MDPI

Article

Multi-Scale Polar Object Detection Based on Computer Vision

Shifeng Ding [1], Dinghan Zeng [1], Li Zhou [2,*], Sen Han [1], Fang Li [2] and Qingkai Wang [3]

1 School of Naval Architecture and Ocean Engineering, Jiangsu University of Science and Technology, Zhenjiang 212100, China; 15001945469@163.com (S.D.); 211210101417@stu.just.edu.cn (D.Z.); hansen@just.edu.cn (S.H.)
2 School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University, Shanghai 200030, China; fang.li@sjtu.edu.cn
3 State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116024, China; wangqingkai@dlut.edu.cn
* Correspondence: zhouli209@sjtu.edu.cn

Abstract: When ships navigate in polar regions, they may collide with ice masses, which may cause structural damage and endanger the safety of their occupants. Therefore, it is essential to promptly detect sea ice, icebergs, and passing ships. However, individual data sources have limits and should be combined and integrated to obtain more thorough information. A polar multi-target local-scale dataset with five categories was constructed. Sea ice, icebergs, ice melt ponds, icebreakers, and inter-ice channels were identified by a single-shot detector (SSD), with a final mAP value of 70.19%. A remote sensing sea ice dataset with 15,948 labels was constructed. The You Only Look Once (YOLOv5) model was improved with Squeeze-and-Excitation Networks (SE), Funnel Activation (FReLU), Fast Spatial Pyramid Pooling, and Cross Stage Partial Network (SPPCSPC-F). In the detection stage, a slicing operation was performed on remote sensing images to detect small targets. Simulated sea ice data were included to verify the model's generalization ability. Then, the improved model was trained and evaluated in an ablation experiment. The mAP, recall (R), and precision (P) values of the improved YOLOv5 were 75.3%, 70.3, and 75.4%, with value increases of 3.5%, 3.4%, and 1.9%, respectively, compared to the original model. The improved YOLOv5 was also compared with other models such as YOLOv3, Faster-RCNN, and YOLOv4-tiny. The results indicated that the performance of the proposed model surpassed those of the other conventional models. This study achieved the detection of multiple targets on different scales in a polar region and realized data fusion, avoiding the limitations of using a single data source, and provides a method to support polar ship path planning.

Keywords: computer vision; single-shot detector (SSD); You Only Look Once (YOLOv5); multi-source data; polar object; remote sensing image; sea ice

1. Introduction

Ice along the Arctic shipping waterways is gradually thawing under the influence of global warming, and new shipping routes to polar areas are becoming available [1]. This could greatly reduce the navigation time and increase safety [2]. Glacial surges, fog, and ice flow will affect the navigation safety and may result in collisions with ice and ship damage. Therefore, it is important to promptly detect sea ice, icebergs, and passing ships to avoid ship–ice and ship–ship collisions [3]. A detection system should provide information about the position and size of the objects on navigation routes, so as to support polar ship path planning and make ship navigation safer and more energy efficient.

Field observation focuses mostly on ships and buoys. As described, visual observation was combined with field measurements [4], determining for instance, ice thickness through the on-site drilling of ice samples. However, on-site detection in the harsh polar environment is challenging, and data collection is limited [5,6]. In recent years, image processing and remote sensing technology have been applied to the acquisition of polar information,

and indirect detection techniques have been developed [7]. Methods such as ship walk observation, shipborne radar observation, and unmanned aircraft observation are used for local-scale detection, while active and passive microwave remote sensing is mainly used for large-scale observations [8].

For local-scale environmental information, shipboard cameras are commonly used to acquire and analyze optical images. Weissling et al. [9] developed a ship-based, ice condition imagery acquisition, processing, and analysis system. Worby et al. [10] evaluated the ice distribution characteristics in the Antarctic based on 20,000 images acquired during Antarctic ship voyages. In addition, researchers are studying how to apply machine learning and deep learning to polar target detection. Li et al. [11] proposed a two-stream radiative transfer model for ponded sea ice. The upwelling irradiance from the pond surface was determined and then its spectrum was transformed into RGB color space. Cai et al. [12] employed convolutional neural networks to detect sea ice by instance segmentation using a simulation ice pool dataset and estimated ice size and concentration.

For large-scale ice detection, passive and active microwave remote sensing images are mostly used. Some algorithms for calculating ice concentration were proposed, including NASA Team (National Aeronautics and Space Administration), Bootstrap, and ASI (ARTIST Sea Ice) [13]. For the identification and classification of sea ice, techniques such as the maximum likelihood method, SVM (support vector machines), Markov random field model, and neural networks have been utilized. Belchansky et al. [14] used SSM/I (Special Sensor Microwave/Image) bright temperature data and remote sensing ice images acquired by the ERS and Okean satellites as inputs to train neural networks. Karvoven et al. [15] segmented and classified six types of ice from Synthetic Aperture Radar (SAR) images using an impulse-coupled neural network. Ressel et al. [16] utilized an artificial neural network to classify ice, and the results demonstrated that the method was resistant to image noise. However, generally, the models used were not modified and improved according to the characteristics of the remote sensing ice images to be analyzed.

The detection based on shipboard optical images is characterized by high resolution, rapidity, and the ability to provide rich information [17], but it cannot allow a continuous monitoring of the environment and is affected by adverse weather conditions. The detection based on remote sensing images can be applied to wide polar regions and is independent of the weather conditions, but its spatial distribution is relatively low, and it is not sufficiently accurate to distinguish small targets. Most studies focused on ship detection rather than on ice detection, and those that investigated ice detection systems mainly used a single data source consisting of remote sensing or optical images.

In this paper, we combined data of local-scale optical images and remote sensing images to integrate their specific strengths. Polar datasets at different scales were constructed. The SSD model was used for polar target detection at the local scale. For remote sensing detection, the YOLOv5 model was improved according to the characteristics of the sea ice, and ablation and comparison experiments were conducted to verify the model. We performed a slicing operation on the images to ensure that small sea ice targets could be detected and we constructed hybrid datasets to verify the proposed model.

2. Polar Multi-Target Detection at the Local Scale

2.1. Target Detection

The region proposal method and the end-to-end method are based on two primary detection deep learning algorithms. Overfeat, R-CNN (Region-CNN), Faster R-CNN [18], etc., are involved in the region proposal-based method while YOLO and SSD are part of the end-to-end-based method [19]. The region proposal-based method has a significant advantage in detection accuracy with respect to the end-to-end-based method because it includes "two steps" and is more accurate for target localization and classification. On the other hand, it has a significant disadvantage in the detection of speed because it requires a long time to generate the region proposal. The end-to-end-based detection method directly extracts features for object localization and classification using convolution. The SSD relies

on the RPN (Region Proposal Network) mechanism of the Faster R-CNN, which combines the detection speed of the end-to-end method with the detection accuracy of the region nomination method. Therefore, in this paper, we chose the SSD model for polar multi-target detection on the local scale.

2.2. SSD Model

The SSD model consists of two major components, a base network and additional network layers, as shown in Figure 1. The base network uses the structure of Visual Geometry Group (VGG 16) and converts the last two fully connected layers into convolutional layers, Conv4_3 and Fc7. The additional network layers include four sets of convolutional layers: Conv6_2, Conv7_2, Conv8_2, and Conv9_2. The SSD detection model operations are as reported below.
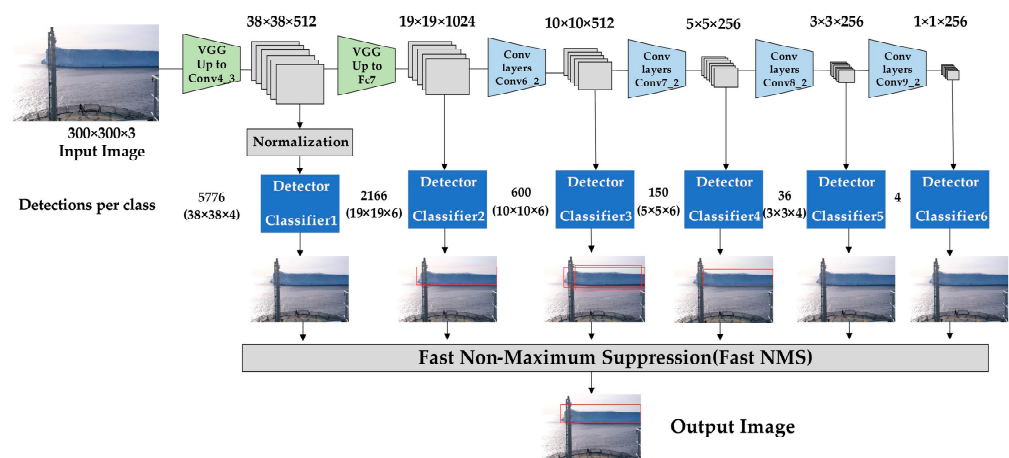

Figure 1. Structure of the SSD model.

Firstly, the input image is converted to a three-channel RGB (Red Green Blue) image with a resolution of 300×300 or 500×500. The image is fed into the network to extract multi-scale feature information, and the scales of each feature layer are 38×38, 19×19, 10×10, 5×5, 3×3 and 1×1.

Then, target feature extraction is performed through six feature layers of different scales. Default boxes are generated for each point of the feature map, and the number of default boxes is different for each layer.

Finally, all the generated default boxes are integrated, analyzed by non-maximum suppression (NMS) and filtered with an intersection over union (IOU) higher than 0.5. The final output contains information about the location, category, and confidence level of the target.

The SSD is characterized by its efficiency as a single-stage detector, performing detection directly in a single forward pass without the need for region proposals, which results in a faster detection compared to other models. It leverages multi-scale features and default boxes and can detect objects of various sizes. These advantages make SSD an effective model widely applied in practical scenarios.

### 2.3. Construction of a Local-Scale Polar Multi-Target Dataset

Due to the lack of a publicly accessible dataset for polar targets, constructing a new dataset is an important step. A total of 650 images were obtained through searching, deweighting, annotation, and review to create a local-scale polar multi-objective dataset. Some of the images were downloaded from The Norwegian Institute (https://icewatch.met.no, accessed on 19 August 2022). The dataset was divided into 5 categories, namely, sea ice (first-year ice), icebreakers, icebergs, inter-ice waterways, and melting pools on ice. Labellmg, an image annotation tool, was used to label the images as fy, icebreaker, iceberg, channel, and

pool, respectively [20,21]. Finally, the dataset is randomly divided into training and testing sets at the ratio of 8:2. The details are shown in Table 1.

Table 1. Local-scale polar multi-target dataset.

Category Name	Label Name	Image Number	Label Number
Sea ice	fy	150	2446
Icebreakers	icebreaker	150	160
Icebergs	iceberg	150	167
Inter-Ice Waterways	channel	100	100
Melting pools on ice	pool	100	558

The majority of the images were captured by shipboard cameras and UAVs (Unmanned Aerial Vehicles), and the photographed scenes corresponded to polar ship navigation scenarios. Some of the sample images are shown in Figure 2.

(a) (b)

(c) (d)
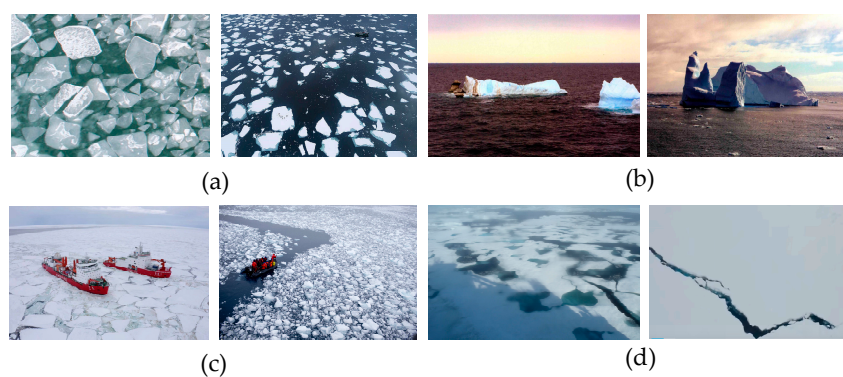
Figure 2. Local-scale polar multi-target dataset. (**a**) Sea ice; (**b**) icebergs; (**c**) icebreakers; (**d**) melt pond and inter-ice waterway.

2.4. Results

The model training and testing configurations are shown in Table 2. The detailed training parameters are shown in Table 3.

Table 2. Configurations and versions.

Configuration	Version
Operating System	Window10
Central processing unit CPU	Intel Xeon W-2255
Graphics GPU	NVIDIA Quadro P620
Deep Learning Platform	Pytorch
Pytorch version	1.10.2
CUDA version	11.3
CUDNN version	8.2.1
Python version	3.9

Table 3. Training parameters.

Parameters	Values
num_calsses	4
learning_rate_base	0.002
batch_size	4
momentum	0.937
num_workers	4
epoch	1000
weight_decay	0.0005

The steps in the training were as follows. Firstly, the training process was mainly used to predict the results and calculate the loss value by the forward propagation algorithm. Secondly, the parameter gradient value was calculated by backward propagation, and the parameters were optimized and updated. Finally, the training was completed by iterating the gradient descent algorithm to the maximum number of iterations. The process was stopped when the model reached loss convergence; then, a model was generated for the subsequent training and target detection tasks.

Average precision (AP), F1 score, and mean average precision (mAP) were determined to evaluate the detection accuracy [22]. The precision (P) value can quantify the effectiveness of sample classification, and the recall (R) value can evaluate the capacity to detect positive samples. Considering only precision or only recall is not sufficient to evaluate a model; so, the F1 score was used to harmonize P and R. The calculation of mAP can be divided into two steps: the first step consists of the calculation of the AP (average precision) of each category, while the second step involves determining the sum of the average precision values of each category and then its average value to obtain mAP. These parameters were calculated according to Equations (1)–(5):

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{F1} = 2\frac{P \cdot R}{P + R} \tag{3}$$

$$\text{AP} = \int_0^1 P \cdot RdR \tag{4}$$

$$\text{mAP} = \frac{\sum_{i=1}^{k} \text{AP}_i}{k} \tag{5}$$

where TP (true positives) is the number of correctly classified positive samples, FP (false positives) is the number of incorrectly classified positive samples, TN (true negatives) is the number of correctly classified negative samples, and FN (false negatives) is the number of incorrectly classified negative samples; k is the category number.

After training, the model was applied to the test set, and finally, an mAP value of 70.19% was obtained. The accuracy of the icebreaker category was the highest at 92%, followed by those of the iceberg category, which was 85%, and of the fy (first-year ice) category, which reached 77%. The accuracies of channel and pool were the lowest, 52% and 45%, respectively, due to the low number of images or labels for these two categories. The detection results for each category are shown in Figure 3. Some of the test results are shown in Figure 4. The SSD model works well for the detection of large targets at close range, but it is not effective in detecting small targets at a distance.
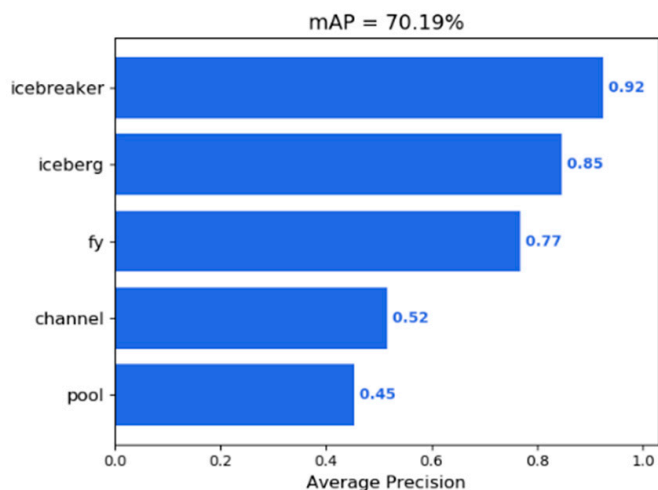
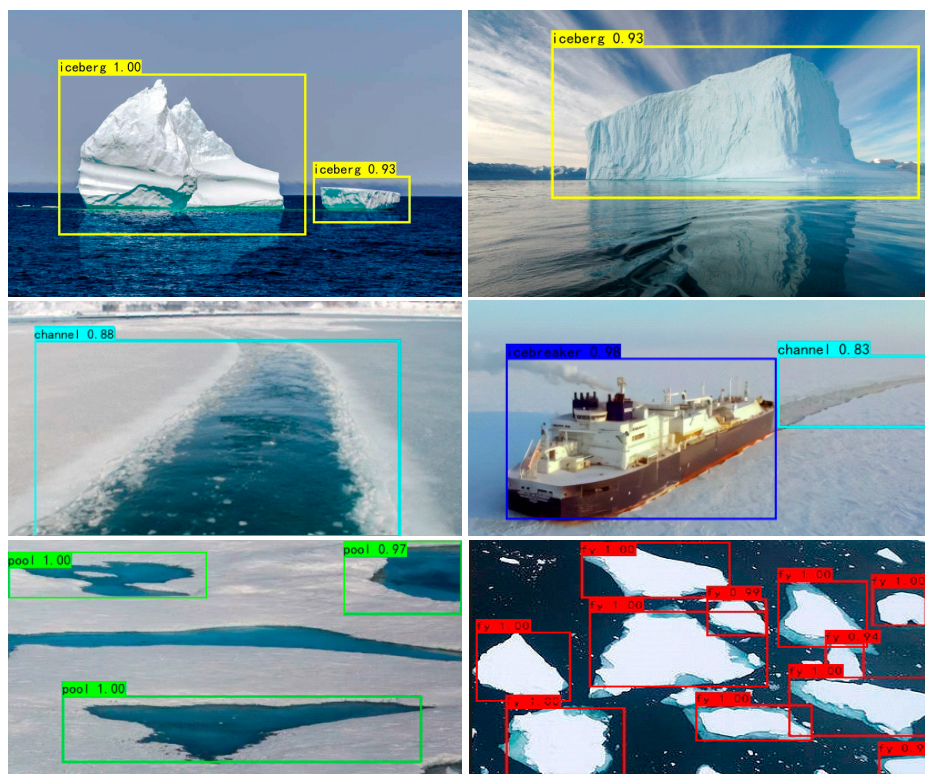**Figure 3.** Detection results for each category.

Figure 4. Example of the test results.

3. Sea Ice Detection by Remote Sensing

The detection on a local scale does not fully meet the requirements of navigating in polar regions, and using a single data source has certain limitations. The ship optical cameras cannot obtain large-scale and long-time series images and cannot monitor non-navigable areas. If sea ice in remote sensing images can be identified and located, and data fusion between local-scale and large-scale data can be performed, the advantages of different data sources can be fully utilized [23].

3.1. Introduction of the YOLOv5 Model

In remote sensing images, the ice masses appear very small and densely clustered, and the SSD model is not able to analyze them. After improving its accuracy and efficiency, the YOLOv5 model was applied to the detection of ice through remote sensing. The backbone,

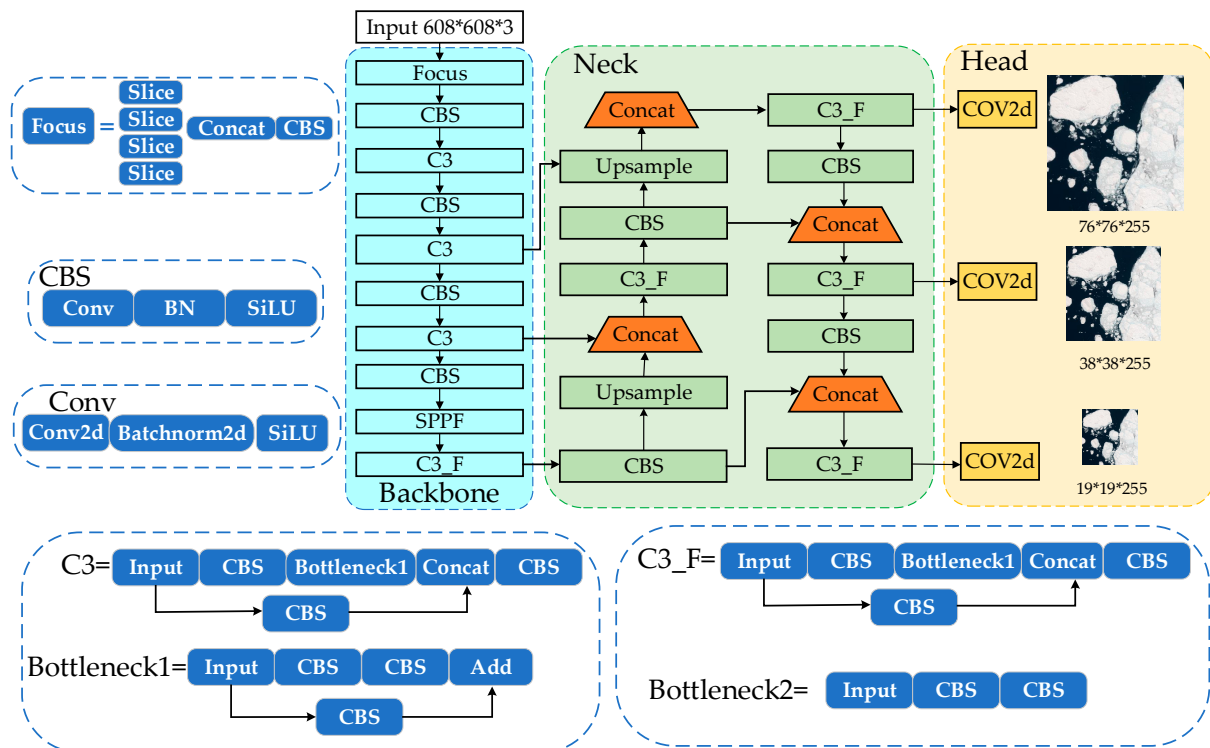neck, and head are the three basic structural components of the YOLOv5 model, as shown in Figure 5.

Figure 5. Structure of the YOLOv5 model.

The YOLOv5 backbone utilizes CSPDarknet as the backbone for extracting features from images, which is composed of cross stage partial networks. The focal module is responsible for efficiently downsampling the images. It is designed to transmit the images through the channel while maintaining primitive information. The backbone layer incorporates the utilization of the C3, C3_F, and Spatial Pyramid Pooling Fast (SPPF) modules. The C3 and C3_F modules can enhance the extraction of image features and augment the overall speed.

The neck module in YOLOv5 utilizes PANet to produce a feature pyramid network. These aggregated features are subsequently forwarded to the head module for prediction. The neck layer integrates the structures of the feature pyramid network (FPN) and the path aggregation network (PAN). Deep-feature images possess a higher degree of semantic information but a lower degree of location information, whereas shallow-feature images exhibit the reverse characteristics. The FPN model can transmit semantic information from a deep-feature image to a shallow-feature image. In contrast, PAN can transmit location information from a shallow-feature image to a deep-feature image. The integration of FPN and PAN enables the consolidation of parameters across various detection layers.

The YOLOv5 head is composed of layers that produce predictions from the anchor box. The head can be categorized into the loss function and non-maximum suppression (NMS). The binary cross entropy loss function is employed for the computation of classification loss and confidence loss, whereas the complete IoU (CIoU) loss function is utilized for the estimation of location loss. The CioU loss function incorporates three crucial parameters: the overlap area, the distance from the center, and the aspect ratio. NMS is employed to eliminate redundant detection while retaining the candidate box with the highest prediction probability as the ultimate prediction box.

3.2. Improved YOLOv5 Model

The YOLOv5 model was improved in three aspects. Firstly, the Squeeze-and-Excitation Networks (SE) attention module was added to the backbone of the original model. Secondly, the Fast Spatial Pyramid Pooling and Cross Stage Partial Network module (SPPCSPC-F) were used to augment the characterization capabilities. Finally, Funnel Activation (FReLU) was introduced to replace the Sigmoid-Weighted Linear Unit (SiLU) and improve the accuracy of ice detection.

3.2.1. Squeeze-and-Excitation Networks (SE) Attention Mechanism

Due to the large size of the remote sensing images and the small size of the ice targets, it is easy to lose some useful information. The Squeeze-and-Excitation Networks (SE) attention mechanism was added to the YOLOv5 backbone [24]. The SE module was inserted after the convolutional layers. The module consists of two operations: squeeze and excitation. It is integrated to adaptively adjust the importance of each channel by learning their weights. The structure of the SE attention mechanism is shown in Figure 6.
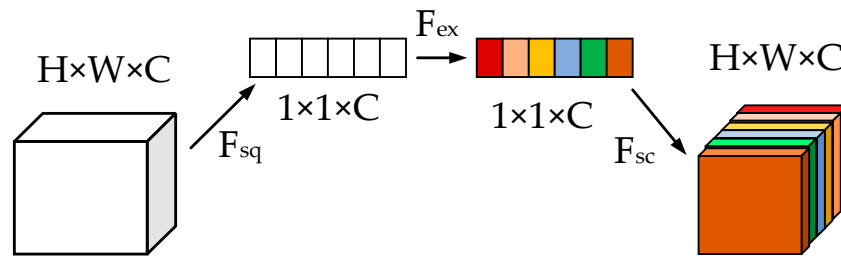

Figure 6. Structure of the SE attention mechanism.

In the squeeze phase (Fsq), global average pooling is applied to the input feature map, compressing it from three dimensions to one dimension. This one-dimensional tensor captures global information for each channel. In the excitation phase (Fex), a set of fully connected layers operates on the output of the squeeze phase. These layers model the importance of each channel and generate a channel attention vector. Finally, a rescale operation (Fsc) normalizes the weights and multiplies them onto each feature channel.

3.2.2. SPPCSPC-F

The Spatial Pyramid Pooling Fast (SPPF) is a module designed to enhance feature representation. SPPF is the improved version of Spatial Pyramid Pooling (SPP) and is faster than SPP under the same conditions. The structure of the SPPF module is shown in Figure 7.
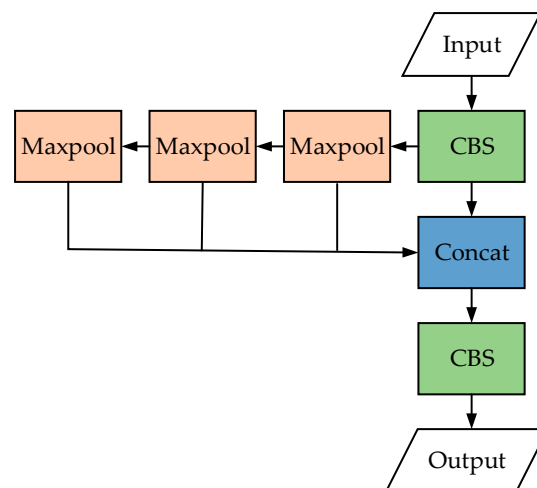

Figure 7. Structure of the SPPF module.

The input feature map passes through three 5×5 maximum pooling layers, and three different sizes of receptive fields are obtained. Although maximum pooling can expand the receptive field, it will reduce the resolution of the feature map and cause the loss of some useful information. SPPCSPC is a structural module that combines the concepts of SPP and Cross Stage Partial Network (CSP) [25]. In this paper, we present the SPPCSPC-F to replace the SPPF concerning the idea of SPPCSPC. The structure of SPPCSPC-F is shown in Figure 8.
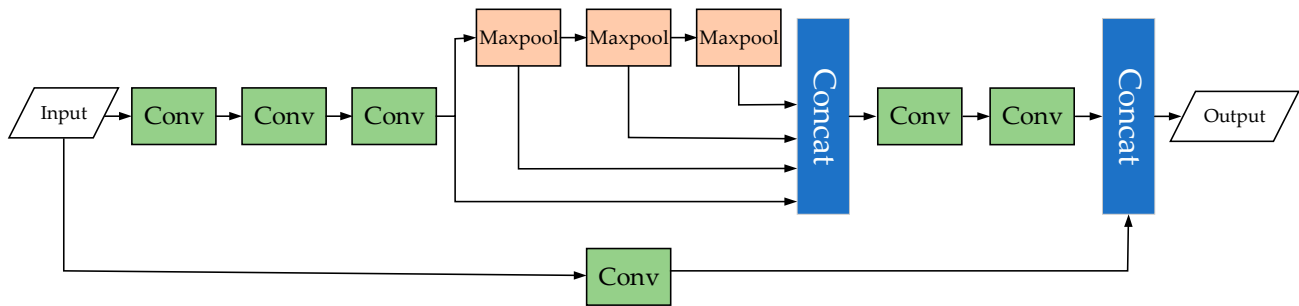

Figure 8. Structure of the SPPCSPC-F module.

The input feature map is passed through the SPPCSPC-F module, with one path performing convolutional operations to extract lower-level features, and the other path preserving the original features. Next, the module performs multi-scale pooling operations on the feature map to capture features with different receptive fields. Finally, the fused features are further processed by subsequent convolutional layers. The order of pooling is modified to increase the speed while keeping the feeling field constant.
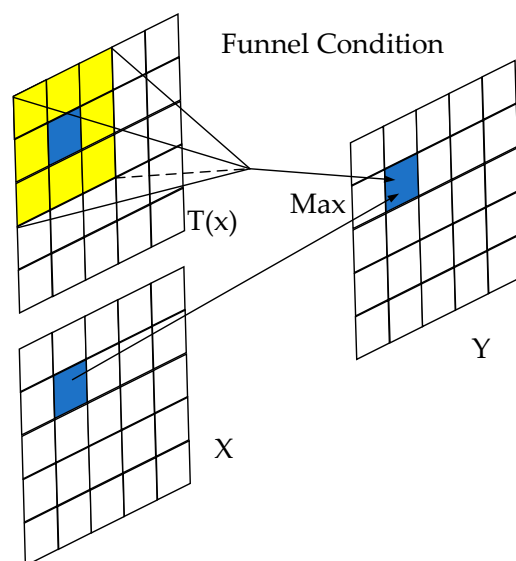
3.2.3. FReLU Activation Function

In the YOLOv5, the Sigmoid-Weighted Linear Unit (SiLU) is used as the activation function. When the input values move away from zero, the derivative of the SiLU can approach zero, leading to gradient saturation. It is difficult for the network to converge or cause training instability. The FReLU was used to replace the SiLU. The FReLU activation function incorporates learnable parameters, enabling the network to adaptively adjust the shape of the activation function through learning [26]. This flexibility enhanced the model's learning capacity and improved its adaptation to the sea ice characteristics. Combining SE attention with FReLU enables YOLOv5 to extract high-quality features, concentrate on key objects, reduce overfitting, and improve generalization ability, especially for detecting small objects in polar regions. The FReLU is defined by Equations (6) and (7):

$$f(x_c, i, j) = \max(x_c, i, j, T(x_c, i, j)) \tag{6}$$

$$T(x_c, i, j) = x_{c,i,j}^w \cdot p_c^w \tag{7}$$

where $T(\cdot)$ denotes the funnel condition, $x_{c,i,j}^w$ denotes a $k_h \times k_w$ Parametric Pooling Window centered on $x_{c,i,j}^w$, p_c^w denotes the coefficient on this window which is shared in the same channel, and $(\cdot)$ denotes dot multiply. The FReLU activation function is shown in Figure 9.

Figure 9. FReLU activation function.

### 3.3. Construction of a Remote Sensing Sea Ice Dataset

The remote sensing sea ice dataset was mainly derived from the Google Earth (http://earthengine.google.com/, accessed on 25 December 2022) and the Northwestern Polytechnical University (NWPU) datasets [27]. A total of 600 images, obtained after deduplication, annotation, and review, constituted the dataset. The tag name was ice, and the number of tags was 15,948. It was randomly divided into a training set and a test set at the data ratio of 8:2. Some of the sample images in the dataset are shown in Figure 10.
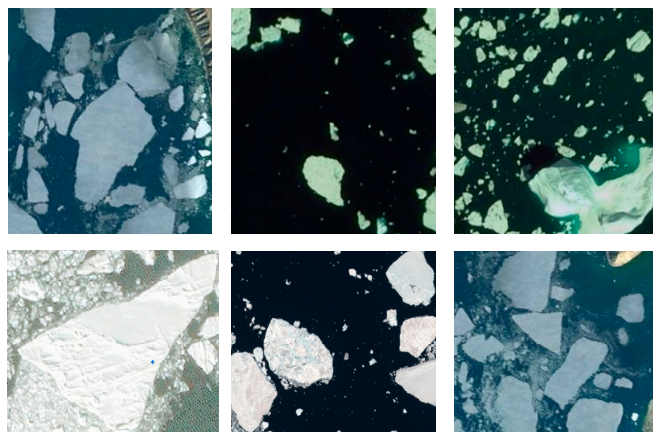
Figure 10. Remote sensing sea ice dataset.

Neural networks need a large amount of data and a high data quality to improve their performance and robustness. The YOLOv5 uses Mosaic, adaptive cutout, and other data processing methods for data enhancement [28].

The main idea of Mosaic is to randomly crop and scale several images and then randomly arrange and splice them to form a single image, to enrich the dataset and improve the training speed of the network. In the normalization operation, several images are calculated at one time, which can reduce the demand for computer memory. The data augmentation process is shown in Figure 11.

(a)

(b)
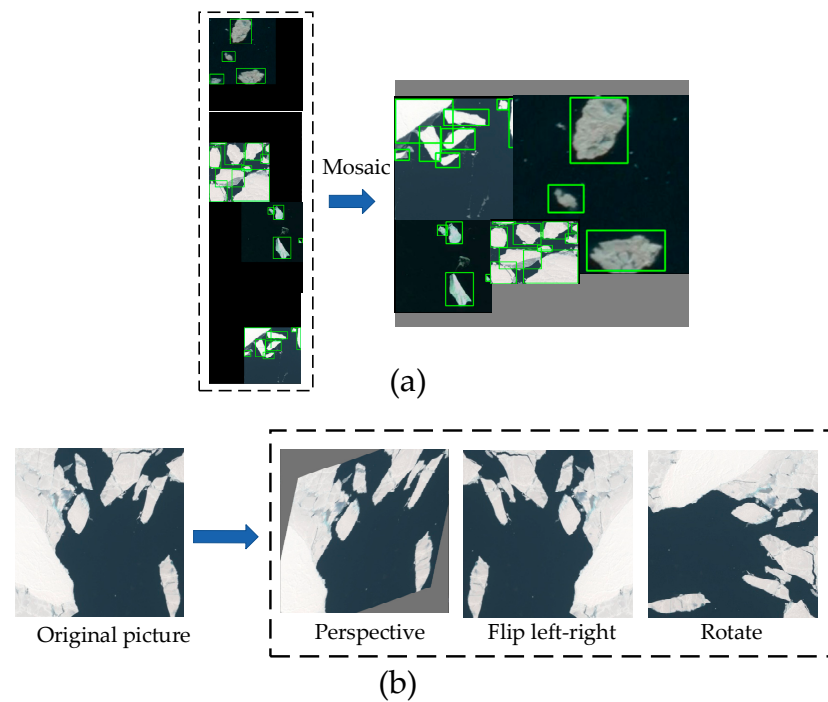
Figure 11. Data augmentation. (**a**) Mosaic; (**b**) perspective, flip left–right and rotation processing.

There are many challenges in the detection of remote sensing images, as some targets are relatively small in size and usually clustered together. If the images are directly sent into the network for detection, many small targets cannot be effectively identified.

To solve this problem, in the detection stage, a sliding window was used to cut a specified-size (such as a 416×416) image as the input. The cutout adjacent images had a 15% overlap. The slicing operation on the remote sensing image is shown in Figure 12. The purpose of the overlap is to ensure that every region is completely detected. Although this causes duplicate detection, overlapping sections can be filtered out by the NMS. Finally, the results of each cutout image were combined to obtain the detection results.
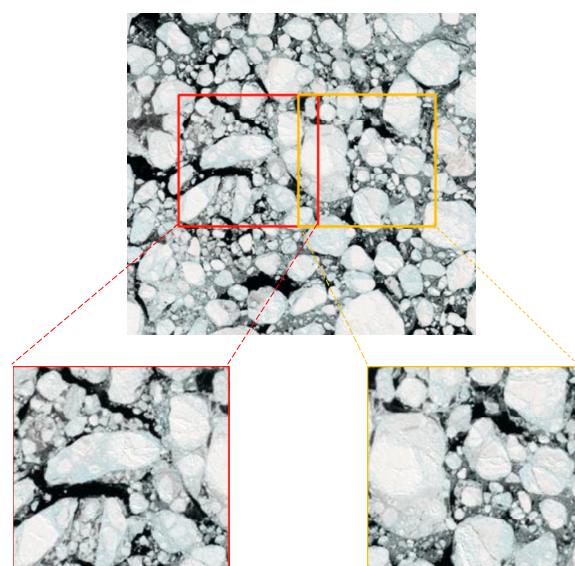
Figure 12. Slicing operation.

In order to verify the accuracy of the improved YOLOv5 model, we combined simulated sea ice images and real sea ice images into a hybrid dataset. The simulated images

were constructed as follows. Firstly, we built a large flat ice field. Secondly, we fragmented the flat ice field to obtain a broken ice field. The Voronoi diagram is morphologically similar to an ice field with large pieces of broken ice and consists of a set of continuous polygons formed by the perpendicular bisectors of lines connecting two neighboring points. We used the RayFire plug-in of 3ds Max to fragment the flat ice field according to the Voronoi diagram, as shown in Figure 13a. Finally, the size of the broken ice field was reduced by 80% to enlarge the gaps between the ice blocks, as shown in Figure 13b.

(a)　　　　　　　　　　　　　(b)
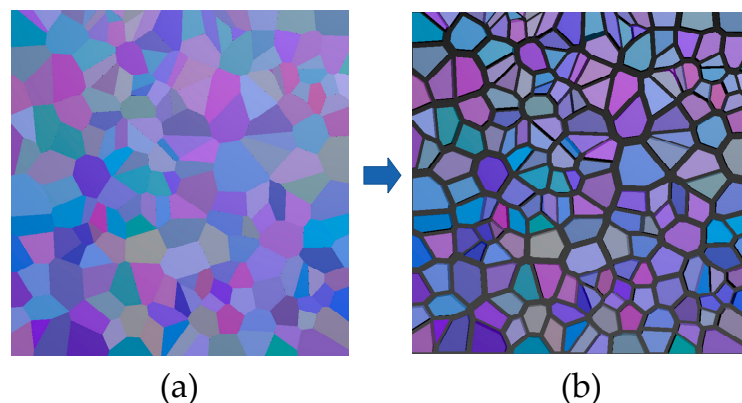
Figure 13. Sea ice modeling process. (**a**) Broken sea ice field; (**b**) operation of enlarging the gaps between the ice blocks.

3.4. Results

3.4.1. Ablation Study

The ablation study was conducted to facilitate the comparison of the different improvement methods. They were trained with the same configuration used in the local-scale polar objection. The epoch was set as 300, the initial learning rate was 0.001, the momentum parameter was 0.9, the weight decay parameter was 0.0005, and the NMS threshold was 0.5. The evaluation was carried out after every 30 training epochs. The results are shown in Table 4.

Table 4. Ablation study results.

Method	P	R	F1	mAP
YOLOv5	0.719	0.684	0.701	0.719
YOLOv5+SE	0.731	0.701	0.716	0.738
YOLOv5+CSPCF	0.737	0.688	0.712	0.743
YOLOv5+FReLU	0.723	0.706	0.714	0.747
YOLOv5+SE+SPPCSPC-F+FReLU	0.753	0.703	0.727	0.754

In Table 5, it can be observed that the mAP of the original YOLOv5 model was 0.719, the lowest among those of the evaluated models. The implementation of SE resulted in an increase in the mAP to 0.738, i.e., by 1.9%. The inclusion of SPPCSPC-F resulted in a 2.4% increase in the mAP, which reached the value of 0.743. However, the R value was relatively low, i.e., 0.688. The inclusion of FReLU resulted in a 2.8% increase in the map, to the value of 0.747. When adding SE, SPPCSPC-F, and FReLU, the mAP was improved by 3.5%, reaching the highest value among those of all the examined models.

Similarly, the P, R, and F1-scores of the original YOLOv5 model were 0.719, 0.684, and 0.701. However, for the proposed method, the P, R, and F1-scores were 0.753, 0.703, and 0.727, that is, they increased by 3.4%, 1.9%, and 1.8%, respectively. Therefore, the improved YOLOv5 model revealed superior accuracy and enhanced performance in the domain of remote sensing sea ice detection.

Table 5. The performance of different models in sea ice detection.

Method	P	R	F1	mAP
YOLOv3	0.858	0.407	0.552	0.604
YOLOv4-tiny	0.757	0.548	0.636	0.648
Faster-RCNN	0.641	0.632	0.636	0.655
YOLOv5	0.719	0.684	0.701	0.719
Ours	0.753	0.703	0.727	0.754

3.4.2. Contrast Study

In order to further validate the advantage benefits and efficacy of the improved YOLOv5 model, incorporating the three mentioned modules, a comparative experiment was conducted. We compared the improved model with other conventional models, such as Faster-RCNN, YOLOv3, and YOLOv4-tiny; the values of loss and mAP are shown in Figure 14.
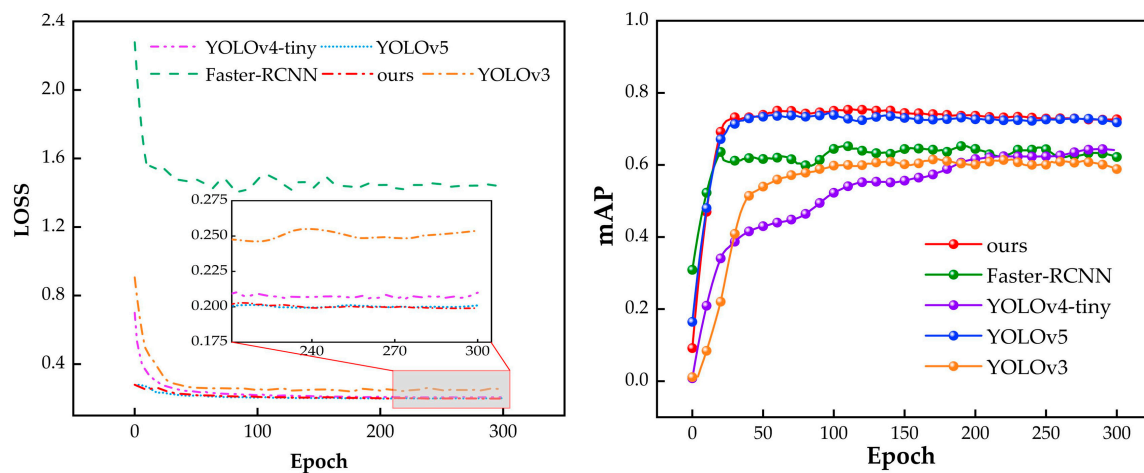

Figure 14. The values of loss and mAP of the different models tested.

In the first 40 epochs, the loss of each model fell quickly, indicating that the training did not achieve a stable state. When the training is stable, the loss in the curve is flat rather than sharp. The loss of our model was lower than that of the others when training reached a steady stage. The mAP rose sharply in the first 80 epochs. All models tended to become more stable after 250 training epochs, and the mAP of our model was the highest.

The values of the evaluation indicators are shown in Table 5. Compared with those of YOLOv3, YOLOv4-tiny, Faster-RCNN, and original YOLOv5, the mAP of YOLOv3 was the lowest, at 60.4%, whereas the mAP of our model was the highest, at 75.4%. YOLOv3 and YOLOv4-tiny showed a higher P value but a lower R value, which indicated that these two models largely miss their ice targets when detecting sea ice. Based on the above results, the improved YOLOv5 can better perform in sea ice detection.

The improved YOLOv5 was used to test a remote sensing image with a resolution of 3660 × 3660. Since some sea ice targets were too dense, the confidence degree was hidden in the results. The detection results of the original YOLOv5 are shown in Figure 15a. Figure 15b shows zoomed-in local images using the original YOLOv5, in which the number of detected sea ice masses was 14 and 55. Figure 15c shows zoomed-in local views of the image detected by the improved YOLOv5, in which the number of detected sea ice masses was 53 and 88. When using the improved YOLOv5, the number of detected ice targets increased by 39 and 33 units, and most of them were small.
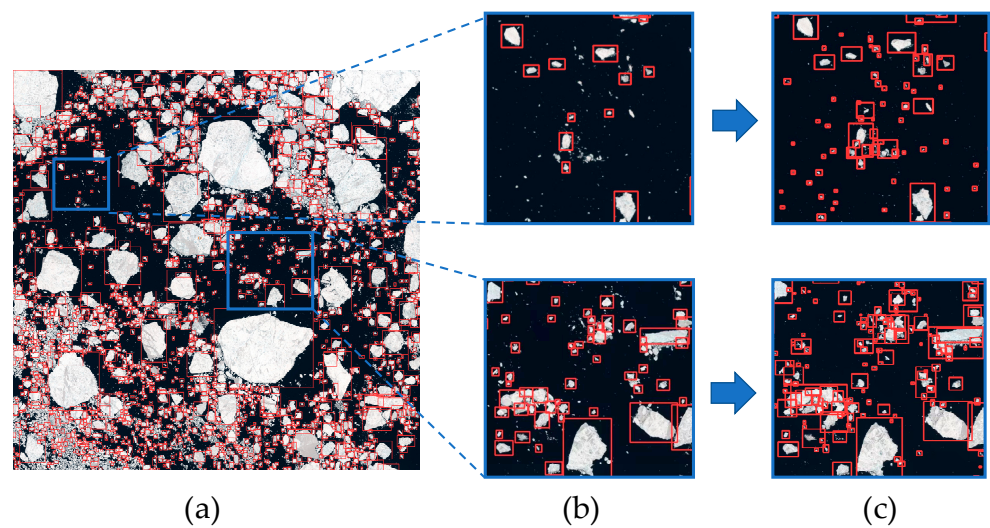
Figure 15. Comparison of the detection results. (**a**) Detection results for the original YOLOv5; (**b**) localized zoomed-in views using the original YOLOv5; (**c**) localized zoomed-in view using the improved YOLOv5.

The results with the confidence degree are shown in Figure 16. Both a real image and a simulated sea ice image are presented. The results demonstrated that the improved YOLOv5 model was able to detect ice targets in simulated sea ice images with strong generalization ability and robustness.
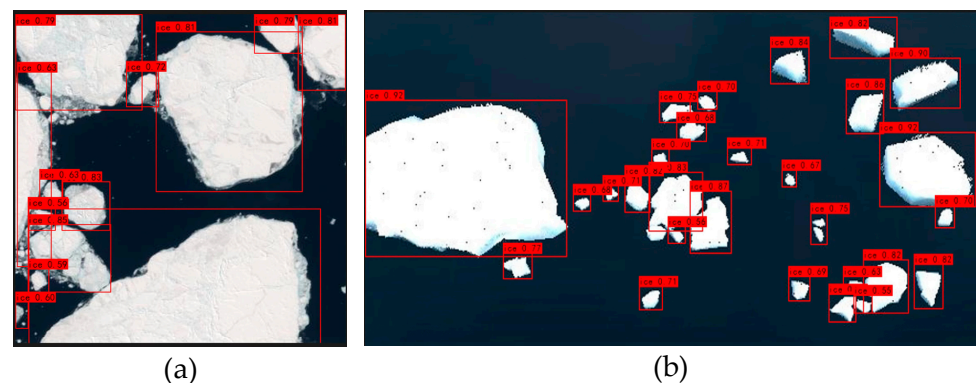
Figure 16. Test detection results using the hybrid dataset. (**a**) Detection result for a real ice image; (**b**) detection result for a simulated ice image.

Local scale detection covers from tens to hundreds of meters. Correspondingly, remote sensing scale detection covers from tens to hundreds of kilometers. If a ship navigates in the polar regions using only local-scale data, the planned path may be optimal at the local scale but not on the whole, as it could be unnecessarily long. If only remote sensing data are used, the planned path may be the best on a large scale, but it may miss some obstacles that will jeopardize the safety of ship navigation on a local scale. In this paper, local-scale and remote sensing data were combined to take advantage of their respective strengths. Our results indicated that the use of this combination for the detection of obstacles can improve the safety and efficiency of polar navigation.

4. Discussion

The instability of polar condition makes navigation difficult. Sea ices which float on the surface are difficult to detect and are prone to collision with the hull or the propeller. In this paper, polar datasets at different scales were constructed. The SSD model was used for multi-target detection at the local scale. For remote sensing images, hybrid datasets were

constructed and a slicing operation was performed, the YOLOv5 model was improved and tailored to detect sea ices. Ablation and comparison experiments were conducted to verify the proposed model.

For the source of data, most studies mainly adopt remote sensing or optical image as a single data source. For example, Li et al. [29] who developed a novel method to extract sea ice cover using Sentinel-1 data based on the support vector machine (SVM). Xu et al. [30] proposed a Recurrent Attention Convolutional Neural Network (RA-CNN) to classify different ships. In this paper, the fusion of remote sensing and optical images is used to take advantage of the complementary strengths.

For ice detection, some studies did not change their model according to the characteristic of ices. Moreover, many studies used only real datasets to verify the accuracy of their model. For example, Frederik et al. [31] proposed a deep learning model based on YOLOv3 for distinguishing icebergs and ships. Markus et al. [32] detected the ice on rotor blades. In this paper, the YOLOv5 mode was improved to ensure that small ices can be detected. The hybrid dataset was constructed to verify the proposed model and the results showed that the model had a good generalization ability.

Although this study successfully detected multi-scale polar objects, it still has some limitations. The lower detection accuracy of some categories on the local scale was due to the small amount of data. The datasets used can be expanded to increase the accuracy [33]. This study focused on rectangular detection boxes; if more detailed sea ice information is needed, in the future, the ice images can be processed with instance segmentation [34].

5. Conclusions

In order to avoid the limitations caused by the use of a single dataset, we constructed multi-scale datasets by combining data from different sources. The SSD model was used to detect local-scale targets, and the improved YOLOv5 model was used to detect remote sensing sea ice targets. The following conclusions can be drawn:

1. The SSD model can be used for the detection of polar targets on a local scale. The dataset it uses includes sea ice, icebergs, icebreakers, ice melt ponds, and inter-ice waterways; the mAP can reach 70.19%; icebergs and icebreakers were detected with the highest average accuracy of 84% and 81%.
2. An improved YOLOv5 model was obtained through Squeeze-and-Excitation Networks (SE), Funnel Activation (FReLU), Fast Spatial Pyramid Pooling, and Cross Stage Partial Network (SPPCSPC-F). The utilization of SE and SPPCSPC-F allowed the characteristics of objects to be strengthened, thereby augmenting the overall detection efficacy and precision of the model. The FReLU activation function was used to enhance the learning capacity and enable better adaptation to sea ice characteristics. A slicing operation was performed on remote sensing images to detect small ice masses. Simulated ice images were included to verify the precision of the proposed model.
3. In comparison to other conventional models such as Faster-RCNN, YOLOv3, and YOLOv4-tiny, the proposed model demonstrated higher accuracy, with an mAP of up to 75.4%, which verified its generalization ability and robustness. The proposed method is tailored to detect remote sensing sea ice, compared to the original model, the mAP value increased 3.5%.
4. For future research, large and diverse polar datasets need to be established. These datasets should contain polar images from different seasons, weather conditions, and periods, so that the model can better adapt to changes in the polar environment. Additionally, improved detection can provide support to avoid polar ship collision with ice masses and improve navigation path planning. It can also provide some help for the calculation of the ice pressure load for ships [35,36].

References

1. Chuah, L.F.; Mokhtar, K.; Ruslan, S.M.M.; Abu Bakar, A.; Abdullah, M.A.; Osman, N.H.; Bokhari, A.; Mubashir, M.; Show, P.L. Implementation of the energy efficiency existing ship index and carbon intensity indicator on domestic ship for marine environmental protection. *Environ. Res.* **2023**, *222*, 115348. [CrossRef] [PubMed]
2. Zuo, Q.; Qian, L.; Xu, X.; Yan, J.; Cheng, L.; Zhang, Z. Navigation strategy and economic research of the northeast passage in the Arctic. *Chin. J. Polar Res.* **2015**, *27*, 203.
3. Lu, Y.; Gu, Z.; Liu, S.; Chuang, Z.; Li, Z.; Li, C. Scenario-based optimization design of icebreaking bow for polar navigation. *Ocean Eng.* **2021**, *244*, 110365. [CrossRef]
4. Yu, M.; Lu, P.; Li, Z.; Li, Z.; Wang, Q.; Cao, X.; Chen, X. Sea ice conditions and navigability through the Northeast Passage in the past 40 years based on remote-sensing data. *Int. J. Digit. Earth* **2020**, *14*, 555–574. [CrossRef]
5. Oloruntobi, O.; Mokhtar, K.; Gohari, A.; Asif, S.; Chuah, L.F. Sustainable transition towards greener and cleaner seaborne shipping industry: Challenges and opportunities. *Clean. Eng. Technol.* **2023**, *13*, 100628. [CrossRef]
6. Mahadi, C.M.H.C.; Mokhtar, K.; Chuah, L.F.; Chan, S.R.; Suhrab, M.I.R.; Mubashir, M.; Asif, S.; Show, P.L. An organisational search and rescue performance assessment for a cleaner environment. *Clean. Eng. Technol.* **2023**, *14*, 100641. [CrossRef]
7. Ogishima, A.; Saiki, K. Development of a micro-ice production apparatus and NIR spectral measurements of frosted minerals for future lunar ice exploration missions. *Icarus* **2021**, *357*, 114273. [CrossRef]
8. Anderson, S. Remote Sensing of the Polar Ice Zones with HF Radar. *Remote Sens.* **2021**, *13*, 4398. [CrossRef]
9. Weissling, B.; Ackley, S.; Wagner, P.; Xie, H. EISCAM—Digital image acquisition and processing for sea ice parameters from ships. *Cold Reg. Sci. Technol.* **2009**, *57*, 49–60. [CrossRef]
10. Worby, A.; Comiso, J. Studies of the Antarctic sea ice edge and ice extent from satellite and ship observations. *Remote. Sens. Environ.* **2004**, *92*, 98–111. [CrossRef]
11. Lu, P.; Leppäranta, M.; Cheng, B.; Li, Z.; Istomina, L.; Heygster, G. The color of melt ponds on Arctic sea ice. *Cryosphere* **2018**, *12*, 1331–1345. [CrossRef]
12. Cai, J.; Ding, S.; Zhang, Q.; Liu, R.; Zeng, D.; Zhou, L. Broken ice circumferential crack estimation via image techniques. *Ocean Eng.* **2022**, *259*, 111735. [CrossRef]
13. Shi, L.; Liu, S.; Shi, Y.; Ao, X.; Zou, B.; Wang, Q. Sea Ice Concentration Products over Polar Regions with Chinese FY3C/MWRI Data. *Remote. Sens.* **2021**, *13*, 2174. [CrossRef]
14. Belchansky, G.I.; Douglas, D.C.; Alpatsky, I.V.; Platonov, N.G. Spatial and temporal multiyear sea ice distributions in the Arctic: A neural network analysis of SSM/I data, 1988–2001. *Geophys. Res. Atmos.* **2004**, *109*, C10017. [CrossRef]
15. Karvonen, J. Baltic Sea Ice Concentration Estimation Based on C-Band Dual-Polarized SAR Data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 5558–5566. [CrossRef]
16. Ressel, R.; Frost, A.; Lehner, S. A Neural Network-Based Classification for Sea Ice Types on X-Band SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3672–3680. [CrossRef]
17. Mei, H.; Lu, P.; Wang, Q.; Cao, X.; Li, Z. Study of the spatiotemporal variations of summer sea ice thickness in the pacific arctic sector based on shipside images. *Chin. J. Polar Res.* **2021**, *33*, 37–48. [CrossRef]
18. 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. [CrossRef]
19. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016. [CrossRef]
20. Morozov, E.G.; Krechik, V.A.; Frey, D.I.; Zamshin, V.V. Currents in the Western Part of the Weddell Sea and Drift of Large Iceberg A68A. *Oceanology* **2021**, *61*, 589–601. [CrossRef]
21. Morozov, E.; Zuev, O.; Zamshin, V.; Krechik, V.; Ostroumova, S.; Frey, D. Observations of icebergs in Antarctic cruises of the R/V "Akademik Mstislav Keldysh". *Russ. J. Earth Sci.* **2022**, *22*, ES2001. [CrossRef]

22. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
23. Li, W.; Liu, L.; Zhang, J. Fusion of SAR and Optical Image for Sea Ice Extraction. *J. Ocean Univ. China* **2021**, *20*, 1440–1450. [CrossRef]
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018. [CrossRef]
25. 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, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475. [CrossRef]
26. Qiu, S.; Xu, X.; Cai, B. FReLU: Flexible Rectified Linear Units for Improving Convolutional Neural Networks. In Proceedings of the 2018 24th International Conference on Pattern Recognition, Beijing, China, 20–24 August 2018; pp. 1223–1228. [CrossRef]
27. Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [CrossRef]
28. Zhao, B.; Wu, Y.; Guan, X.; Gao, L.; Zhang, B. An Improved Aggregated-Mosaic Method for the Sparse Object Detection of Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 2602. [CrossRef]
29. Li, X.-M.; Sun, Y.; Zhang, Q. Extraction of Sea Ice Cover by Sentinel-1 SAR Based on Support Vector Machine With Unsupervised Generation of Training Data. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3040–3053. [CrossRef]
30. Xu, Z.; Sun, J.; Huo, Y. Ship images detection and classification based on convolutional neural network with multiple feature regions. *IET Signal Process.* **2022**, *16*, 707–721. [CrossRef]
31. Hass, F.S.; Arsanjani, J.J. Deep Learning for Detecting and Classifying Ocean Objects: Application of YoloV3 for Iceberg–Ship Discrimination. ISPRS Int. *J. Geo-Information* **2020**, *9*, 758. [CrossRef]
32. Kreutz, M.; Alla, A.A.; Eisenstadt, A.; Freitag, M.; Thoben, K.-D. Ice Detection on Rotor Blades of Wind Turbines using RGB Images and Convolutional Neural Networks. *Procedia CIRP* **2020**, *93*, 1292–1297. [CrossRef]
33. Wu, S.; Wang, J.; Liu, L.; Chen, D.; Lu, H.; Xu, C.; Hao, R.; Li, Z.; Wang, Q. Enhanced YOLOv5 Object Detection Algorithm for Accurate Detection of Adult Rhynchophorus ferrugineus. *Insects* **2023**, *14*, 698. [CrossRef]
34. Zhou, L.; Cai, J.; Ding, S. The Identification of Ice Floes and Calculation of Sea Ice Concentration Based on a Deep Learning Method. *Remote. Sens.* **2023**, *15*, 2663. [CrossRef]
35. Zhou, L.; Diao, F.; Sun, X.; Ding, S.; Zhu, A.; Song, M.; Han, Y. Numerical simulation of ice-breaking loads on ships in ice areas based on the circumferential cracking method. In Proceedings of the 19th China Marine (Shore) Engineering Symposium (Previous), Ningbo, China, 10 July 2019; pp. 197–202.
36. Xie, C.; Zhou, L.; Ding, S.; Liu, R.; Zheng, S. Experimental and numerical investigation on self-propulsion performance of polar merchant ship in brash ice channel. *Ocean Eng.* **2023**, *269*, 113424. [CrossRef]