# EGISD-YOLO: Edge Guidance Network for Infrared Ship Target Detection

Weida Zhan, Cong Zhang, Shufang Guo, Jinxin Guo, and Mingkai Shi

***Abstract*—In the marine field, infrared detection technology is of great significance for timely localization and detection of ships in security missions. However, since infrared ship targets are often in the environmental conditions of small pixel occupancy, low contrast and complex background, it poses a great challenge for multi-ship detection, classification, and localization tasks. Therefore, to address these problems we propose an edge information-guided infrared ship target detection (EGISD-YOLO) network, in which a dense-csp structure is designed to improve the csp module of YOLO to increase the reusability of the backbone feature information, and in addition, to address the noise and interference generated by the image in the complex background, a deconvolutional channel attention module (DCA) is designed to link the contextual language to the image, which relates the contextual semantics to obtain the local information of the target. Crucially, we propose an edge-guided structure that takes the edge information of low-level features as a cue to fuse with deep-level features to enrich the target contour and thus improve the target localization ability, so that the network still possesses robustness under low-contrast conditions, and finally, we add a small-size prediction head at the end of the network to further increase the detection ability of weak targets. The proposed EGISD-YOLO is experimentally demonstrated to have better detection performance for infrared ship targets.**

***Index Terms*—Attention mechanism, edge guidance, infrared target detection, ship image.**

## I. Introduction

In RECENT years, infrared detection technology has been developing rapidly, and the characteristics of noncontact as well as passive detection have made it a research hotspot in the field of civil reconnaissance and detection, which has been widely concerned by countries all over the world. Infrared detection technology has been widely used in early warning system, air defense system, and sea defense system in the marine field. In the sea defense system, the marine ship target detection can not only detect the ship target to prevent collision, but also provide technical support for the sea search and rescue, rescue and other work, so it has been widely concerned. Compared with visible light, infrared has a strong antiinterference ability, 24-h continuous operation, no need to make up light at night and other characteristics, and can better detect other target objects at the same time to hide themselves.

However, due to the unique characteristics of its target factors, the accurate detection and identification of infrared ships is extremely challenging. In general, the maritime ship target because of the imaging distance, resulting in the target in the whole image occupies small pixels, and because the infrared image texture features less, and the contrast between the background is low, it is very easy to submerge in the complex sea clutter, these characteristics make infrared ship small target detection more difficult. In the harbor scene, due to the close shooting distance, the number of docked ship targets is large and each target occupies more pixels of the whole image, these different scenes and target sizes bring some difficulty to the detection of marine ship targets, as shown in Fig. 1 (The images are from the infrared ship dataset introduced in Section III below.). The accuracy and real-time performance of the current infrared ship target detection methods cannot fully meet the needs of the sea defense scene, so the infrared ship target detection technology is still the focus and difficulty of the current research.

Earlier traditional target detection methods, usually using algorithms based on hand-designed features, detect targets by sliding a window over the image and using classifiers or regressors [1], [2], [3], but mostly due to their limited feature representation capability, reliance on manually labeled bounding boxes, lack of robustness to handle scales and rotations, difficulty in dealing with target occlusion and partial visibility, and relatively low training and inference efficiency.

Nowadays, with the rapid development of deep learning, it brings great help to the field of target detection. Currently the mainstream use of deep convolutional neural networks (CNNs) to learn features [4], [5], [6], [7], [8], this type of method has a strong feature expression ability and generalization, in response to the ship target is in the port and other complex scenes, the performance of the performance of the ship is better, and does not need to manually extract the features but directly from the original image to detect the target, so it can greatly improve the localization accuracy and efficiency of ship detection.

At present, the methods for detecting ship targets are not limited, and there is no lack of some traditional and deep learning methods combined. Articles [9], [10], [11], and [12] mainly used methods based on morphological reconstruction and multiscale filter filtering, analyzed the ship features one

Manuscript received 4 February 2024; revised 15 March 2024 and 7 April 2024; accepted 10 April 2024. Date of publication 16 April 2024; date of current version 30 May 2024. This work was supported by the Jilin Provincial Science and Technology Department Development Plan through Project Vehicle multi-band intelligent fusion enhanced imaging and processing system development under Grant 20240302029GX. *(Corresponding author: Weida Zhan.)*

Weida Zhan, Cong Zhang, Jinxin Guo, and Mingkai Shi are with the School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun 130013, China (e-mail: zhanweida@cust.edu.cn; 2022100778@mails.cust.edu.cn; guojinxin@mails.cust.edu.cn; 2576750867@qq.com).

Shufang Guo is with the Baicheng Meteorological Bureau, Baicheng 137000, China (e-mail: 971451700@qq.com).

Digital Object Identifier 10.1109/JSTARS.2024.3389958

Fig. 1. Example images of infrared ships. We give three infrared ship targets in Fig. 1, the first image shows a weak target under the effect of long-range imaging, the second image shows a target near the harbor under low contrast, and the third image shows multiple targets under the complex background near the harbor. Ship targets are indicated by red borders and the first image is zoomed in at the top right corner.

by one, enriched the features while weakening the impact of the target texture information is insufficient, and interspersed with thresholding to suppress the background clutter, this type of method is more stringent in the selection mechanism and parameters, which leads to the weaker generalization ability. Therefore, articles [13], [14], [15], and [16] used the joint task of traditional image grayscale processing and neural networks to extract target features after improving the contrast between the background and the ship's target, which alleviates the difficulty of the network's decision-making on the target, and has a higher generalization ability and detection efficiency.

Although the existing methods have carried out profound research on the segmentation of ship background and target and multiscale target features, for the classification task of multiple ships, optimizing the feature extraction network and feature fusion methods to improve the detection accuracy will increase the computational complexity at the same time, so taking into account the detection accuracy, generalization and real-time is the focus of the research and the difficulty of the ship detection needs to pay attention to. As an important basis for localization and segmentation in neural networks, target edge features can provide a more accurate description of the shape and structure of the target with their corner information, but existing research has paid little attention to this aspect. Thus, we have made an in-depth exploration of network methods with the ability of edge-guided semantic information, looking for optimization measures from feature localization measures in other domains (which are further described in Section II) to solve the above difficulties.

Therefore, to address the above difficulties and shortcomings, an infrared ship target detection network research method is proposed, and its main work can be summarized as follows.

1) Introducing edge feature-guided network structure in the yolo-v5 target detection network, fusing the semantic information in the low-level features that is conducive to localization into the deep-level network, assisting the network to understand the shape and contour of the target to better achieve classification and localization effects.
2) A dense-csp (DCSP) module is designed to replace the original csp module of BACKBONE, which improves gradient propagation and enhances feature reusability by means of dense connections.
3) To address the important target and background information in the original image that may be lost as the network deepens, a dilated convolutional channel attention (DCA) module is proposed to enable the network to better fuse the contextual semantics while capturing the abstract texture details and prediction distortion.
4) Additional small-size prediction heads are added to the YOLO network to improve the detection of weak target ships and classify the targets in more detail.

The rest of this article is organized as follows. Section II briefly describes the scheme of edge information guidance and the choice of building the deep learning network. Section III describes the dataset selection, Section IV describes the design of the proposed algorithm in detail, and Section V analyzes the findings of the comparison and ablation experiments and compares them with the state-of-the-art methods. Finally, Section VI concludes this article.

## II. Related Work

### A. Edge-Guided Scheme

In the field of image processing and computer vision, the early research on image edge information is mostly used for image enhancement or target contour acquisition, e.g., using sobel operator [17], canny operator [18], and other methods of image gradient computation and thresholding to recognize the edge information. However, with the development of deep learning, based on the image edge features contain rich information about the shape and contour of the object, and have the quality of assisting in obtaining the semantics of target localization and segmentation in neural networks, so they are widely used in the research of target segmentation and target detection.

In target segmentation research, the articles [19], [20], [21], [22], [23], and [24] used edge information as a guide to sharpen the target, and edge information was mostly integrated from the contextual semantics to obtain the localization information to assist in the fusion of the synthesized high and low level feature information and thus achieve segmentation, whereas the article [24] obtained the edge semantics by using mask-guided pyramid networks; in target detection research, the articles [25], [26] used the encoder part of different scales to progressively fuse features with the target significant edge extraction network to form a U-shaped structure to merge the object features

and enhance the edges to cope with the rough boundaries of the object;

However, most of the current research in the direction of target detection is aimed at the edges of significant targets, and it is still worthwhile to further improve and explore the structure of the guidance network for infrared ship images with low contrast and blurred contour boundaries.

### B. Deep Learning Scheme

Early traditional target detection methods, such as haar feature and AdaBoost cascade methods [1], directional gradient histogram based methods [2], feature transform-based methods [3], but mostly due to some shortcomings such as inadequate feature representation, sensitivity to illumination variations, complex background interference, difficulty in view angle change, target occlusion problem, and low resolution, which limit their effectiveness in infrared accuracy and robustness in ship target detection tasks.

Deep learning, as an important branch in the field of computer vision, has achieved remarkable results in tasks such as target detection, segmentation, and classification. Such methods can learn advanced feature representations from a large amount of image data by constructing a deep neural network model, so as to realize the understanding and analysis of images.

Currently, the mainstream detection networks are categorized into two detection classes: 1) single-stage and 2) two-stage. Single-stage detection algorithms such as YOLO [6], SSD [7], RetinaNet [27], etc., require only one process of feature extraction, have a faster detection speed, and are very suitable for scenes with real-time requirements, but the relative detection accuracy is lower; two-stage detection algorithms are representative of the RCNN family of algorithms (Faster RCNN [28], Mask RCNN [29], etc.), first use a number of candidate frames generated for feature extraction and classification, and then adjust the positioning of the candidate frames that are classified as the target to complete the regression task, which has higher accuracy but cannot meet the real-time demand of such a task as maritime monitoring.

Based on recent learning studies, we concluded that networks using the yolo model as a framework [30], [31], [32] perform well in a variety of detection tasks. For example, the paper [33] optimized IoU loss and module calculation methods based on yolo algorithm to apply them to embedded device deployment and ship detection. The paper [34], [35] designed structures based on yolov3 and yolov5, respectively, and added attention modules to improve the detection accuracy of remote sensing ship images. In this paper [36], [37], yolov5 was used to conduct an in-depth study on the lightweight method of SAR ship image detection model. The end to end regression method is used to predict the boundary boxes and categories by using image segmentation, which has strong real-time and global perception ability, and is suitable for the detection and classification of infrared ship targets. Therefore, after analyzing and comparing various structures, we finally selected yolov5 as the network backbone of the proposed method.

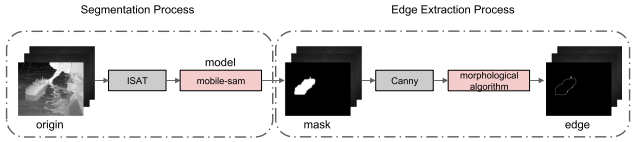


Fig. 2. Segmentation and edge extraction process for image targets.

## III. Dataset

The dataset used [38] is provided by Yantai Arrow Optoelectronics Technology Company for the study of infrared ship target recognition. The database collects more than 8000 infrared images, which are acquired in different scenarios using infrared panoramic radar imaging devices with different resolutions and focal lengths. The infrared wavelength belongs to the thermal infrared band and ranges from 3 to 14 micrometers. The spatial resolution of the images is between 0.3 m and 5 m. The resolutions of the images are 384×288, 640×512, and 1280×1024. The database records a total of 8000 images of cruise ships, bulk carriers, warships, sailboats, kayaks, container ships, and fishing boat targets at sea, in ports, and waterfront areas in different scenes, at different times of the day, and at different resolutions. The images in the dataset were labeled using liner, bulk carrier, warship, sailboat, canoe, container ship, and fishing boat as the labels for each type of ship target, and the targets were labeled using rectangular boxes. The main purpose of this database is to be used to study the target detection and recognition technology in the real world infrared sea defense field, so a database of infrared ship target detection in real sea defense scenarios has been established for verifying the effectiveness of infrared target detection algorithms in practical applications. We used 7402 images from the dataset for training, 1000 for validation, and 1000 for testing. And extensive experiments were conducted on this dataset using mainstream detection algorithms.

In addition, in order to study and analyze the image characteristics and algorithm features in detail, as shown in Fig. 2, we used segmentation algorithms and edge extraction methods to extract the mask image and edge image, respectively, specifically in the segmentation process, we convert the xml truth labels of the dataset into the corresponding labels in yolo format, and separate the infrared ship target from the background by using the mobile-sam model in the SAM segmentation method [39], [40] on the ISAT platform through manual supervision, and converted into a mask image, then the mask is used to extract the target edges using Canny operator and morphological operations to smooth the edges, and finally the edge truth image is generated to compare with the designed network and evaluate the corresponding metrics.

The infrared ship target detection method in this study will fully utilize the advantages and features of this dataset with the following benefits and implications. Real scenario validation: This dataset is collected under real sea defense scenarios, covering a wide range of practical environments such as sea, harbor, and waterfront. Therefore, our infrared ship target detection method can be validated in real scenarios, and the validation results are more effective for practical applications; diversity:

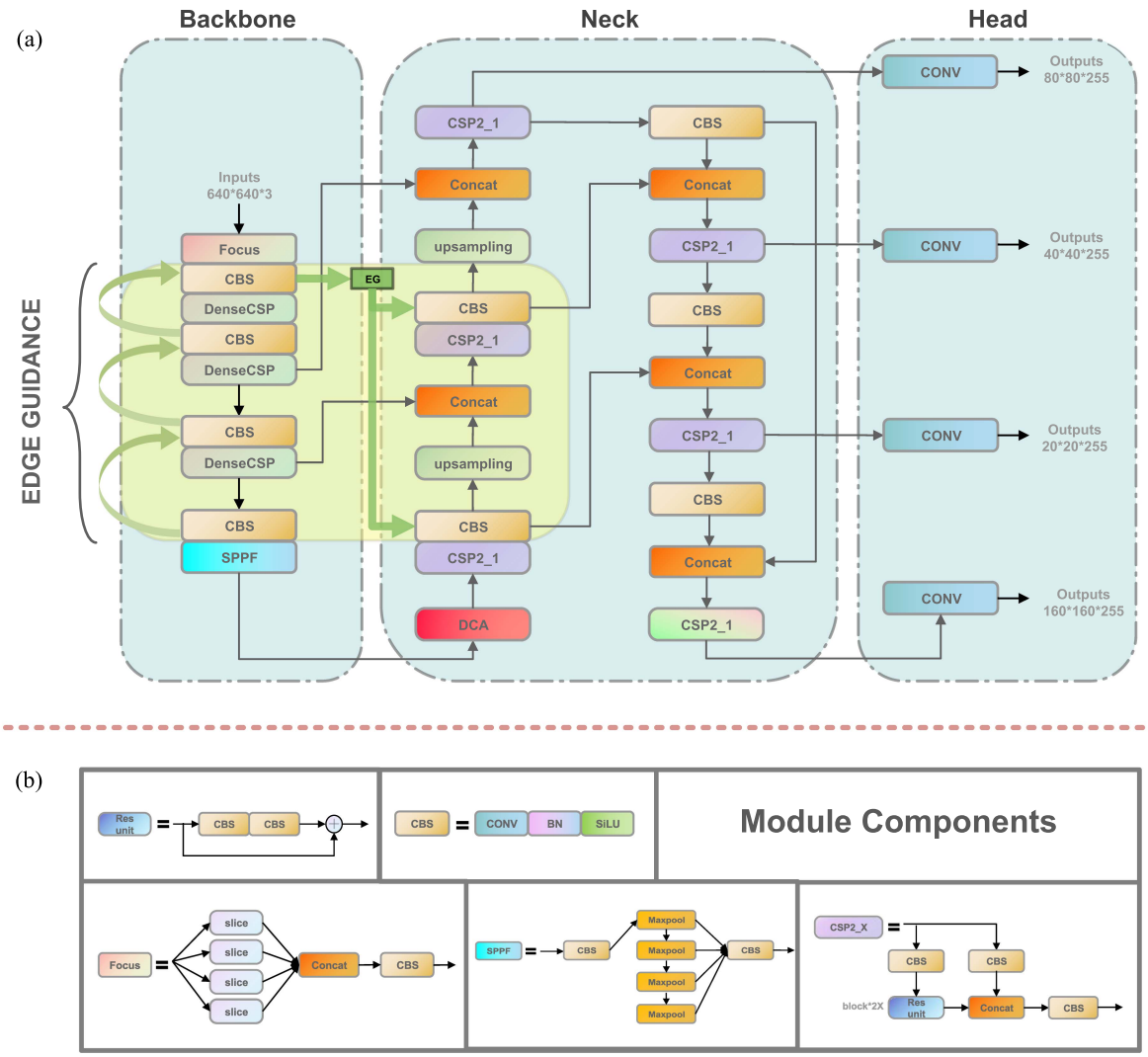


Fig. 3. Segmentation and edge extraction process for image targets.

The dataset collects infrared images with different resolutions and focal lengths, and contains seven types of ship targets, and this diversity allows our method to be tested and evaluated on different types of ship targets, which increases the method's ability to generalize; large-scale data: The dataset contains more than 8000 infrared images, which provides rich training and testing data. Therefore, it can effectively improve the performance and accuracy of our infrared ship target detection method; labeling accuracy: The ship targets in the dataset are labeled with rectangular frames. The xml tag contains the number, size, and position information of each ship type, so the accuracy of edge and mask image generation is guaranteed. This labeling accuracy enables us to perform accurate target detection and identification and evaluate the effectiveness of our method on tasks such as target localization and bounding box regression.

## IV. Methods

In this section, we present an introduction to the proposed edge information-guided infrared ship target detection (EGISD-YOLO), describing the overall architecture and main innovations of the proposed network model in the following subsections.

### A. Overall Architecture

The network architecture of EGISD-YOLO is illustrated in Fig. 3(a). The overall consists of three components: Backbone, Neck, and Predictor Head. The yolo-v5s model is used as the basic framework of the network, and its module composition is illustrated in Fig. 3(b). Compared with several other models of yolo-v5, this model has a smaller number of parameters and faster detection speed, which ensures the real-time detection of ship targets and facilitates us to improve the network.

First, in the Backbone part, we replace all the original CSP modules with the new Dense CSP module, which increases the feature expressiveness through feature multiplexing and gradient flow. Then between Backbone and Neck, as marked by the yellow part in Fig. 3(a), an edge-guiding structure is designed, which starts from the CBS module at the bottom of

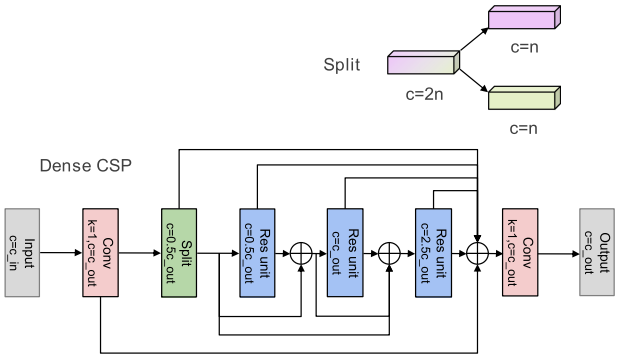


Fig. 4. Dense CSP module.

Backbone and performs a jumper connection to the CBS module at the top in turn, and after integrating to the largest CBS in the sensory field, the edge information of the feature is imported into two different scales of the feature in the Neck by an EG module layers, which enhances the localization ability for targets submerged in the background. In addition, a channel attention mechanism is added to the Neck, which echoes the fusion feature of SPPF, to enhance the target semantic attention without significantly increasing the computational effort. Finally, a fourth prediction head is newly added to the Head section for focusing on weak targets and reducing the misdetection and omission rates of the detection results. Compared with the original yolo-v5 structure, our model oriented to infrared ship targets can show better detection results.

### B. Dense CSP Module

The structure of the CSP module can be referred to the CSP2_X shown in Fig. 3(b), which splits the input features into two branches for processing, which promotes feature reuse and information flow; the cross-layer connection ensures the inference speed of the model and the stability of gradient propagation. The Dense structure [41], [42] can increase the nonlinear transformations and feature combinations inside the module to better capture the complex relationships and feature interactions in the input data, and further enhance the feature expressiveness. Therefore we combine the two to construct the Dense CSP module, whose structure is shown in Fig. 4.

The input features first go through a $1 \times 1$ convolutional layer to change the number of feature channels, and then the channels are split into two by the split operation, which is connected to the Res unit of the classical Dense structure, and the tail uses concat and convolutional layers to integrate the feature information for output. The process can be expressed as follows:

$$F_1 = conv_{1\times1}(F_{in}) \tag{1}$$

$$F_2 = Split(F_1) \tag{2}$$

$$F_3 = Dense(R_1, R_2, R_3) \tag{3}$$

$$F_{\text{out}} = conv_{1\times1}\left(cat\left(F_1, F_2, F_3\right)\right) \tag{4}$$

where $F_{\text{in}}$ and $F_{\text{out}}$ denote the input and output features, respectively, $R_i(i = 1, 2, 3)$ represents the Res unit block, $cat$ represents the concat operation, and $Dense$ represents the classical Dense connection structure.

### C. Deconvolution Channel Attention Module

The channel attention mechanism can play a role in suppressing the large amount of noise and interference present in the background of the infrared ship image, and the important information can be separated by giving higher weights to the channels with high attention to the feature targets through the network adaptive computation. However, we found through experiments that for images in the environment of a wide variety of ships and weak targets, these target features lack local semantic information, and we need to expand the receptive field to obtain nonlocal contextual information.

Combining the above conditions and inspired by the literature [43], we articulate an algorithm combining channel attention and inverse convolution after Backbone's SPPF module, first, we feed the input features $Feature_1$ into the channel attention structure shown in Fig. 5(a), transform the global information into vector representation by pooling and MLP, and then model the nonlinear relationship of the one-dimensional features, which is generated by weighting with the initial $Feature_2$ features; Then entering the inverse convolution part of the attention in Fig. 5(b), the process can be represented as follows:

$$P_1 = Dconv_{3\times3}(Feature_2) \tag{5}$$

$$P_2 = Dconv_{3\times3}\left(conv_{1\times1}\left(cat\left(P_1, Feature_2\right)\right)\right) \tag{6}$$

$$Feature_3 = conv_{1\times1}\left(cat\left(P_1, P_2, Feature_2\right)\right) \tag{7}$$

where, $P_1$, $P_2$ represents the feature maps obtained after the first and second deconvolution operations, $Dconv_{3\times3}$ represents the dilation convolution with a convolution kernel size of 3, with dilation rates of 2 and 4 (to obtain rich sensory field information while ensuring the computational efficiency of the network), and the number of channels is 1/2 and 1/4 of the input, $Feature_3$ is the output feature, respectively. The use of dilation convolution after the attention structure can effectively avoid the problem of information loss caused by the meshing effect, in addition, reversing the primary and secondary relationship between the initial and dilation features can also lead to information bias, so we prevented the omission of information by taking the original features as the dominant features in the design of the channels.

### D. Edge-Guided Network Architecture

Ship targets in infrared images are usually characterized by low contrast, blurring, and similarity to the background, and thus have been a challenge and hotspot for research in this field, where the lack of clear target contours makes it impossible for general networks to obtain clear target feature localization and key semantic information from shallow networks. Inspired by the ideas of the article [25], [26], the target detection method of mask image segmentation can strengthen the target localization information through edge features running through the network, so it is experimentally tested in our study.

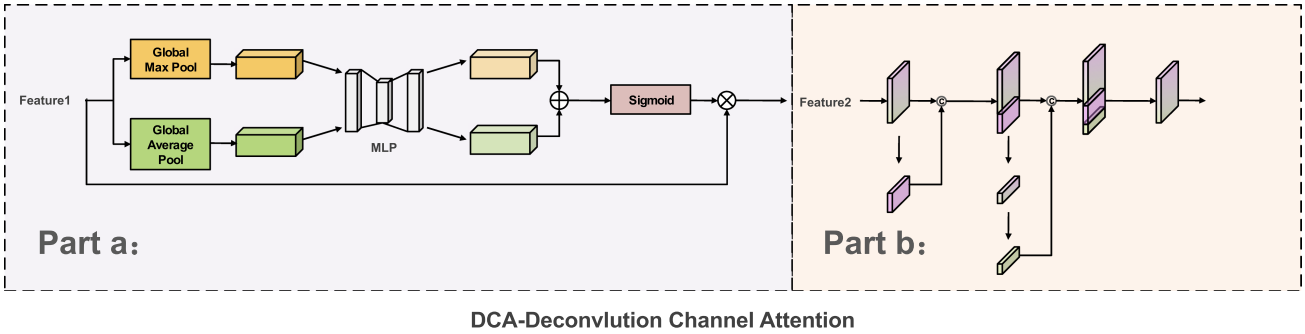


Fig. 5. Inverse convolution channel attention module. Part a: Inverse Convolutional Channel Attention Module Channel Attention Part. Part b: Attention module anti-convolution part.

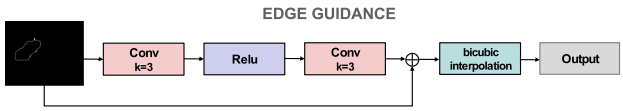


Fig. 6. Edge information processing module EG for edge-guided structures.

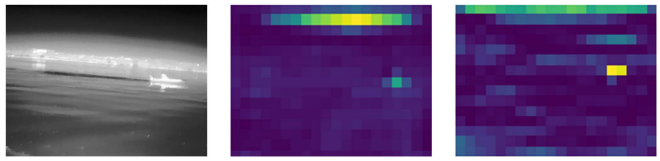

Fig. 7. Comparison of feature maps of infrared ship images. The left is the original image of the infrared ship image, the middle is the feature map at the next level of the fusion edge, and the right is the fused guided output feature map.

In order to build an information bridge between different depths of the network, we design an edge-guided structure, as shown in the yellow area in Fig. 3(a), where the Backbone part is bottom-up along the direction of the green arrow, and the deep features converge to the shallow features in turn. Then, through the weight allocation to the multiscale semantics of the low-level contour and localization semantics as the main information, after the concise processing of the EG module, the edge semantics will be sent to the Neck and the high-level features to achieve the fusion of the fusion of the semantics from the bottom layer with a large sense of the field of the positional information, resulting in the enhancement of features to obtain a clearer target contour. The process can be described as follows:

$$\widetilde{C}_3 = C_3 + up\left(Swi\left(\mu, C_4\right)\right) \tag{8}$$

$$\widetilde{C}_i = C_i + up\left(Swi\left(\mu, \widetilde{C}_{i+1}\right)\right), \quad i = 1, 2 \tag{9}$$

where $C$ represents the features output from the CBS module, Backbone part of the CBS block from top to bottom as $C_1$ to $C_4$ sequentially, on behalf of the use of bilinear interpolation $up$ performs the feature upsampling operation, and $Swi$ represents the convolutional layer used to change the feature channel, $\mu$ for the convolutional layer control parameters. The gain of the low-level edge information is maximized by incrementally weighting the channels with the previous level, after obtaining the fused semantic information of the multilevel features, we threshold the features at this level using morphological operations and filter with canny operator to obtain enhanced edges, and then feed the obtained output edges to the EG module.

The EG module for modifying the edge features is shown in Fig. 6, which further integrates the information and increases the edge saliency by a simple combination of $3 \times 3$ convolutional layer filtering and ReLU function, and adjusts the feature size at the tail using bicubic interpolation for the final feed into the Neck. Therefore, as shown in Fig. 7, the highlighted part is the focus area of the target, and the features before fusion lack the sensitivity to features in the background. However, the target features guided by edge information significantly shift the attention from the fuzzy background to the ship target, which further confirms the effectiveness of our method.

### E. Dim Target Prediction Head

In addition to the original three target prediction heads of yolo-v5, we add a new weak target prediction head to help the model cope with the small targets in ship images and the problem of missed detection and wrong detection that occurs easily in dense scenes, and the multiscale perception capability can better learn and predict the location and bounding box of the weak targets to provide more accurate target localization. As shown in the Neck section of Fig. 3(a), similar to the other three prediction heads, the outputs of the CBS blocks are concatenated using two scales, and then introduced into the Head after a CSP structure, with the difference that we replace the last two layers of CSP2_X with bicubic interpolation and global average pooling, avoiding overfitting while maintaining the structural invariance of the target features.

The new predictive Head feature map scale size is set to $160 \times 160$, which is a quarter of the original image size, in addition to subsequent experiments showing that fine-grained feature representations are beneficial to the detection results.

## V. Result and Discussion

In this section, we perform a qualitative-quantitative evaluation of the proposed method using the infrared ship dataset

mentioned in Section III and deeply analyze the experimental results. We first describe the evaluation metrics used by the algorithm in Section V-A, followed by the implementation and experimental configuration details of the algorithm in Section V-B. In Section V-C, we show the quantitative results of the algorithm, comparing it with other state-of-the-art ship detection methods on the same dataset in order to prove its effectiveness; meanwhile, in Section V-D we show the qualitative results of the algorithm, with a visual example of how the algorithm performs on different images, highlighting the successful ship detection methods. performance, highlighting successful ship detection and situations where challenges may be encountered. In addition, in Section V-E we perform ablation analysis, evaluate the impact of different components on the performance of the algorithm and discuss their importance in the overall performance, and finally, in Section V-F we conclude with a discussion of possible improvements and shortcomings of the method.

### A. Evaluation Metrics

Considering the characteristics of infrared ship targets, the following commonly used detection metrics are used to measure the accuracy and comprehensiveness of the algorithm, so as to objectively evaluate the target detection performance:

1) *Precision:* The precision rate is the ratio of the number of samples correctly detected as positive classes to the number of all samples detected as positive classes. In target detection, the precision rate measures how many of the detected targets are true targets.
2) *Recall:* Recall is the ratio of the number of samples correctly detected as positive classes to the number of samples of all true positive classes. In target detection, recall measures how many true targets were correctly detected.
3) *Average Precision (AP):* Average precision is the area between precision and recall calculated at different thresholds. In target detection, the corresponding precision and recall are calculated according to different confidence thresholds, and then the area under their curves is calculated.
4) *Mean Average Precision (mAP):* Mean average precision is the average of the mean precision of all categories. It is a measure of the average precision of multiple categories and is used to evaluate the overall target detection performance.

### B. Implementation Details

To be fair, all our experiments were run using the same server with CPU Xeon(R) Platinum 8255 C and GPU RTX 2080 Ti 11 GB, and we implemented the proposed EGISD-YOLO on the framework of Pytorch 2.0 and cuda version 11.8. The images are adaptively scaled to 640×640 pixels in the model, the initial learning rate of the network is set to 0.003, the weight decay is set to 0.0005, the batch size is set to 16, the optimizer uses Adam, and all the CNN-based detection models are trained on the dataset for 150 epochs.

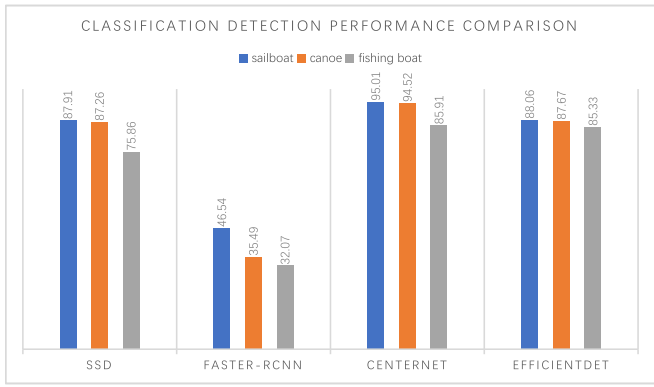


Fig. 8. Accuracy evaluation for sailboat, canoe, and fishing boat categories (I).

### C. Quantitative Results

In order to fairly evaluate the performance of EGISD-YOLO, we used seven state-of-the-art target detection methods for comparison, including SSD [7] using a single-stage multiscale feature layer, Faster-RCNN [28] with a two-stage detector and shared features, CenterNet [44] with a centroid detection approach, and EfficientDet [45] with structural extensions and bidirectional weighted feature capability of EfficientDet, RetinaNet [27] that uses Focal Loss to optimize the classification problem, and the Yolo family (Yolo-v5, Yolo-v7) that excels in global sensing and real-time detection. In order to fully demonstrate the performance and evaluation level of the participating detection methods in our experiments, we chose to ignore the traditional methods with poor performance, and adapted RetinaNet and CenterNet, which are based on the Tensorflow platform, to Pytorch's deep learning framework to ensure the computing speed and stability of the models.

As shown in Table I, EGISD-YOLO and other methods are evaluated in terms of the most intuitive precision rate, the recall rate for verifying the authenticity of the detection results, the average precision evaluation under different thresholds $mAP_{50}$ and $mAP_{\{0.5:0.95\}}$, among them, $mAP_{50}$ for the broader threshold performance evaluation and $mAP_{\{0.5:0.95\}}$ in favor of the strict IoU range control; in addition, the model size and the detection rate are introduced to measure the model complexity and real-time performance. The data boldly labeled in the table represent the best results for each column, and it can be observed that EGISD-YOLO achieves the highest precision and recall rates of 96.3% and 91.2% without significant increase in complexity compared to the baseline model, which also maintains a $mAP$ high level, and the detection rate is basically the same as that of the baseline.

In addition, for the sailboat, canoe, and fishing boat categories, which have fewer sample targets in the dataset, the model learning task is more difficult, and it is a great challenge to improve the detection of these categories, which requires the model to show superior target classification ability in training. Therefore, we analyze their accuracy metrics separately to observe the classification performance of the network. A comparison of the accuracy of the three categories is shown in Figs. 8 and 9 below, which shows that most of the methods have significantly lower

TABLE I
QUANTITATIVE EVALUATION OF EGISD-YOLO AND OTHER METHODS

| Methods | Precision | Recall | $mAP_{50}$ | $mAP_{\{0.5:0.95\}}$ | Model size | Speed/s |
|---|---|---|---|---|---|---|
| SSD | 0.915 | 0.596 | 0.753 | 0.745 | 93.6M | 0.042 |
| FasterRCNN | 0.655 | 0.746 | 0.718 | 0.709 | 108.0M | 0.099 |
| CenterNet | 0.959 | 0.667 | 0.838 | 0.820 | 124.0M | 0.018 |
| EfficientDet | 0.931 | 0.616 | 0.697 | 0.693 | 15.0M | 0.047 |
| RetinaNet | 0.927 | 0.608 | 0.694 | 0.690 | 139.0M | 0.035 |
| Yolov5 | 0.932 | 0.898 | **0.940** | 0.654 | **13.8M** | **0.012** |
| Yolov7 | 0.917 | 0.806 | 0.889 | **0.877** | 142.4M | 0.032 |
| ours | **0.963** | **0.912** | 0.937 | 0.841 | 16.6M | 0.015 |

TABLE II
QUANTITATIVE EVALUATION OF EGISD-YOLO AND OTHER METHODS

| Structure | DCSP | DCA | EG | Head | Precision | Recall | $mAP_{50}$ | $mAP_{\{0.5:0.95\}}$ |
|---|---|---|---|---|---|---|---|---|
| Net1 | | | | | 0.932 | 0.898 | 0.940 | 0.654 |
| Net2 | ✓ | | | | 0.937 | 0.892 | 0.935 | 0.644 |
| Net3 | | ✓ | | | 0.930 | 0.902 | 0.927 | 0.702 |
| Net4 | | | ✓ | | 0.936 | 0.905 | 0.940 | 0.778 |
| Net5 | ✓ | ✓ | | | 0.940 | 0.905 | 0.937 | 0.720 |
| Net6 | ✓ | | ✓ | | 0.933 | 0.904 | 0.939 | 0.779 |
| Net7 | | ✓ | ✓ | | 0.942 | 0.910 | 0.940 | 0.805 |
| Net8 | ✓ | ✓ | ✓ | | 0.961 | 0.912 | 0.941 | 0.815 |
| Net9 | ✓ | ✓ | ✓ | ✓ | 0.963 | 0.915 | 0.937 | 0.841 |

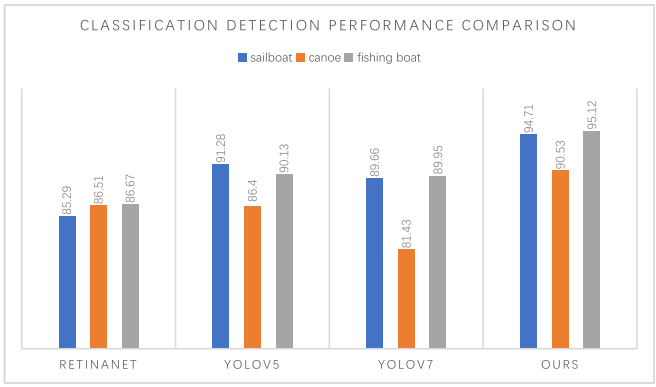


Fig. 9. Accuracy assessment for sailboat, canoe, and fishing boat categories (II).

accuracy when dealing with these kinds of targets, implying that there will be more probability of missed and wrong detections in the prediction, whereas our method still exhibits a high detection correctness, which side-by-side shows the excellent robustness of the algorithm.

### D. Qualitative Results

As shown in Fig. 10, we use the proposed EGISD-YOLO algorithm with other methods to conduct prediction experiments on four representative scenes in the dataset prediction set and analyze the qualitative results obtained. The first column in the figure is the original image, and we can observe that most algorithms show different degrees of misdetection and omission in the first image where the scene is more complex, which leads to serious bias due to the large number of targets, and the difficulty in recognizing the difference between the background and the targets that are in the vicinity of the port in networks with weak localization capabilities such as SSD and Retinanet. In contrast, our method not only correctly recognizes the targets, but also achieves high confidence scores under the precise guidance of edge information. In the second image, we selected an image with low contrast between the background and the target, which is a considerable challenge for the algorithm to evaluate the confidence of the target, e.g., in Centernet, the confidence of the target drops to about half, which may result in the loss of the target for the task of strict IoU threshold processing. In the third and fourth images, we continue to examine the accuracy of the network's classification and prediction frames, and our algorithms both achieve accurate processing and judgments and show excellent computing speed.

### E. Ablation Studies

In this section, we will discuss the effectiveness of the structure of each component of the EGISD-YOLO algorithm and verify the possibility of optimizing the algorithm in turn using different combinations of forms. As shown in Table II, the first column represents the nine network structures obtained by combining the four modules in columns 2–4 with the baseline model, and the evaluation metrics remain the same as those used in the quantitative experiments: Precision, Recall, and $mAP$.

In the experiments, we observe that the edge-guided modules, when acting alone or in combination with other structures, show stable and beneficial performance in detection, especially for precision and $mAP_{\{0.5:0.95\}}$, which increased by 18.9% in Net4 $mAP_{\{0.5:0.95\}}$ compared to the baseline model, and by 20.9% in the comparison $mAP_{\{0.5:0.95\}}$ between Net2 and Net6, which fully illustrates the role of the edge-feature-guided mechanism in promoting the classification ability of the model. ability; moreover, in Net8, it can be seen that the combination of the three structures DCSP, DCA, and EG is the most

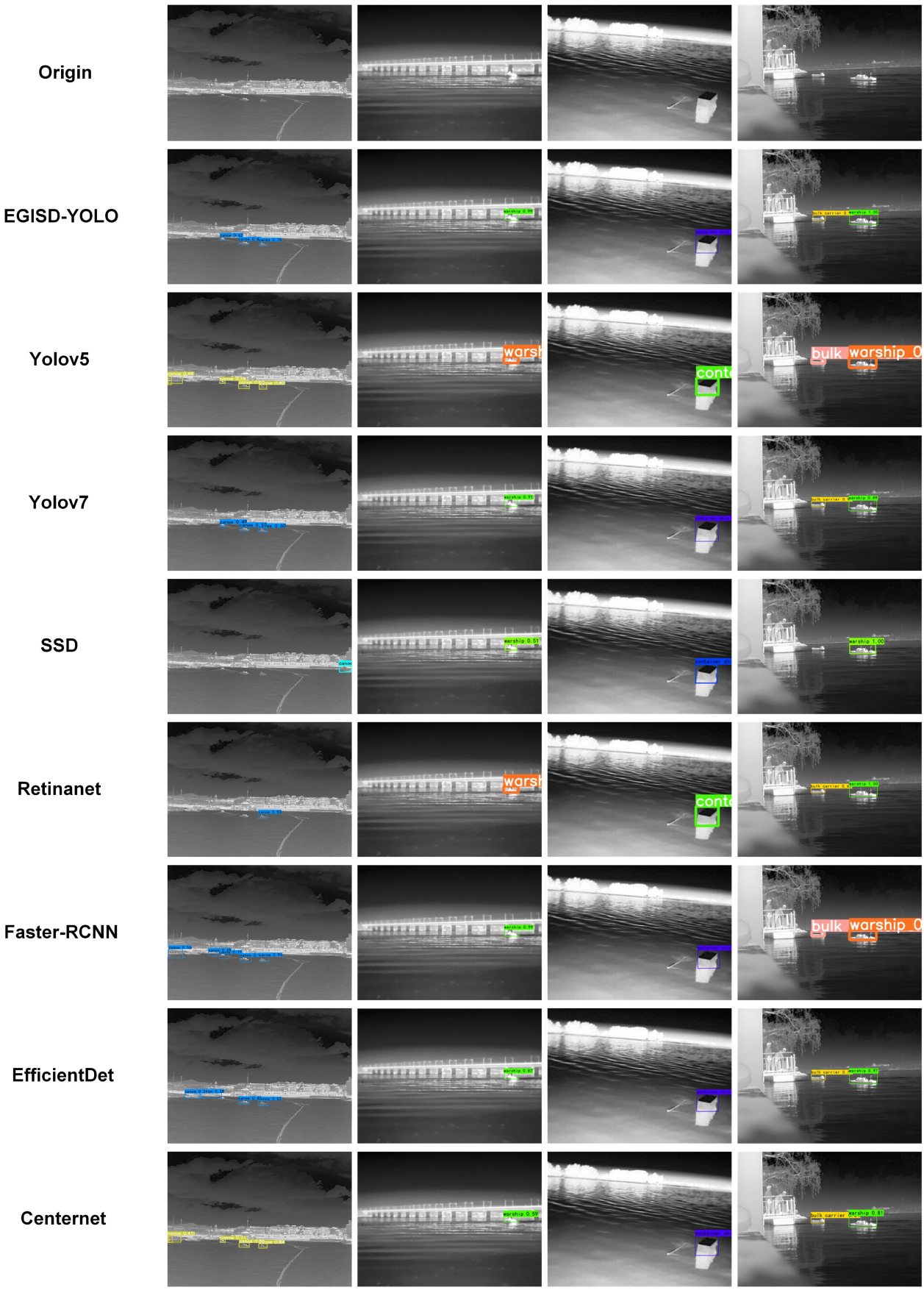


Fig. 10. Qualitative results of the EGISD-YOLO algorithm compared to other methods.

significant improvement for the network accuracy rate, which increases by 3% compared with the baseline model, proving that there is a good fit between our model structures, and that the way of fitting multiscale features is also worth further digging and exploring.

### F. Discussion

In the study of the effect of mining edge guidance mechanism on the localization and classification performance of the object detection network, we observe that it has a promotion effect on the improvement of indicators, but how sharp classification ability this mechanism can provide under different IoU thresholds is still a problem that needs to be considered and discussed. In addition, the mask image of edge features can expand the richness of the data set by making the truth value. We can also try to use it as a sample to improve the loss, and explore its role in the semantics of target positioning and the impact of recall changes.

## VI. Conclusion

In this article, an infrared ship target detection network EGISD-YOLO based on edge-guided structure is proposed. First, a DCSP module is designed to improve the CSP module to capture the complex relationships in the input data and optimize the feature expression by using the Dense structure to increase the nonlinear transformation and feature combination inside the module; then, to address the large amount of noise and interference in the background of the infrared ship image, an inverse convolutional channel attention is proposed to expand the feature sensing field to obtain nonlocal contextual information, and to give the nonlocal contextual information. localized contextual information and give more weights to the channels with high attention. It should be emphasized that, in order to solve the problem of low contrast, blurring, and similarity to background of ship targets in infrared images, an edge information guiding mechanism is designed to converge the deep features into the shallow features, and to emphasize the edge and localization semantic information through the weight assignment in order to obtain a clearer target contour from the underlying features and to enhance the localization and classification ability of the features. Finally, on the basis of the original target prediction head of the network, a prediction head specialized in detecting weak and small targets is added to better learn and predict the location and bounding box of small targets in the ship images and to cope with the problem of easy omission and misdetection in infrared images. Numerous experiments in the article demonstrate that our EGISD-YOLO achieves leading performance in infrared ship classification detection. In future work, we will further explore more effective algorithmic frameworks in pursuit of higher detection technology performance.

## References


[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. I–I.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 91–99, 2015.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[7] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Springer, Oct. 11–14, 2016, pp. 21–37.

[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[9] Y. Li, Z. Li, Y. Zhu, B. Li, W. Xiong, and Y. Huang, "Thermal infrared small ship detection in sea clutter based on morphological reconstruction and multi-feature analysis," *Appl. Sci.*, vol. 9, no. 18, 2019, Art. no. 3786.

[10] P. Wu et al., "SARFB: Strengthened asymmetric receptive field block for accurate infrared ship detection," *IEEE Sensors J.*, vol. 23, no. 5, pp. 5028–5044, Mar. 2023.

[11] X. Wang and C. Chen, "Ship detection for complex background SAR images based on a multiscale variance weighted image entropy method," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 184–187, Feb. 2017.

[12] W. Mo and J. Pei, "Nighttime infrared ship target detection based on two-channel image separation combined with saliency mapping of local grayscale dynamic range," *Infrared Phys. Technol.*, vol. 127, 2022, Art. no. 104416.

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

[14] L. Li, G. Liu, Z. Li, Z. Ding, and T. Qin, "Infrared ship detection based on time fluctuation feature and space structure feature in sun-glint scene," *Infrared Phys. Technol.*, vol. 115, 2021, Art. no. 103693.

[15] X. Chen, C. Qiu, and Z. Zhang, "A multiscale method for infrared ship detection based on morphological reconstruction and two-branch compensation strategy," *Sensors*, vol. 23, no. 16, 2023, Art. no. 7309.

[16] J. Liu, H. Chen, and Y. Wang, "Multi-source remote sensing image fusion for ship target detection and recognition," *Remote Sens.*, vol. 13, no. 23, 2021, Art. no. 4852.

[17] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[18] X. Sheng, C. Kang, J. Zheng, and C. Lyu, "An edge-guided method to fruit segmentation in complex environments," *Comput. Electron. Agriculture*, vol. 208, 2023, Art. no. 107788.

[19] J. Jin, W. Zhou, R. Yang, L. Ye, and L. Yu, "Edge detection guide network for semantic segmentation of remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5000505.

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[21] X. Yin, X. Li, P. Ni, Q. Xu, and D. Kong, "A novel real-time edge-guided lidar semantic segmentation network for unstructured environments," *Remote Sens.*, vol. 15, no. 4, 2023, Art. no. 1093.

[22] Y. Chong, X. Chen, and S. Pan, "Context union edge network for semantic segmentation of small-scale objects in very high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2020, Art. no. 6000305.

[23] Y. Ni, J. Liu, J. Cui, Y. Yang, and X. Wang, "Edge guidance network for semantic segmentation of high resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9382–9395, 2023.

[24] W. Zhang, H. Fan, X. Xie, Q. Wang, and Y. Tang, "Mask guidance pyramid network for overlapping cervical cell edge detection," *Appl. Sci.*, vol. 13, no. 13, 2023, Art. no. 7526.

[25] J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "EGNet: Edge guidance network for salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8779–8788.

[26] X. Zhou et al., "Edge-guided recurrent positioning network for salient object detection in optical remote sensing images," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 539–552, Jan. 2023.

[27] C. Li and S. Wang, "Infrared offshore ship dataset," 2021. [Online]. Available: http://iray.iraytek.com:7813/apply/Sea_shipping.html
[28] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6569–6578.
[29] S. Pan, Y. Tao, C. Nie, and Y. Chong, "PEGNet: Progressive edge guidance network for semantic segmentation of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 637–641, Apr. 2021.
[30] S. Yuan et al., "IRSDD-YOLOv5: Focusing on the infrared detection of small drones," *Drones*, vol. 7, no. 6, 2023, Art. no. 393.
[31] Y. Fan et al., "Application of improved YOLOv5 in aerial photographing infrared vehicle detection," *Electronics*, vol. 11, no. 15, 2022, Art. no. 2344.
[32] W. Wu et al., "Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image," *PLoS One*, vol. 16, no. 10, 2021, Art. no. e0259283.
[33] J. Li and J. Ye, "Edge-YOLO: Lightweight infrared object detection method deployed on edge devices," *Appl. Sci.*, vol. 13, no. 7, 2023, Art. no. 4402.
[34] Y. Da, X. Gao, and M. Li, "Remote sensing image ship detection based on improved YOLOv3," in *Proc. 7th Int. Conf. Intell. Comput. Signal Process.*, 2022, pp. 1776–1781.
[35] L. Li, L. Jiang, J. Zhang, S. Wang, and F. Chen, "A complete YOLO-based ship detection method for thermal infrared remote sensing images under complex backgrounds," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1534.
[36] M. F. Humayun, F. A. Nasir, F. A. Bhatti, M. Tahir, and K. Khurshid, "YOLO-OSD: Optimized ship detection and localization in multiresolution SAR satellite images using a hybrid data-model centric approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5345–5363, 2024.
[37] Y. Guo, S. Chen, R. Zhan, W. Wang, and J. Zhang, "LMSD-YOLO: A lightweight YOLO algorithm for multi-scale SAR ship detection," *Remote Sens.*, vol. 14, no. 19, 2022, Art. no. 4801.
[38] M. Alfaraj, Y. Wang, and Y. Luo, "Enhanced isotropic gradient operator," *Geophysical Prospecting*, vol. 62, no. 3, pp. 507–517, 2014.
[39] C. Zhang et al., "A comprehensive survey on segment anything model for vision and beyond," 2023, *arXiv:2305.08196*.
[40] W. Ji, J. Li, Q. Bi, W. Li, and L. Cheng, "Segment anything is not always perfect: An investigation of sam on different real-world applications," 2023, arXiv:2304.05750.
[41] T. Deng, G. Mao, Z. Zhou, and Z. Duan, "Road vehicle detection and recognition algorithm based on densely connected convolutional neural network," *J. Comput. Appl.*, vol. 42, no. 3, 2022, Art. no. 883.
[42] L. Yang et al., "CondenseNet V2: Sparse feature reactivation for deep networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3569–3578.
[43] Y. Han, J. Liao, T. Lu, T. Pu, and Z. Peng, "KCPNet: Knowledge-driven context perception networks for ship detection in infrared imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2022, Art. no. 5000219.
[44] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10781–10790.
[45] Z. Lu, V. Rathod, R. Votel, and J. Huang, "RetinaTrack: Online single stage joint detection and tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14668–14678.

**Weida Zhan** received the Ph.D. degree in optical engineering from the Changchun University of Science and Technology, Changchun, China, in 2011.

He is currently a Professor with the School of Electronic Information Engineering, Changchun University of Science and Technology. His research interests include artificial intelligence, image processing, and infrared imaging techniques.

**Cong Zhang** received the bachelor's degree in engineering from the North University of China, Taiyuan, China, in 2021. He is currently working toward the master's degree in control science and engineering from the Changchun University of Science and Technology, Changchun, China.

His research interests include computer vision, with an emphasis on infrared object detection, and object tracking applications.

**Shufang Guo** graduated from Lanzhou Meteorological School, Lanzhou, China, in 1996, and from the Changchun University of Technology, Chanchun, China, in 2004, majoring in computer and application.

She is currently Deputy Senior Meteorological Engineer with Baicheng Meteorological Bureau, Baicheng, China. Her research interests include meteorological science and computer vision.

**Jinxin Guo** was born in 1998 in Gansu, China. He received the bachelor's degree in communication engineering from Beihua University, Jilin, China. He is currently working toward the Ph.D. degree with the Changchun University of Science and Technology, Changchun, China.

His research interests include model antagonism and infrared dim target detection.

**Mingkai Shi** received the bachelor's degree in engineering from Beijing Jiaotong University, Beijing, China, in 2021. He is currently working toward the master's degree in control science and engineering from the Changchun University of Science and Technology, Changchun, China.

His research interests include computer vision, pattern recognition, and intelligent control.