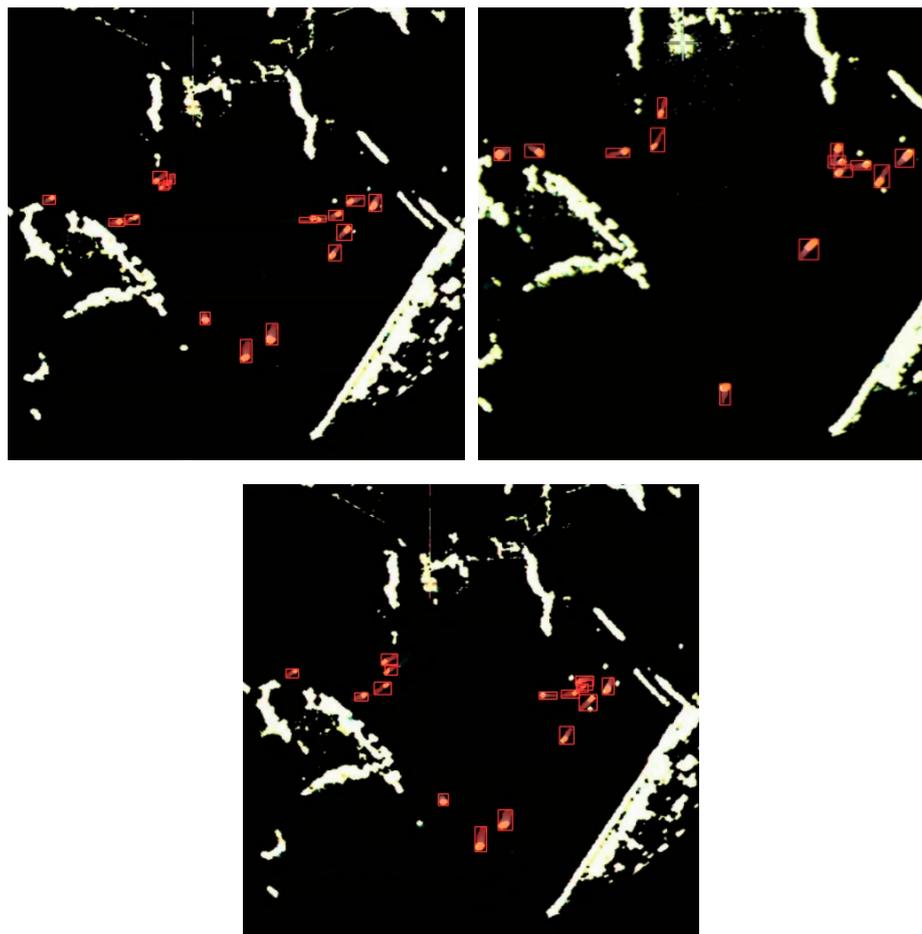
(a) Segmentation of long-wake ships
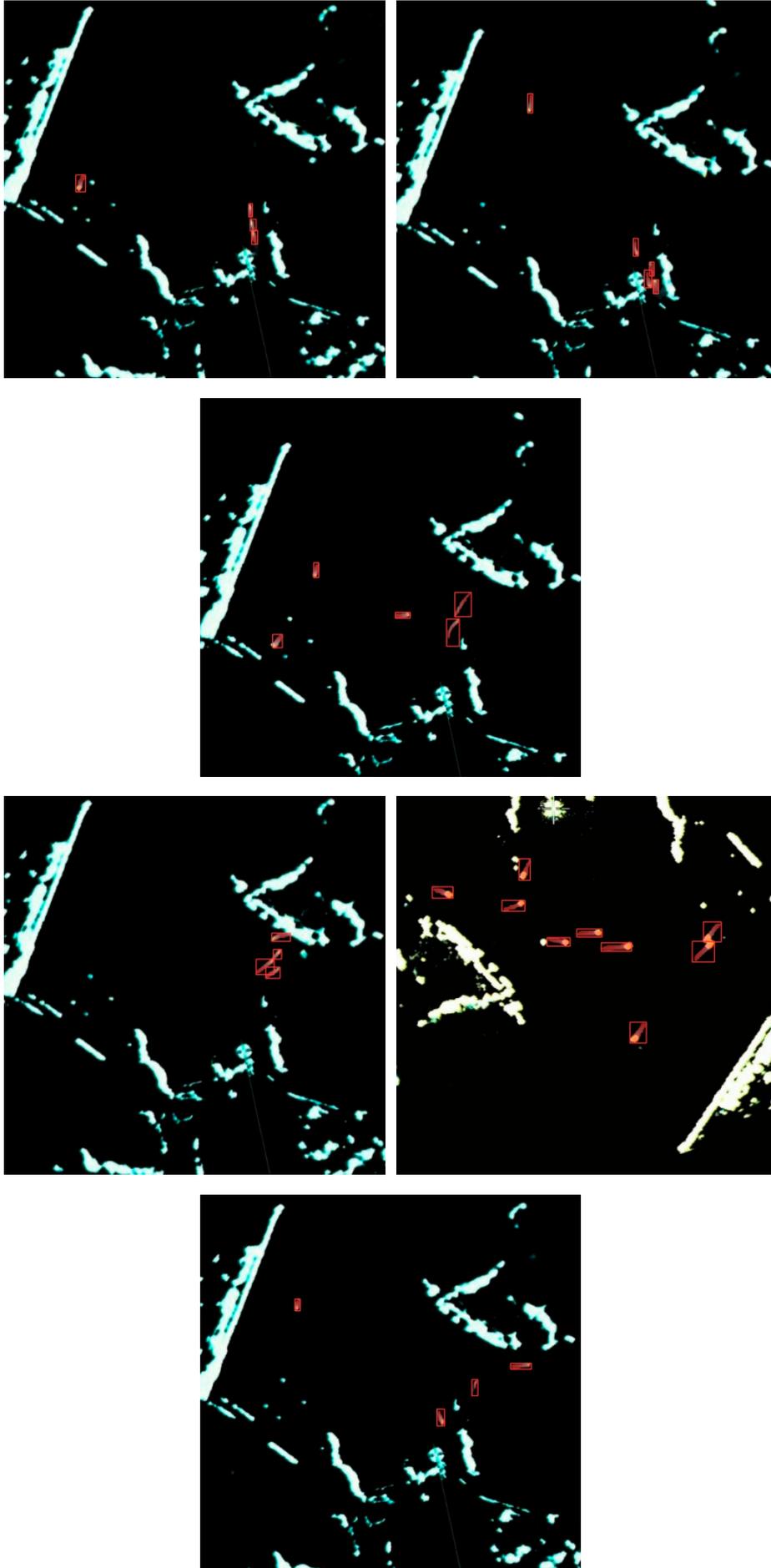
(b) Segmentation of short-wake ships
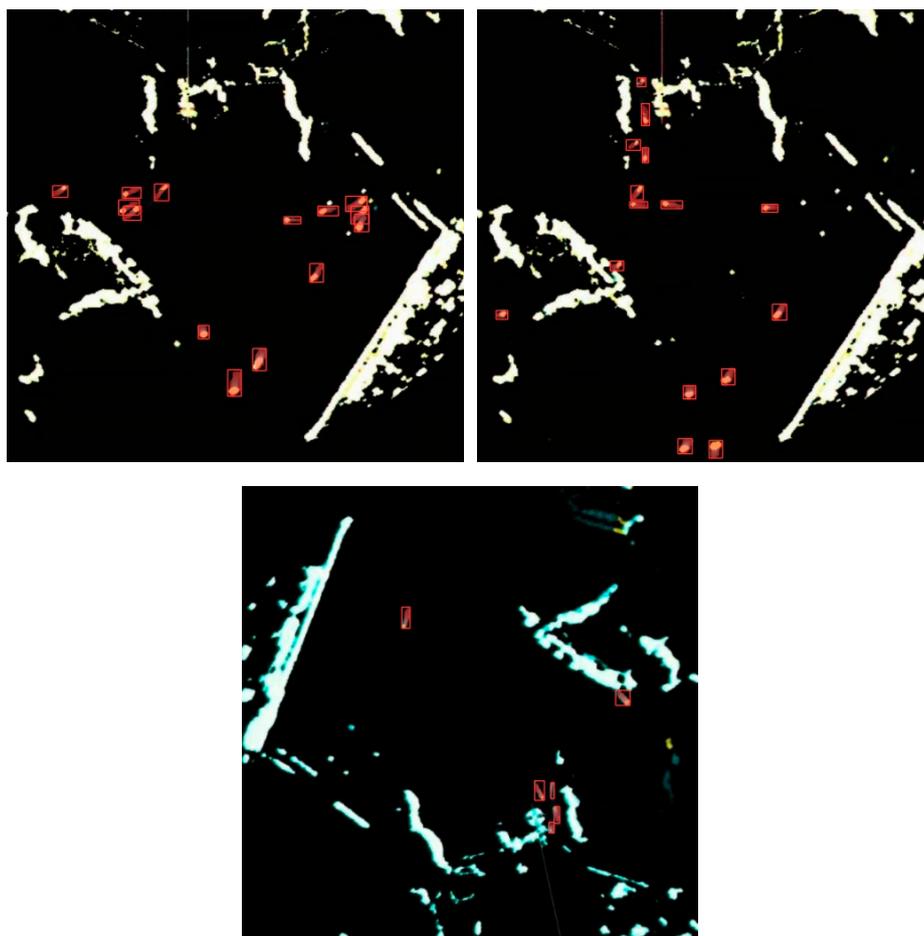
(c) Segmentation of dense ships

Figure 11. Segmentation results of MrisNet under marine radar images.

4.4.4. Comparisons of small-scale ship segmentation

Instance segmentations of small objects has been a prominent research direction in the field of visual recognition tasks. To evaluate the performance of different algorithms in the segmentation of small-scale ships, this section conducted experiments in two scenarios, i.e., cross navigation and dense navigation. These scenarios involve a large number of ships with short trails, and factors such as islands, reefs, and atmospheric disturbances significantly affect the model's performance. To enhance the credibility of the findings, this experiment compared the performance of MrisNet with the standard YOLOv5(L) and YOLOv8(L). As shown in Figure 12, YOLOv8(L) exhibited a higher misidentification rate but fewer omissions in recognition. YOLOv5(L) demonstrated relatively poor accuracy in recognizing small-scale ships, leading to more omission issues and a tendency to misidentify ships traveling in opposite directions. In comparison to the comparative algorithms, MrisNet achieved more accurate localization of small-scale ships, enabling finer segmentation of ship contours and demonstrating favorable segmentation accuracy for tiny objects. This result highlights that lightweight algorithm, through the construction of a rational network structure, can effectively extract ship pixels and contour features from radar images, thereby significantly improving the performance of the model in small-scale instance segmentation tasks.
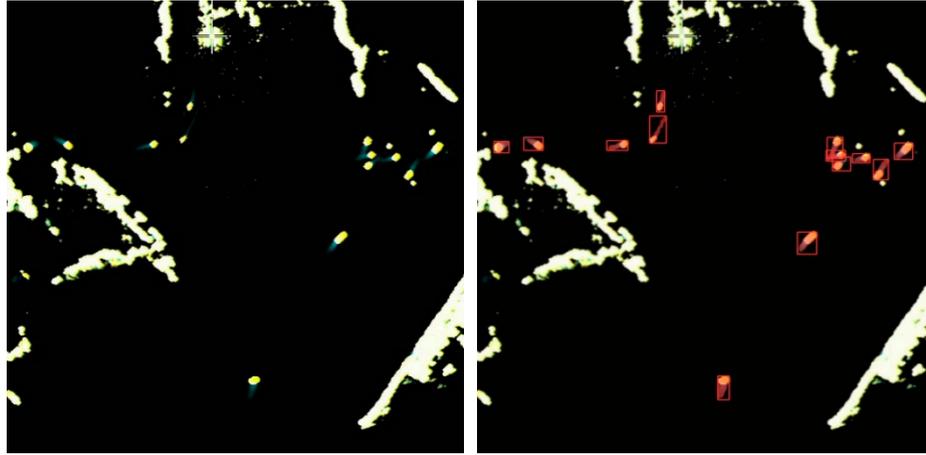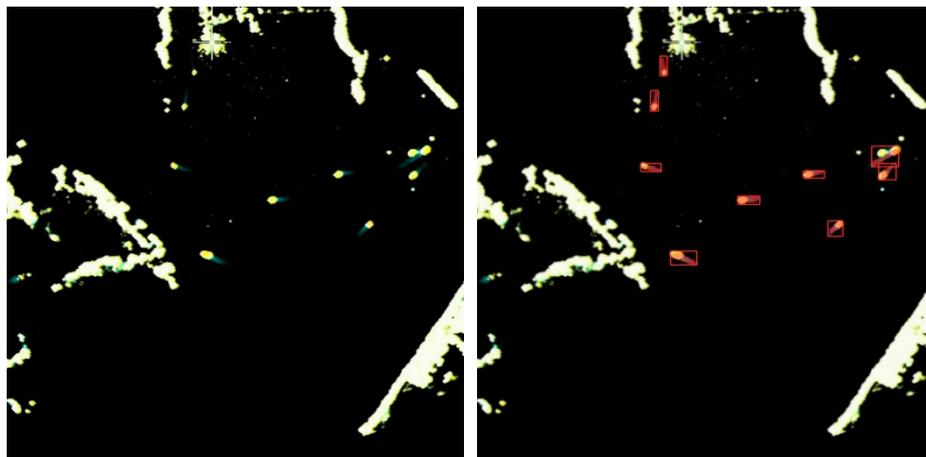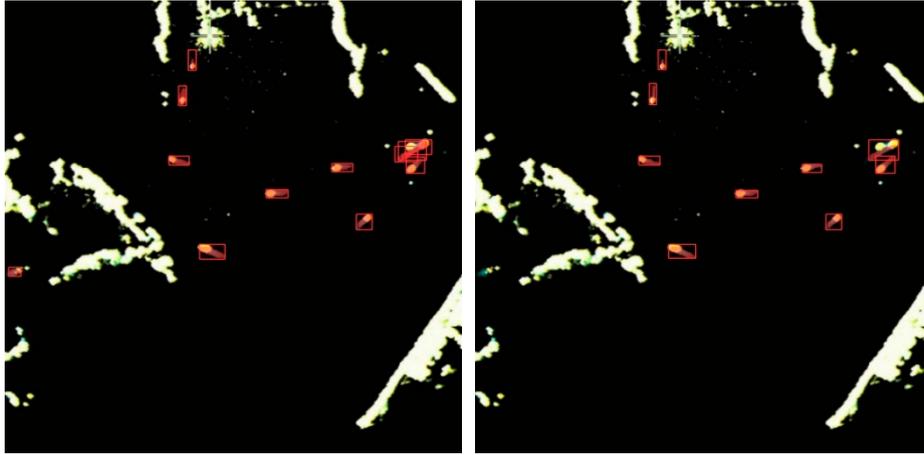
Original image   MrisNet
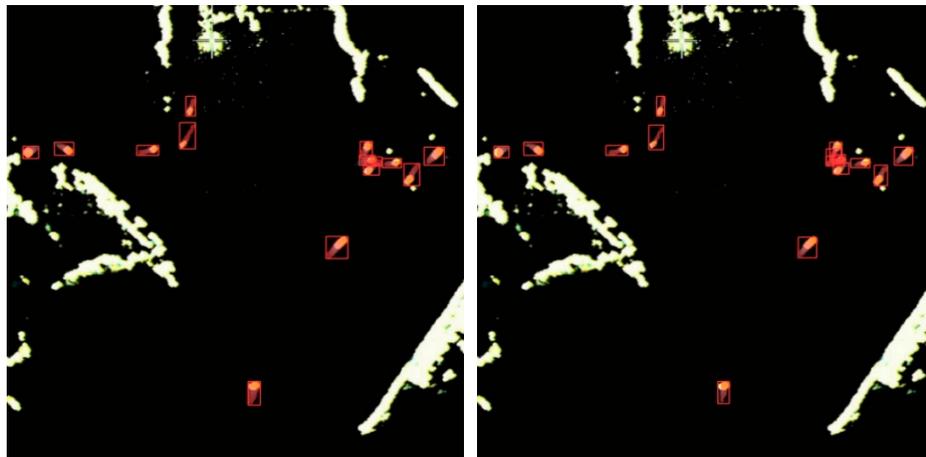
YOLOv8 YOLOv5

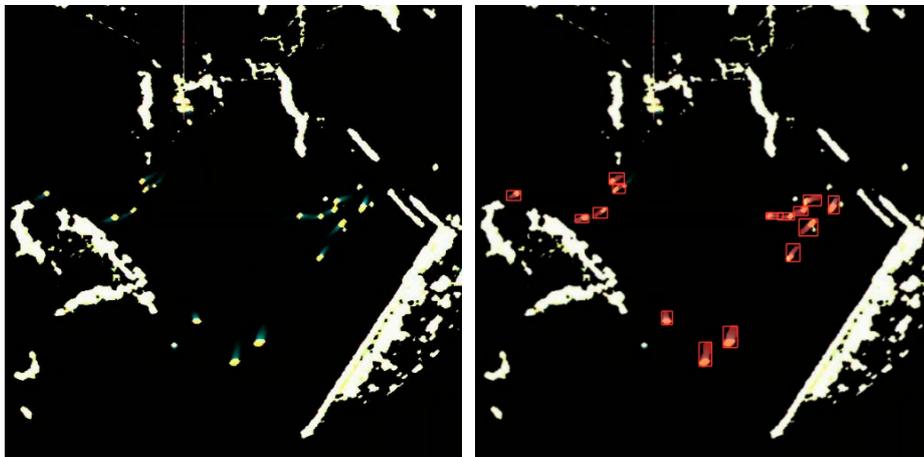(a) Segmentation comparison in cross navigation scenario
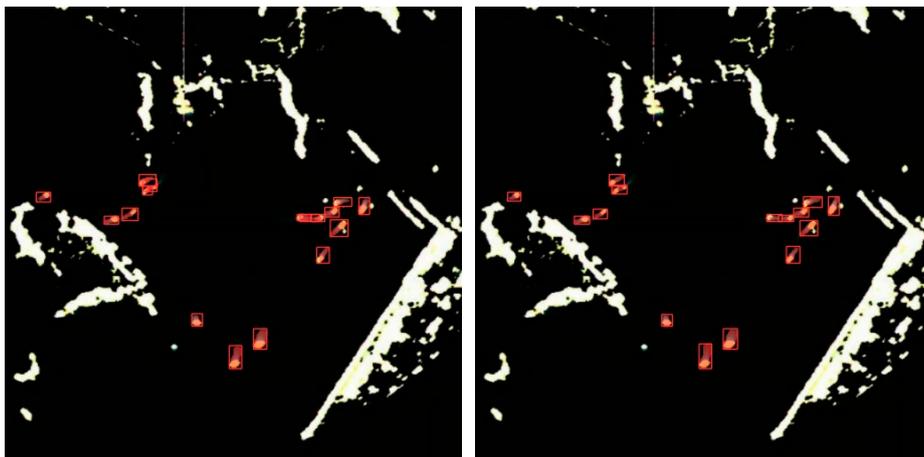

Original image   MrisNet

YOLOv8 YOLOv5

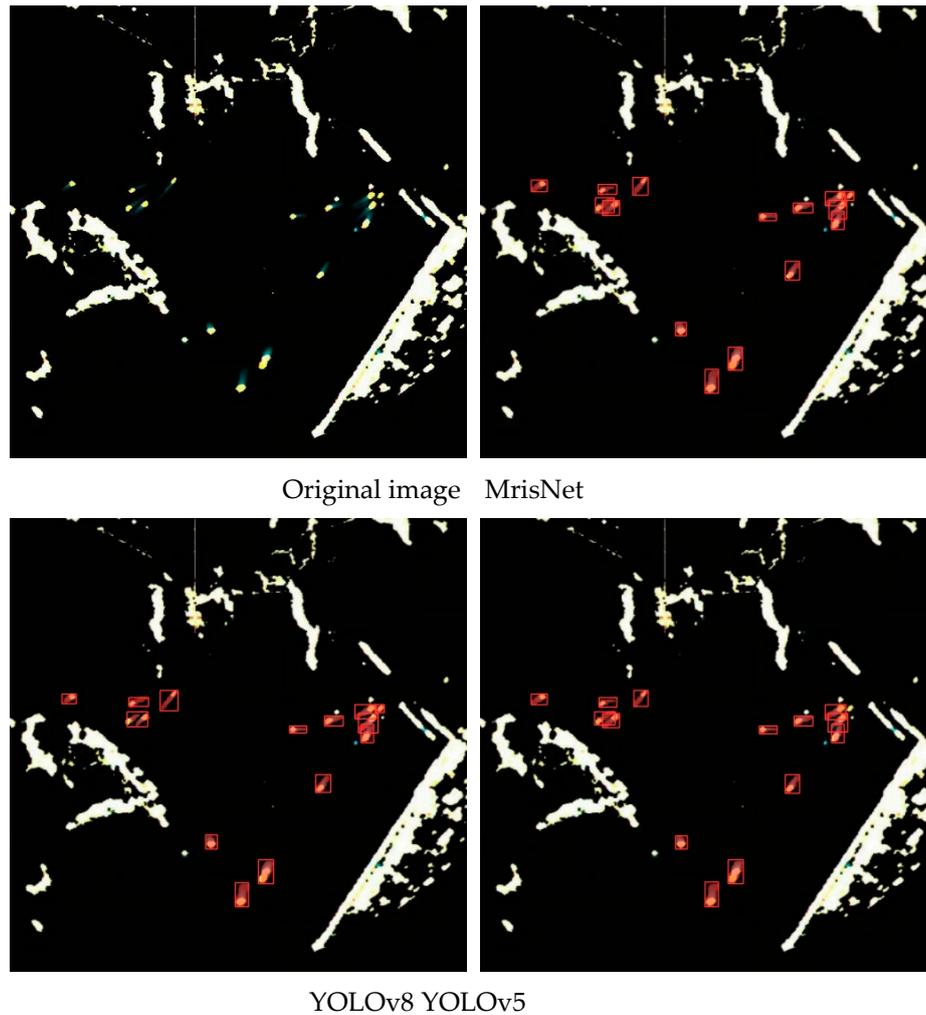(b) Segmentation comparison in interference scenario

Original image   MrisNet

YOLOv8 YOLOv5

(c) Segmentation comparison in dense navigation scenario

Original image   MrisNet

YOLOv8 YOLOv5
(d) Segmentation comparison in parallel sailing scenario

Figure 12. Comparisons of various algorithms for small-scale ship segmentation.

4.4.5. Ship identification in extreme environments

To evaluate the capability of MrisNet in ship segmentation under extreme environments, a subset of extreme scenario images was constructed by selecting samples from the RadarSeg dataset. This subset consists of 200 images, categorized into two identification scenarios, i.e., dense ship and tiny ship identification. For comparison, YOLACT, YOLOv8(S), and Mask R-CNN were also tested on the same images and threshold settings. The experimental results, as shown in Table 3, indicate that MrisNet exhibited encouraging performance. It achieves relatively higher ship segmentation precision and fewer misidentification errors, and performs better in terms of recall rate. Therefore, MrisNet demonstrates satisfactory adaptability to ship segmentation in extreme scenarios.

Table 3. Comparisons of experimental data in extreme scenarios.

| Algorithms	Detected Ships	True Ships	False Alarms	Recall	Pr
YOLACT	1296	1117	179	0.8007	0.8619
YOLOv8(S)	1288	1170	118	0.8387	0.9084
Mask R-CNN	1389	1299	90	0.9312	0.9352
MRNet	1382	1301	81	0.9326	0.9414

Figure 13 showcases the instance segmentation of MrisNet on ship radar images under various types of interference signals. The experimental images were extracted from the aforementioned

subset of extreme scenario images, covering three common target types, i.e., short-trail ships, long-trail ships, and dense ships, which are prevalent in radar images. In this experiment, the original radar images were subjected to three types of noise processing, i.e., salt-and-pepper noise, gaussian noise, and speckle noise. The research found that speckle noise and gaussian noise had a more significant impact on radar images, causing noticeable interference with trail features. This led to a decrease in the positional precision of long-trail ships and the potential omission of identification for small-scale ships, as their trail features became confounded with the image background. In contrast, salt-and-pepper noise had a minor impact on the effective ship features. The experimental results demonstrated that even under the interference of salt-and-pepper noise, MrisNet was able to achieve precise segmentation of all ships. However, under the influence of speckle noise and Gaussian noise, the confidence of ship segmentation by MrisNet slightly decreased, and the positioning accuracy of ships in dense navigation scenarios also decreased. Nevertheless, in the majority of scenarios, MrisNet maintained an effective ship segmentation performance.
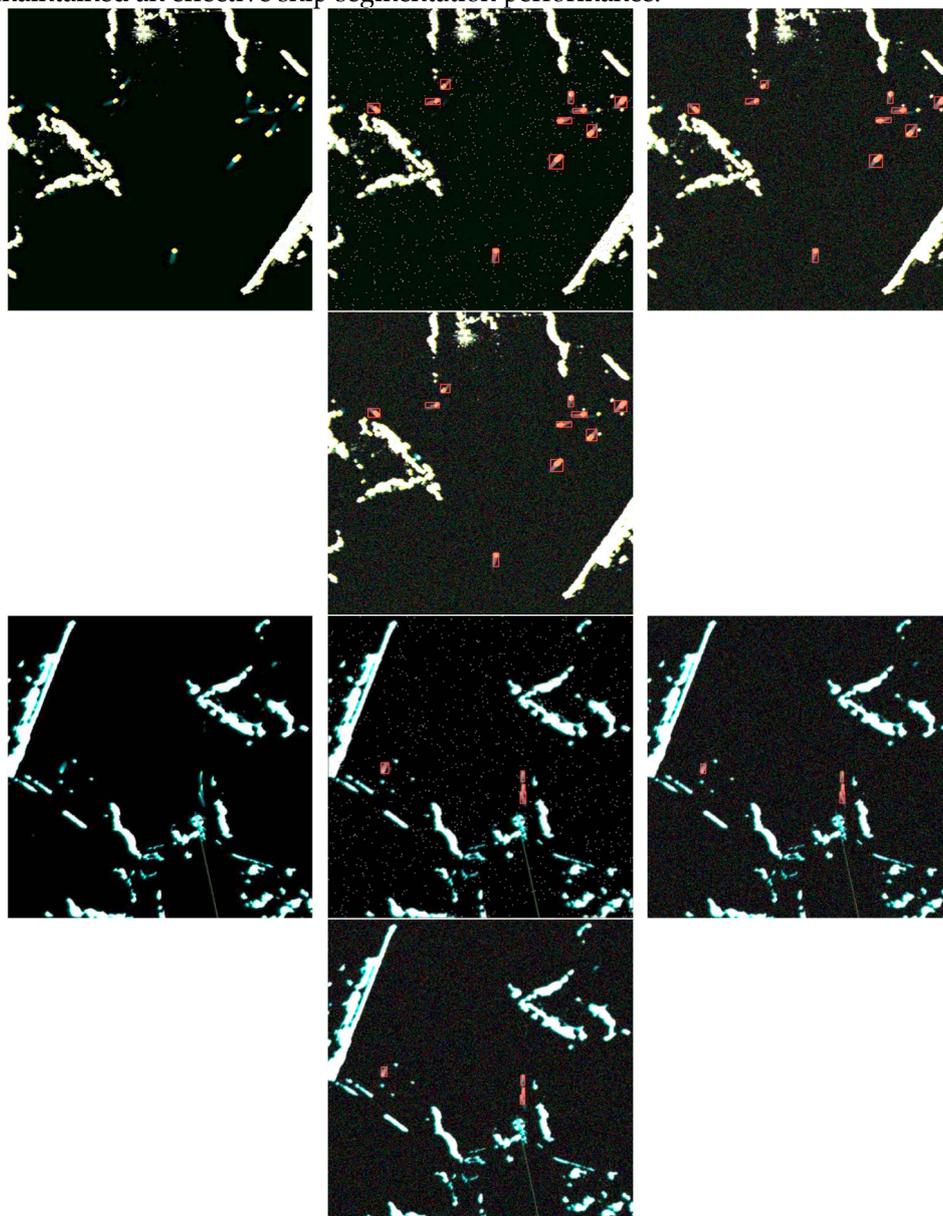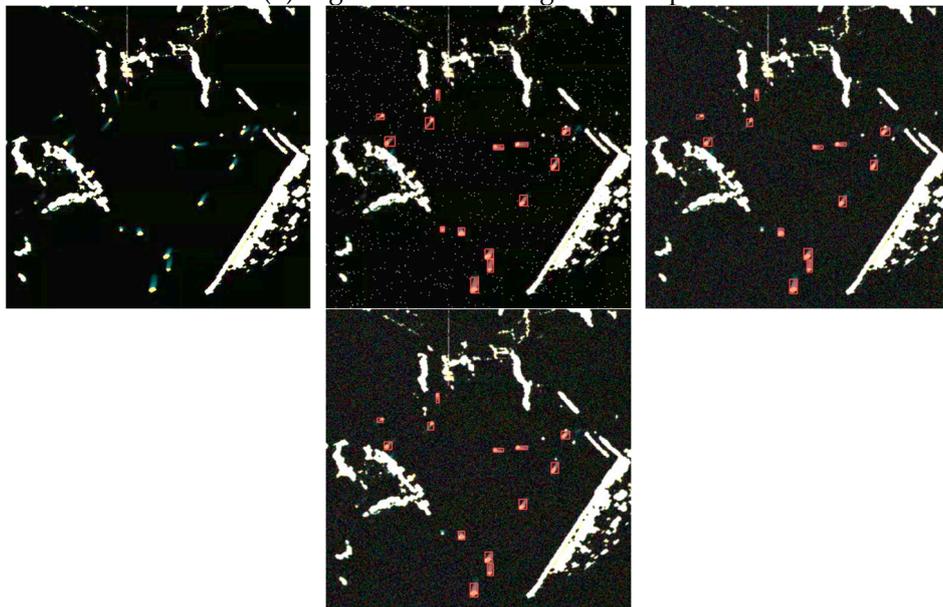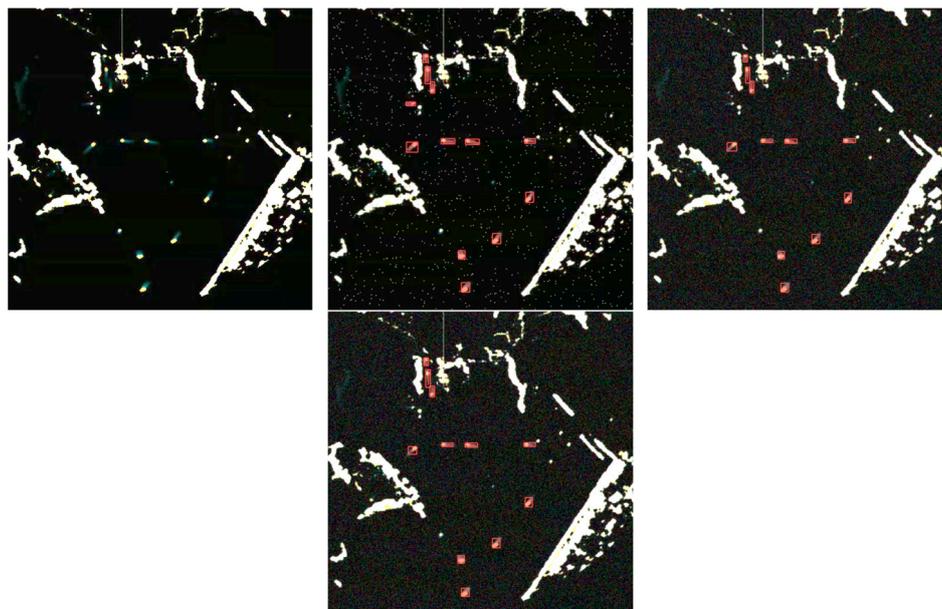

Original image Salt-and-Pepper noise Gaussian noise Speckle noise
(a) Segmentation of short-wake ships
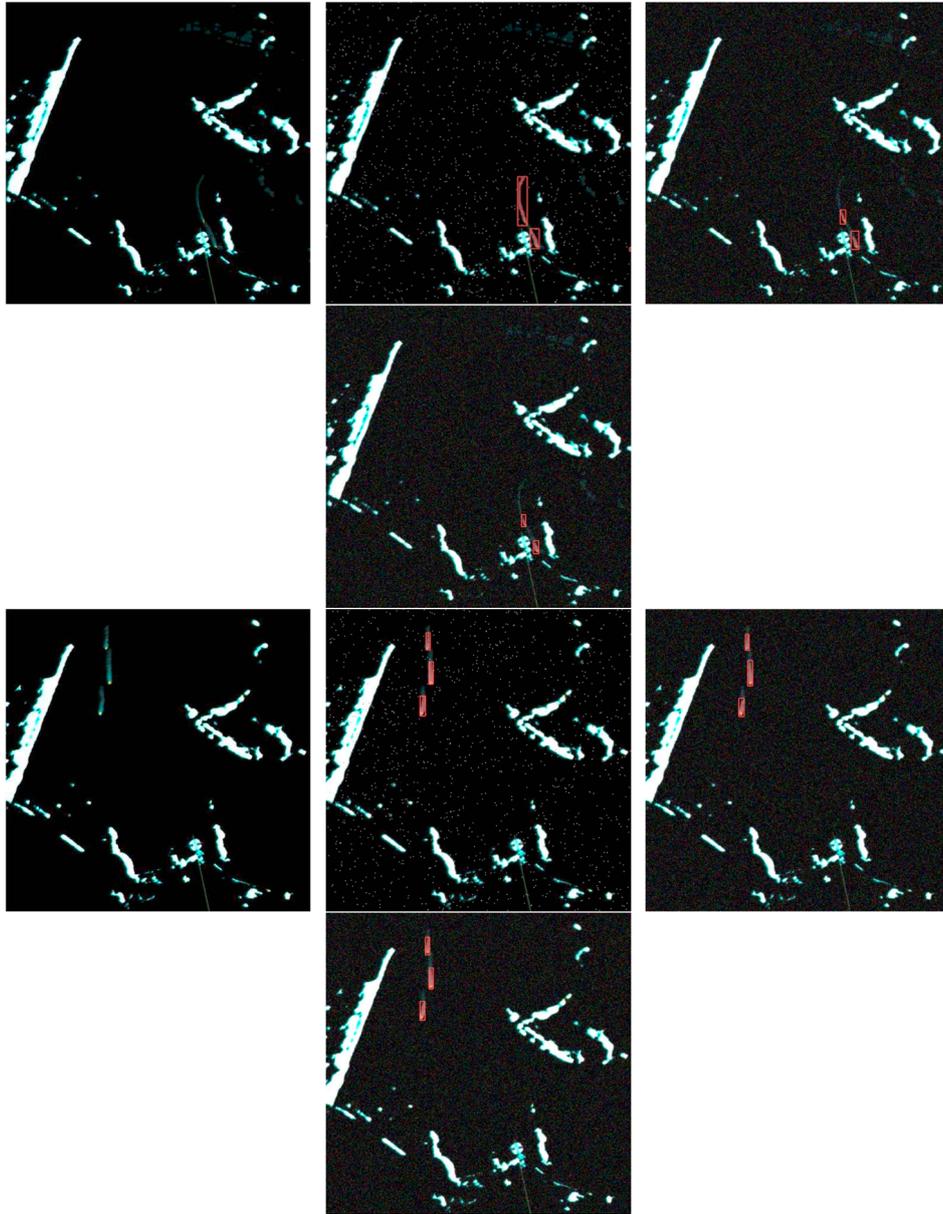
Original image Salt-and-Pepper noise Gaussian noise Speckle noise

(b) Segmentation of long-wake ships

Original image Salt-and-Pepper noise Gaussian noise Speckle noise
(c) Segmentation of dense ships

Figure 13. Ship segmentation results of MrisNet under various noises.

5. Conclusions and Discussions

This paper presents a ship segmentation algorithm called MrisNet for marine radar images. MrisNet proposes a novel feature extraction network called FasterYOLO, which incorporates efficient convolutional units to accurately extract key features of ships in radar images. Furthermore, a simple, parameter-free attention module is introduced to infer the three-dimensional attention weights of feature maps, optimizing the deep feature output of the network and enhancing the ship segmentation at a higher semantic level. Additionally, the algorithm integrates a self-attention mechanism based on the Transformer structure into the feature fusion network to model and extract long sequence feature of images more accurately, aiding in distinguishing ship instances from the background environment. In the prediction structure, the algorithm improves the calculation of position loss for the predicted boxes by utilizing the EIoU function, thereby extracting more precise positional information of ships.

Experimental results have demonstrated that MrisNet outperforms common algorithms in ship segmentation on marine radar images. MrisNet achieves accurate segmentation of long-tail and short-tail ships in various scenarios, with recall and precision reaching 94.8% and 95.2%, respectively. Particularly, it exhibits acceptable performance in scenarios involving tiny ships, dense navigation, and complex interference. Furthermore, MrisNet exhibits advantages in terms of model parameter size and real-time computational cost. It has a parameter size of 13.8M and a calculation consumption of 23.5G, significantly reduced compared to deep YOLO series, making it more suitable for deployment in maritime monitoring devices. Considering the relatively limited image samples, future research will focus on expanding the RadarSeg dataset to cover a wider range of ship navigation scenarios. Additionally, the next steps in research will involve introducing more interference factors to enhance the model's robustness.

Author Contributions: Conceptualization, F.M., Z.K. and C.C.; methodology, F.M. and Z.K.; software, Z.K.; validation, Z.K.; formal analysis, F.M. and C.C.; investigation, J.S. and J.D.; resources, J.D.; data curation, J.S.; writing—original draft preparation, Z.K.; writing—review and editing, F.M. and Z.K.; visualization, Z.K.; supervision, F.M. and C.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Access to the data will be considered upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wei, Y.; Liu, Y.; Lei, Y.; Lian, R.; Lu, Z.; Sun, L. A new method of rainfall detection from the collected X-band marine radar images. *Remote Sens.* **2022**, *14*, 3600. https://doi.org/10.3390/rs14153600.
2. Li, B.; Xu, J.; Pan, X.; Chen, R.; Ma, L.; Yin, J.; Liao, Z.; Chu, L.; Zhao, Z.; Lian, J. Preliminary investigation on marine radar oil spill monitoring method using YOLO model. *J. Mar. Sci. Eng.* **2023**, *11*, 670. https://doi.org/10.3390/jmse11030670.
3. Wen, B.; Wei, Y.; Lu, Z. Sea clutter suppression and target detection algorithm of marine radar image sequence based on spatio-temporal domain joint filtering. *Entropy* **2022**, *24*, 250. https://doi.org/10.3390/e24020250.
4. He, W.; Ma, F.; Liu, X. A recognition approach of radar blips based on improved fuzzy c means. *Eurasia J. Math. Sci. Technol. Educ.* **2017**, *13*, 6005-6017. https://doi.org/10.12973/eurasia.2017.01048a.
5. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. https://doi.org/10.1109/tpami.2016.2577031.
6. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22-29 October **2017**; pp. 2961-2969.
7. Chen, X.; Su, N.; Huang, Y.; Guan, J. False-alarm-controllable radar detection for marine target based on multi features fusion via CNNs. *IEEE Sens. J.* **2021**, *21*, 9099-9111. https://doi.org/10.1109/JSEN.2021.3054744.
8. Chen, X.; Mu, X.; Guan, J.; Liu, N.; Zhou, W. Marine target detection based on Marine-Faster R-CNN for navigation radar plane position indicator images. *Front. Inf. Technol. Electron. Eng.* **2022**, *23*, 630-643. https://doi.org/10.1631/fitee.2000611.
9. Wang, Y.; Shi, H.; Chen, L. Ship detection algorithm for SAR images based on lightweight convolutional network. *J. Indian Soc. Remote Sens.* **2022**, *50*, 867-876. https://doi.org/10.1007/s12524-022-01491-1.
10. Li, S.; Fu, X.; Dong, J. Improved ship detection algorithm based on YOLOX for SAR outline enhancement image. *Remote Sens.* **2022**, *14*, 4070. https://doi.org/10.3390/rs14164070.
11. Zhao, C.; Fu, X.; Dong, J.; Qin, R.; Chang, J.; Lang, P. SAR ship detection based on end-to-end morphological feature pyramid network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4599-4611. https://doi.org/10.1109/jstars.2022.3150910.
12. Zhang, T.; Zhang, X. HTC+ for SAR ship instance segmentation. *Remote Sens.* **2022**, *14*, 2395. https://doi.org/10.3390/rs14102395.
13. Zhao, D.; Zhu, C.; Qi, J.; Qi, X.; Su, Z.; Shi, Z. Synergistic attention for ship instance segmentation in SAR images. *Remote Sens.* **2021**, *13*, 4384. https://doi.org/10.3390/rs13214384.
14. Yang, X.; Zhang, Q.; Dong, Q.; Han, Z.; Luo, X.; Wei, D. Ship instance segmentation based on rotated bounding boxes for SAR images. *Remote Sens.* **2023**, *15*, 1324. https://doi.org/10.3390/rs15051324.
15. Shao, Z.; Zhang, X.; Wei, S.; Shi, J.; Ke, X.; Xu, X.; Zhan, X.; Zhang, T.; Zeng, T. Scale in scale for SAR ship instance segmentation. *Remote Sens.* **2023**, *15*, 629. https://doi.org/10.3390/rs15030629.
16. Sun, Y.; Su, L.; Yuan, S.; Meng, H. DANet: Dual-branch activation network for small object instance segmentation of ship images. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 6708-6720. https://doi.org/10.1109/TCSVT.2023.3267127.
17. Sun, Y.; Su, L.; Luo, Y.; Meng, H.; Li, W.; Zhang, Z.; Wang, P.; Zhang, W. Global Mask R-CNN for marine ship instance segmentation. *Neurocomputing* **2022**, *480*, 257-270. https://doi.org/10.1016/j.neucom.2022.01.017.
18. Guo, M.; Guo, C.; Zhang, C.; Zhang, D.; Gao, Z. Fusion of ship perceptual information for electronic navigational chart and radar images based on deep learning. *J. Navig.* **2019**, *73*, 192-211. https://doi.org/10.1017/s0373463319000481.
19. Mao, D.; Zhang, Y.; Zhang, Y.; Pei, J.; Huang, Y.; Yang, J. An efficient anti-interference imaging technology for marine radar. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1-13. https://doi.org/10.1109/TGRS.2021.3068787.
20. Zhang, C.; Fang, M.; Yang, C.; Yu, R.; Li, T. Perceptual fusion of electronic chart and marine radar image. *J. Mar. Sci. Eng.* **2021**, *9*, 1245. https://doi.org/10.3390/jmse9111245.
21. Dong, G.; Liu, H. A new model-data co-driven method for radar ship detection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1-9. https://doi.org/10.1109/tim.2022.3169573.

22. Zhang, F.; Wang, X.; Zhou, S.; Wang, Y.; Hou, Y. Arbitrary-oriented ship detection through center-head point extraction. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1-14. https://doi.org/10.1109/tgrs.2021.3120411.

23. Zhang, T.; Wang, W.; Quan, S.; Yang, H.; Xiong, H.; Zhang, Z.; Yu, W. Region-based polarimetric covariance difference matrix for PolSAR ship detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1-16. https://doi.org/10.1109/tgrs.2022.3146385.

24. Qi, X.; Lang, P.; Fu, X.; Qin, R.; Dong, J.; Liu, C. A regional attention-based detector for SAR ship detection. *Remote Sens. Letters* **2022**, *13*, 55-64. https://doi.org/10.1080/2150704x.2021.1987574.

25. Yin, Y.; Cheng, X.; Shi, F.; Zhao, M.; Li, G.; Chen, S. An enhanced lightweight convolutional neural network for ship detection in maritime surveillance system. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5811-5825. https://doi.org/10.1109/JSTARS.2022.3187454.

26. Zhang, Y.-F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146-157. https://doi.org/10.1016/j.neucom.2022.07.042.

27. Jocher, G. YOLOv5 by Ultralytics. Available online: https://github.com/ultralytics/yolov5 (accessed on 21 September 2023)

28. Chen, J.; Kao, S.-h.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; Chan, S.-H.G. Run, don't walk: chasing higher flops for faster neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 18-22 June 2023; pp. 12021-12031.

29. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (ICML), Vienna, Austria, 18-24 July 2021; pp. 11863-11874.

30. Li, Y.; Yao, T.; Pan, Y.; Mei, T. Contextual transformer networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2022, *45*, 1489-1500. https://doi.org/10.1109/TPAMI.2022.3164083.

31. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574-8586. https://doi.org/10.1109/TCYB.2021.3095305.

32. Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.

33. Jocher, G. YOLO by Ultralytics. Available online: https://github.com/ultralytics/ultralytics (accessed on 21 September 2023)

34. 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), Vancouver, Canada, 18-22 June 2023; pp. 7464-7475.

35. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 11-17 October 2021; pp. 10012-10022.

36. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. Yolact: real-time instance segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, 27-28 October 2019; pp. 9157-9166.

37. Wang, X.; Zhang, R.; Kong, T.; Li, L.; Shen, C. Solov2: dynamic and fast instance segmentation. *arXiv* **2020**, arXiv:2003.10152.

38. Peng, S.; Jiang, W.; Pi, H.; Li, X.; Bao, H.; Zhou, X. Deep snake for real-time instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 13-19 June 2020; pp. 8533-8542.

39. Yang, G.; Lei, J.; Zhu, Z.; Cheng, S.; Feng, Z.; Liang, R. AFPN: asymptotic feature pyramid network for object detection. *arXiv* **2023**, arXiv:2306.15988.