

Article A Real-Time Lightweight Detection Algorithm for Deck Crew and the Use of Fishing Nets Based on Improved YOLOv5s Network

Jiaming Wang ^{1,2}, Xiangbo Yin ² and Guodong Li ^{2,*}

- ¹ College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; m210901473@st.shou.edu.cn
- ² Fishery Machinery and Instrument Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200092, China; yinxiangbo@fmiri.ac.cn
- * Correspondence: liguodong@fmiri.ac.cn

Abstract: A real-time monitoring system for the operational status of fishing vessels is an essential element for the modernization of the fishing industry. The operational status of fishing vessels can be identified by using onboard cameras to detect the deck crew and the use of fishing nets. Due to the typically limited processing capacity of shipboard equipment and the significant memory consumption of detection models, general target detection models are unable to perform real-time image detection to identify the operational status of fishing vessels. In this paper, we propose a lightweight real-time deck crew and the use of a fishing net detection method, YOLOv5s-SGC. It is based on the YOLOv5s model, which uses surveillance cameras to obtain video of fishing vessels operating at sea and enhances the dataset. YOLOv5s-SGC replaces the backbone of YOLOv5s with ShuffleNetV2, replaces the feature fusion network with a modified Generalized-FPN, and adds the CBAM attention module in front of the detection head.

Keywords: YOLOv5s; deck crew detection; fishing net detection; deep learning model lightweight

Key Contribution: (1) A significant decrease in the parameters and FLOPs of the detection network, which improves the speed of model detection. (2) Proposing A fishing vessel operation status detection scheme to reduce human supervision costs. (3) Real-time detection network for low computing power devices.

1. Introduction

China is a vast maritime nation with an abundance of marine resources, and the total value of fishery production plays a crucial role in the Chinese economy [1]. The significance of the total value of fisheries production to the agricultural sector cannot be overstated. At present, the marine economy has become one of the favorable engines of the national economy, but the rapid growth of the market demand for aquatic products has caused many fishing vessels to operate illegally during the fishing moratorium and in foreign waters, which has damaged the ecological balance of the sea and seriously undermined the sustainable and healthy development of the ecological environment of fishing waters [2]. Some fishermen in the coastal areas have been working illegally during the closed season. In addition, they are influenced by the traditional idea of "Living by the Sea" and exploit the loopholes of supervision, with "vessels with different certificates", illegal new "three-noes" vessels, and illegal fishing at sea. This is especially true during the fishing moratorium, which can severely disrupt the order of fisheries production and harm the ecological equilibrium of the ocean [3]. In recent years, as the number of fishing vessels and the degree of mechanization and automation of fishing vessels have increased, so has the intensity of fishing, and irresponsible fishing practices have caused a severe decline in marine fishery resources and a crisis in marine fishery production [4,5]. To advance the modernization of the fishing industry and improve the intelligence and information



Citation: Wang, J.; Yin, X.; Li, G. A Real-Time Lightweight Detection Algorithm for Deck Crew and the Use of Fishing Nets Based on Improved YOLOv5s Network. *Fishes* 2023, *8*, 376. https://doi.org/ 10.3390/fishes8070376

Academic Editors: Zhangying Ye, Xinjun Chen and Bin Liu

Received: 11 May 2023 Revised: 6 July 2023 Accepted: 18 July 2023 Published: 20 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). technology on fishing vessels, it is necessary to identify the operational status of fishing vessels at sea. China's maritime rights and interests can be effectively safeguarded by advancing the level of fisheries informatization management construction and eliminating the retrograde development situation [6]. Consequently, it is imperative to investigate onboard camera technology to autonomously identify fishing vessels' operational status. The onboard camera allows for the rapid and accurate identification of deck crew and fishing nets on fishing vessels, which is essential for the automatic identification of the operating status of fishing vessels. The operational status of fishing vessels is identified based on the identification results of the onboard camera.

Manual detection, satellite monitoring, and shipboard video surveillance are the primary methods for identifying the operational status of a fishing vessel. The manual inspection method is primarily applied by law enforcement officers who board fishing vessels to identify their operational status. This manual judgment method is highly accurate and can promptly find out if a fishing vessel has engaged in illegal fishing in a restricted area. But it also requires a significant amount of human labor, and when there are too many fishing boats, there may not be enough staff to monitor the operational status of all fishing boats. The primary method of satellite detection involves deploying a monitoring system for fishing vessels on the fishing vessel. This system transmits the fishing vessel's current status data to the satellite, which then conveys it to the ground-based base station. However, the data capacity on land is limited and costly, and the real-time efficacy is poor. For example, the price and communication fee of Inmarsat equipment is high, and the terminal equipment is large. In some areas such as China's large number of small vessels, the operating sea area is basically not extensive, poor economic capacity, and other characteristics, promoting the application of such fishing vessels in a large area is significantly difficult [7]. The application of intelligent fishing vessels is very difficult. Because of the abundant space for the implementation of intelligent transformation of fishing vessels, the artificial intelligence of fishing vessel operation mode identification technology based on shipboard video is considered, combined with multi-source data to achieve the monitoring of fishing gear and fishing methods, and to improve the level of intelligent control of fishing vessel compliance operations. Zhang Jiaze [8] achieved 95.35% accuracy in the behavioral recognition method of the fishing vessel by installing high-definition camera equipment at four fixed locations on a mackerel fishing vessel and building a 3-2D fusion convolutional neural network to extract and classify the behavioral features of the fishing vessel. Shuxian Wang [9] attached cameras, built a convolutional neural network, and added pooling layers, LSTM long short-term memory modules, and attention modules on Japanese mackerel fishing boats. In the behavior recognition test set of Japanese mackerel fishing vessels, they received an F1 score of 97.12%.

Deep learning algorithms have shown outstanding performance and promising application prospects in the fields of object detection and recognition [10-15]. In boat-based video identification systems, deep learning algorithms have produced greater results in terms of accuracy and real-time performance. But the following issues continue to exist: the computing platform of the fishing vessel's on-board equipment has limited computing power resources and the operating environment of the fishing vessel is complex, with issues such as light changes and field of view occlusion affecting the final detection results, yet the detection speed of complex models cannot meet the real-time requirements of the task, and the network models are too large to be deployed. Most deep learning algorithms with high accuracy require more model parameters and high computational complexity, requiring high computational power of hardware devices and slow detection speed; while deep learning models with fast detection speed are lacking in accuracy. Not much research has been performed on the lightweight detection model for deck crew and the use of fishing nets that balance detection accuracy and detection speed well. To solve this problem, many excellent lightweight and efficient network structures have been proposed such as MobileNet [16–18], EfficientNet [19], PP-LCNet [20], etc. This study proposes a real-time detection algorithm YOLOv5s-SGC based on the YOLOv5s model, using the

lightweight network ShuffleNetV2 [21] 0.5× replaces the YOLOv5s backbone network CSP-Darknet53 [22] to reduce the number of parameters and increase the speed of operation while maintaining accuracy. We used Generalized-FPN [23] to replace the original FPN + PAN in the feature fusion network and the general convolution and C3 modules in the feature fusion network were replaced with GSConv and CSP_GSC modules to further reduce the complexity of the model and the number of parameters, and finally the CBAM attention module was introduced in front of the detection layer to strengthen the feature representation capability of the network. The problem of detection accuracy degradation due to the reduced number of parameters is increased at the cost of a small amount of computation. To provide a real-time and effective detection method for deck crew and the use of fishing net detection of fishing vessels, experiments were conducted on the data set of fishing vessels out at sea operations suggested in this paper, and the detection performance was compared with other lightweight improvement methods.

2. Dataset

2.1. Fishing Vessel Operations Image Dataset

The dataset used for the experiment is self-built, and the source of the data is the video recordings of fishing vessels during their operations at sea. The video of the fishing vessel's operation for 7 days was taken by mounting a camera on the fishing vessel to capture the situation on the fishing vessel's deck. From it, 773 photos with a resolution size of 1920×1080 were extracted. The video capture devices used are two gun cameras that can capture 1080p 22 Hz images, and the captured video data are stored in a 2 TB hard disk, which supports cyclic storage.

This project placed onboard cameras on two trawlers to expand the diversity of the image dataset of fishing vessel operations. Directly on the deck, two gun-mounted cameras were placed, with camera 1 inclined upward to record the activities at the bow and camera 2 angled downward to record the activities at the deck. The collection of images of fishing vessel operations is shown in Figure 1.



(a) Onboard Camera Deploymen



(b) Image Captured by Camera 1



(c) Image Captured by Camera 2

Figure 1. Image of fishing vessel operations.

2.2. Data Labeling

The dataset was annotated using the Roboflow annotation tool, dividing the training set, validation set, and test set in the ratio of 8:0.5:1.5. In the dataset, the crew is labeled Person and the fishing nets are labeled Fishing_Net, with a total of 1531 crew members and 543 fishing nets.

2.3. Data Enhancement

This study specifically incorporates some complex environmental factors that may be encountered in the process of simulating fishing boats at sea, such as changes in light and obstruction of vision, in the construction of the dataset to identify the operational status of fishing boats more accurately. Enhancing the dataset can significantly increase the model's resilience and accuracy because these complicated aspects frequently create increased interference with the identification of fishing vessel operating status in actual applications. A total of 2319 data photos were obtained after the dataset was expanded using a combination of Gaussian noise, random rectangular occlusion, random pixel zeroing, pretzel noise, Gaussian blur, motion blur, changing color temperature, and random contrast. Of these, 1857 images were used for the training set, 345 images were used for the test set, and 117 images were used for the validation set. The data enhancement samples are shown in Figure 2.



(c) Motion Blur+ Change of Colour Temperature + Gaussian Noise + Random Contrast

Figure 2. Data enhancement samples.

3. Proposed Methods

In this study, ShuffleNetV2 is introduced to replace the original backbone network CSP-Darknet53, and deep separable convolution is used instead of traditional convolution to efficiently utilize the feature channels and network capacity while reducing the number of parameters. In the feature fusion network, GhostNet's Ghost module is introduced to reduce the number of parameters and operations of the model by converting the normal convolution operation to generate only some highly differentiated feature maps, and then transforming these feature maps based on cheap linear operations to obtain other similar feature maps. Although the above methods will significantly reduce the number of parameters of the model and speed up the operation speed of the detection network, they will also make the accuracy of detection decrease, so in order to compensate for the loss of accuracy due to the reduction of model parameters, this study introduces a CBAM attention module in front of the detection head. CBAM is an attention mechanism module that combines space and channel to extract positive and effective features in the image, thus improving the accuracy of the network.

3.1. YOLOv5s Model

The YOLOv5 model, the current dominant deep learning framework, can be used for state detection to provide real-time deck crew and the use of fishing net detection on fishing vessels. YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x are four variations of a one-stage structural target detection network that are all essentially the same structurally and have networks that go wider and deeper in that order. The YOLOv5s, the lightest model size, and the fastest detection speed is used in this study as the fundamental model due to the low computational power of hardware devices on fishing vessels. The input, backbone, neck, and head components make up the bulk of the YOLOv5 algorithm.

Although YOLOv5 performs admirably on publicly available datasets and is excellent at feature extraction and target detection, there is still room to improve the performance of its model files on devices with limited processing power. Additionally, it frequently confuses the target with the background of the fishing boat when images taken by the onboard camera are dimly lit or have uneven illumination, leading to false alarms. The network can maintain high-precision target detection while reducing the model's weight size and improving its processing speed by making lightweight modifications and incorporating attention modules, As a result, it can be deployed and perform real-time detection of deck crew and fishing net operations even on low-intelligence fishing vessels.

3.2. The Backbone Based on ShufflenetV2

The sizes of fishing nets and deck crew in the collected fishing vessel operation images in this study belong to large and medium targets in the image. The backbone network CSP-DarkNet53 of YOLOv5s has a deeper network depth and larger model capacity, which provides higher performance in terms of accuracy, but its number of parameters and computational complexity are also larger. To meet the requirements of high accuracy, real-time operation on low computing power devices, and fewer model parameters, the YOLOv5s backbone feature extraction network is replaced by the ShuffleNetV2 network. ShuffleNetV2 introduces the concept of "channel split" based on ShuffleNetV1 [24], and redesigns the basic structure module into two types: the basic unit and the downsampling unit, as shown in Figure 3.



Figure 3. ShuffleNetV2 module composition.

The primary design idea of ShuffleNet is to execute a channel shuffle operation on several channels to address the issue of non-communication between the layers of network feature information produced by group convolution, thereby reducing the computational effort of the network.

In the YOLOv5s-SGC backbone network, to meet the requirements of detection speed and hardware device computing power, a lighter version of ShuffleNetV2 0.5× was used to build a lightweight and efficient feature extraction network to replace the CSP-DarkNet53 feature extraction network in YOLOv5s model. The ShuffleNetV2 0.5× has a total of 18 layers. The structure and parameter settings of the backbone are shown in Table 1. This

Output Repeat **Output Size** Kernel Size Stride Layer Channels Input 640×640 Conv1 320×320 3×3 2 1 24 2 1 24 Maxpooling 160×160 3×3 2 Shuffle_Block 80×80 1 48 1 3 48 Shuffle_Block 80×80 Shuffle_Block 40×80 2 1 96 Shuffle_Block 40×40 1 7 96 192 Shuffle_Block 20×20 2 1 Shuffle_Block 3 192 20×20 1

approach will lose a small amount of model accuracy, but can significantly reduce the

Table 1. The detailed structure of the backbone.

number of parameters and operations of the backbone network.

3.3. Generalized-FPN Structure

In a feature pyramid network, the purpose of multi-scale feature fusion is to aggregate features extracted from the backbone network at different resolutions. The structure of FPN + PAN of YOLOv5s only focuses on feature extraction and does not consider the connection of internal modules. Therefore, we use the modified Generalized-FPN as the feature fusion network for YOLOv5s-SGC, thus reducing the number of network operations and increasing the detection speed while enhancing the crew and fishing net detection accuracy. The Generalized-FPN uses $\log_2 l$ -link to extend the depth while retaining effective feature reuse and adopts a Queen-fusion structure to add more feature fusion, where each node receives inputs at both the same level and neighbor level, and finally, the feature fusion is performed by Concat.

3.4. The Ghost Module

After the feature extraction network, the feature maps for detecting the deck crew and fishing net need to undergo further processing in the feature fusion network. As shown in Figure 4, there will be some photos with high similarity among these feature maps, which is feature map redundancy in the neural network, leading to wasteful computation and parameter increase, from a single image following feature extraction by the neural network. In response to this situation, we introduced GhostNet [25] in the feature fusion section module in GhostNet. In the Ghost module, the first step is to generate a part of the feature map by a small amount of traditional convolution, after which a linear operation is performed on these feature maps, and the newly generated feature map becomes the Ghost feature map. Finally, the two parts of the feature map, and this structure is called Ghost convolution GSConv.

The C3 module and the CSP_GSC module are shown in Figure 5. The C3 module of YOLOv5s consists of a bottleneck and contains a large number of parameters, which makes it computationally intensive. In order to reduce the number of parameters of the neck network and improve the detection speed of the model. The study replaces the normal convolution and C3 modules in the YOLOv5s feature fusion network with the GSConv and CSP_GSC modules to reduce the size of the model. The CSP_GSC module is a replacement of the Conv module of BottleNeck in the C3 module with the GSConv module, so the parameters and computational effort in the CSP_GSC module are reduced. Note that since the Ghost module has two normal convolutions and two depth-wise convolutions in its bottleneck structure, if used extensively and without special optimization, the speed in hardware will be reduced, offsetting the advantages of reducing the number of parameters and FLOPs. As the computing power of the platform grows, the advantages of GSConv become less obvious.GSConv is better suited for edge computing devices due to its smaller

computing consumption and memory footprint. By changing the original operation of generating feature maps using convolutional kernels to retaining only a small number of convolutional kernels and replacing the rest of the convolutional operations with linear operations, the amount of computation and time required to generate feature maps can be significantly reduced [26].





Figure 4. Similar feature maps of YOLOv5s.



Figure 5. C3 module and CSP_GSC module.

3.5. The CBAM Module

Significantly reducing the number of parameters and computational effort of the backbone feature extraction network and the feature fusion network also reduces the detection accuracy of the model. As shown in Figure 6, the heat map of the dataset reveals that the accuracy of the model's target detection can be enhanced by incorporating an attention mechanism into the network, as the fishing nets are primarily located at the periphery or outside of the deck and the crew is primarily located on the deck.





(a) Heatmap of Fishing Nets

(b)Heatmap of Deck Crew

Figure 6. Dataset heatmap.

This study introduces the CBAM attention module [27] into the feature network to accurately obtain the relative position information of the target in the fishing vessel operation image, and to focus on the feature part of the fishing net and crew, thereby reducing the influence of information unrelated to the target, while also increasing the accuracy of target detection [27]. The CBAM module is an efficient and straightforward attention module for feed-forward convolutional neural networks. Given an intermediate feature map, the CBAM module sequentially infers the attention map along two separate dimensions, channel, and space, and finally multiplies the attention map with the input feature map to perform adaptive feature optimization, thereby enhancing the extraction of important features and suppressing target-irrelevant information.

In this paper, the CBAM attention module is added in front of the detection head, which improves the detection accuracy of the model, although it slightly increases the computational complexity and the number of parameters.

After the backbone feature extraction network is replaced with ShuffleNetV2, the feature fusion network is modified to a Generalized-FPN incorporating GSConv and CSP_GSC modules, and the CBAM attention module is added in front of the detection head, the structure of the lightweight real-time detection algorithm YOLOv5s-SGC based on the modified YOLOv5s network for fishing vessel personnel and retrieval nets is shown in Figure 7.



Figure 7. YOLOv5s-SGC model structure.

4. Experiments

4.1. Evaluation Metrics

In this paper, we use common evaluation metrics to judge target detection models: *Precision, Recall, mAP* (mean Average Precision), weight size, FLOPs, and detection speed.

The accuracy rate equals the number of samples correctly identified as positive cases by the network; the recall rate equals the ratio of the number of samples correctly identified as positive cases by the network to the total number of positive samples. *mAP* equals the area under the PR (Precision–Recall) curve, where P is Precision and R is Recall, as calculated as follows:

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

$$Recall = \frac{TP}{TP + FN}$$
(2)

TP represents "the number of positive samples identified by the network as positive", *FP* represents "the number of negative samples identified by the network as positive", and *FN* represents "the number of positive samples identified by the network as negative". *FN* represents "the number of positive samples identified as negative by the network". The *AP* of each category is obtained by integrating its PR curve. The *AP* is obtained by integrating the PR curves of each category, and the *mAP* is obtained by averaging the *AP* of each category:

$$mAP = \frac{1}{M} \sum_{m=1}^{M} AP^{(m)}$$
(3)

In this paper, we use FPS to evaluate the detection speed of the model. Moreover, the number of parameters and FLOPS are also used to evaluate the model. The larger the number of parameters, the higher the capability of the model on the hardware. the FLOPs are used to evaluate the computational power of the model, and the larger the FLOPs, the higher the computational power of the model is required.

4.2. Environment

The operating system used for model training is Windows 10, the CPU model is AMD_Ryzen_7_5800H, the GPU model is NVIDIA GeForce RTX 3070 Laptop, the video memory size is 8G, the memory size is 32G, the deep learning framework used is Pytorch 1.12.0, the programming language is Python 3.7, and the GPU acceleration libraries are CUDA 11.6 and CUDNN 8.3.2.

The operating system used during the running speed test was Windows 11. To simulate running the inspection program on a low computing power fishing boat hardware device, the CPU model was AMD_Ryzen_7_5800H, the memory size was 16 G, the deep learning framework used was Pytorch 1.12.1, and the programming language was Python 3.7.3.

4.3. Model Training

In the model training session, the number of training rounds was set to 150, the Adam optimizer was used, the batch size was set to 16, no pre-training weights were used, the initial learning rate was set to 0.001, the network input image size was 640×640 , using cosine annealing to reduce the learning rate. train_loss and val_losss changes during the training process are shown in Figure 8. The model was trained to 130 rounds when convergence was reached.



Figure 8. Change curves of loss value.

4.4. Results

Table 2 presents a comparison of FLOPs, model parameters, and weight sizes between the YOLOv5s-SGC and YOLOv5 models. Table 3 displays the detection results of the YOLOv5s-SGC and YOLOv5 models on the fishing vessel operation image dataset.

Table 2. Comparison of Yolov5s and Yolov5s SGC model.

Model	Params	FLOPs	Weight
YOLOv5s	7.02×10^{6}	15.8 G	13.7 MB
YOLOv5s-SGC	$0.76 imes 10^{6}$	2.0 G	1.83 MB

Model	Precision	Recall	AP	FPS
YOLOv5s	0.933	0.849	0.901	6.21
YOLOv5s-SGC	0.935	0.781	0.859	9.81

Table 3. Test results of YOLOv5s and YOLOv5s-SGC model.

It is evident that YOLOv5s-SGC outperforms YOLOv5s in terms of the number of parameters and floating point operations based on the data in Tables 2 and 3. The weight size of the method presented in this paper is only 1.83 MB, which is 86.64% less than YOLOv5s's 13.7 MB, and the FLOPs size is only 2.0GFLOPs. With similar accuracy, 6.8% less recall, and 4% less average precision than YOLOv5s, the algorithm presented in this paper obtains 9.81 FPS on CPU devices, a 57.97% improvement in detection speed. The comparative data demonstrates that the algorithm reduces the size and computation of the network model and increases the speed of computation while maintaining a certain level of accuracy, making it more suitable for deployment on hardware devices with limited computational capacity.

Figure 9 shows the confusion matrix for the YOLOv5s model and the YOLOv5s-SGC model. The horizontal and vertical coordinates indicate the FP and FN of the model, respectively, and the values on the diagonal line are the recall of the model's predicted targets. From the figure, it can be seen that YOLOv5s has 86% and 97% detection rates for crew and nets, and 14% of crew and 3% of nets are mistakenly detected as background. In contrast, the detection rates of YOLOv5s-SGC for crew and nets were 82% and 97%, with 18% of the crew and 3% of the nets being falsely detected as background. Overall, the difference in classification performance between YOLOv5s-SGC and YOLOv5s is not significant, but YOLOv5s-SGC has reduced the model size by 86.64% and improved the computing speed by 57.97%, which can be better applied to the real ship operation status detection tasks of deck crew and fishing nets usage.



(a)YOLOv5s

(b)YOLOv5s-SGC

Figure 9. Confusion matrix of YOLOv5s and YOLOv5s-SGC.

4.5. Performance Comparison of Different Models

This part aims to objectively and accurately evaluate the overall performance of the proposed model on the fishing vessel operation dataset, including detection accuracy, model size, FLOPs, and detection speed, and compare it to six other models.

As seen in Table 4, the YOLOv5s-SGC model has slightly lower accuracy than other models, but the number of parameters and FLOPS are the smallest among all models. This makes the detection speed of the YOLOv5s-SGC model second only to the YOLOv5n model among the eight models. In terms of detection speed, the YOLOv5s-SGC model is 61.89% of the YOLOv5n model, but in terms of model size and FLOPs it is only 43.18% and 48.78% of YOLOv5n, which is more suitable for deployment on low-computing shipboard devices. YOLOv5s-SGC achieves an FPS of 9.81 on CPU devices, which can satisfy the requirement for real-time target detection on fishing vessels.

Table 4. '	Test results	of different	models

Model	Params	FLOPs	Weight	FPS	mAP
YOLOv5s	$7.02 imes 10^6$	15.8 G	13.7 MB	6.21	0.901
YOLOv5s-MobileNetV3	$3.54 imes10^6$	6.3 G	7.08 MB	6.54	0.875
YOLOv5s-PP_LCNet	$3.71 imes 10^6$	8.1 G	7.36 MB	5.87	0.885
YOLOv5s-ShuffleNetV2 $1 \times$	$3.61 imes 10^6$	7.5 G	7.22 MB	7.46	0.870
YOLOv5s-GhostNet	$7.28 imes 10^6$	12.8 G	14.3 MB	3.61	0.897
YOLOv5n	$1.76 imes 10^6$	4.1 G	3.65 MB	15.85	0.885
YOLOv5lite	$1.54 imes10^6$	3.6 G	3.27 MB	9.55	0.88
YOLOv5s-SGC	$0.76 imes10^6$	2.0 G	1.83 MB	9.81	0.86

The detection results of each model on some of the test images are shown in Figure 10. These models were all able to identify the crew in the images and fishing net locations, but YOLOv5s-PP_LCNet appeared to misidentify them. The minimal number of parameters and computational effort of YOLOv5s-SGC also caused the model to have a low confidence level for some of the detected targets, but the confidence level was still above 0.5.



(f) YOLOv5n

(g) YOLOv5lite



Figure 10. Detection results of different models on test images.

From the above experimental results, it can be seen that, compared with other network models, the YOLOv5s-SGC model has a better trade-off between the number of parameters, the number of operations, and the accuracy of the model, and performs better in the detection of the deck crew and the use of fishing nets, which can detect the targets of fishing vessels in real-time and provide a basis for identifying the current operation status of fishing vessels.

4.6. Ablation Experiments

Ablation experiments were conducted on the models in this paper. The dataset and experimental environment were the same. Six models, YOLOv5s, YOLOv5s-ShuffleNetV2, YOLOv5s-GFPN, YOLOv5s-GFPN-S, YOLOv5s-GFPN-SG, and YOLOv5s-SGC were obtained by extending from different dimensions, respectively. The structures of different models are shown in Table 5 and the experimental results are shown in Table 6.

Table 5. Design of ablation experiments.

Model	ShuffleNetV2	Ghost Module	Generalized-FPN	СВАМ
YOLOv5s				
YOLOv5s-ShuffleNetV2	\checkmark			
YOLOv5s-GFPN			\checkmark	
YOLOv5s-GFPN-S	\checkmark		\checkmark	
YOLOv5s-GFPN-SG	\checkmark	\checkmark	\checkmark	
YOLOv5s-SGC	\checkmark	\checkmark	\checkmark	\checkmark

Table 6. Results of ablation experiments.

Model	Params	FLOPs	Precision	mAP	Weight	FPS
					0	
YOLOv5s	7.02×10^{6}	15.8 G	0.933	0.901	13.7 MB	6.21
YOLOv5s-ShuffleNetV2	$2.86 imes10^6$	5.6 G	0.919	0.862	5.79 MB	9.23
YOLOv5s-GFPN	$5.65 imes 10^6$	12.9 G	0.949	0.909	11.1 MB	7.39
YOLOv5s-GFPN-S	$1.33 imes10^6$	2.7 G	0.905	0.856	2.94 MB	12.24
YOLOv5s-GFPN-SG	$0.73 imes 10^6$	1.9 G	0.89	0.86	1.77 MB	10.32
YOLOv5s-SGC	$0.76 imes10^6$	2.0 G	0.935	0.859	1.83 MB	9.81

The ablation experiments results can be seen from Table 6. After replacing the backbone network with the ShuffleNetV2 backbone, the model weight was reduced by 7.91 MB and FLOPs by 10.2G and had a 48.63% increase in detection speed, but the mAP was reduced by 3.9%, respectively, compared to the original YOLOv5s. After replacing the feature fusion network with Generalized-FPN, not only did the model size and FLOPs decrease by 20 and 2.9 G, but the accuracy and mAP also improved by 1.6% and 0.8%, while the detection speed increased by 19%. This shows that the improved feature network can improve the model computing speed and accuracy while ensuring accuracy. YOLOv5s-GFPN-S was improved with both ShuffleNetV2 and Generalized-FPN, the weights and FLOPs of the model decreased significantly, resulting in a 97% increase in computing speed, but also a 2.8% and 4.5% decrease in accuracy and mAP. YOLOv5s-GFPN-SG improves on Generalized-FPN using the Ghost module. Compared with YOLOv5s-GFPN-S, the model size is reduced by 1.2MB while maintaining accuracy, but the runtime speed is reduced by 16.5%. Our proposed model YOLOv5s-SGC added the CBAM attention mechanism before the detection layer on the basis of YOLOv5s-GFPN-SG. Compared to YOlOv5s-SG, the weight of the model and FLOPs were only increased by 0.06MB and 0.1G, while precision was increased by 4.5%. Although YOLOv5s has better mAP, considering the parameter size and real-time performance of the model, the improved YOLOv5s-SGC is similar to YOLOv5s in mAP with only 10.83% params and 13.36% model weight, while the FPS results show the real-time performance is much better. In summary, the improved model YOLOv5-SGC in this paper is superior in terms of real-time detection for deck crew and fishing nets.

The above ablation experiments illustrate that replacing the backbone network with ShuffleNetV2 0.5× reduces the number of parameters and increases the running speed by introducing a depth-separable convolution with channel splitting. Replacing the feature fusion network with Generalized-FPN makes it possible to reduce the model size and increase the computational speed while maintaining accuracy. Adding the Ghost module to the feature fusion network and replacing the traditional convolution with cheap linear operations can also significantly reduce the number of parameters and computational complexity of the model and speed up the detection of the model. These methods will make the average accuracy value have a slight decrease, after the introduction of the CBAM

attention mechanism, through the space and channel attention module can compensate for the accuracy loss to a certain extent, and finally both make the model lightweight and ensure that the model has high accuracy. Thus, real-time and accurate detection can be performed on low-intelligence fishing vessel equipment to meet the needs of fishing vessel operation status detection.

5. Discussion

Researchers are increasingly using vision methods to detect the operational status of fishing vessels. Combining deep learning techniques with shipboard video surveillance is an important way to achieve the identification of the operational status of fishing vessels. In this study, we incorporate the features of YOLOv5s, ShuffleNetV2, Generalized-FPN, GhostNet, and CBAM attention modules to develop a lightweight target detection model to detect crew and fishing nets on deck and obtain their number and location information. This approach enables real-time target detection with high accuracy even on low computing power shipboard hardware devices.

We extracted pictures from the video of fishing boats out at sea and annotated the deck crew and fishing nets in the pictures to produce the dataset. Considering the fact that the images captured by the camera during the fishing boat's sea trip may be due to light changes and field of view occlusion, we processed the dataset by data augmentation, using various methods such as Gaussian noise and random rectangular occlusion in random combinations, thus increasing the diversity of the dataset and improving the robustness of the trained model.

Our deployed lightweight target detection model, YOLOv5s-SGC, achieves an FPS value of 9.81 on devices with low computing power and has a significant reduction in the number of parameters, FLOPs, and model size compared to YOLOv5s, with an accuracy of 86%. Based on the current performance of our model, it is expected that the recognition rate and detection speed in practical applications can meet the requirements of real-time detection and provide strong support for the operational status detection of fishing vessels. However, compared with the YOLOv5n model, although the model size and FLOPs are lower, there is still a gap in speed, so there is some room for improvement in our model description. We will further develop a more lightweight target detection model based on our model to better serve the automatic identification of fishing vessel operation status.

6. Conclusions

By installing two cameras on a trawler to record video data of fishing vessel operations at sea and creating an image dataset of fishing vessel operations from the video data, this paper proposes a lightweight real-time detection algorithm for deck crew and fishing nets based on improved YOLOv5s. This algorithm combines the features of YOLOv5s, ShuffleNetV2, Generalized-FPN, GhostNet, and CBAM attention modules YOLOv5s-SGC, which decreases the number of parameters. The model has been proven to be deployable on low computing power devices and to provide real-time accurate target detection of fishing vessels in operational images. It has also been proven to lessen the workload of law enforcement officers by determining the current operational status of fishing vessels based on the model's detection results.

Author Contributions: Conceptualization, J.W. and G.L.; methodology, J.W., X.Y. and G.L.; software, J.W.; validation, J.W.; formal analysis, J.W.; investigation, J.W.; resources, X.Y.; data curation, J.W. and X.Y.; writing—original draft, J.W.; writing—review & editing, J.W. and G.L.; supervision, G.L.; Project administration, G.L.; funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the project entitled "Research on multi parameter fusion recognition and tracking method of single fish target for large aquaculture water body" funded by the National Natural Science Foundation of China (NSFC), China (32073026), was supported by the project "Research and development of long-range stereo fish detection sonar equipment" financed from Sanya Yazhou Bay Science and Technology City Administration, China (SKJC-2020-01-013).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request from the authors.

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

References

- 1. Zou, L. Research on information technology for the regulation of fishing vessels. Farmers Consult. 2017, 24, 347–349.
- 2. Yusuf, M.; Razak, A. Illegal Fishing Eradication: Comparative Study of Indonesia and Malaysia. JL Pol'y Glob. 2018, 71, 120.
- 3. Zou, L. Study on The Supervision of Marine Fishing Vessel of Cangnan County. Master's Thesis, Northwest Agriculture and Forestry University, Xianyang, China, 2018.
- Feng, B.; Chen, X.; Zhu, G. Identification of complementary overfishing and its implications for fisheries management. *Resour. Dev. Mark.* 2010, 26, 20–23.
- Yan, W.; Jiang, Y.; Cai, L.; He, Y.; Wang, F. Analysis of Fishing Yield of Different Types of Fishing Boats on Operation by Grey Theory in Zhoushan. J. Zhejiang Ocean Univ. Nat. Sci. 2018, 37, 468–474.
- Lin, Z. Study on the need for the application of the Fishing Vessel Dynamic Monitoring Information System. *China Agric. Inform.* 2013, 10S, 152. Available online: https://xueshu.baidu.com/usercenter/paper/show?paperid=e8837f3d498802d227f0dcef3c785 c91&site=xueshu_se (accessed on 10 May 2023).
- Zhu, J. A Study on the Application of the Fishing Vessels Dynamic Monitoring Information System in Fishery Management. Master's Thesis, South China University of Technology, Guangzhou, China, 2010.
- Zhang, J.; Zhang, S.; Wang, S.; Yang, Y.; Dai, Y.; Xiong, Y. Recognition of Acetes chinensis fishing vessel based on 3-2D integration model behavior. *South China Fish. Sci.* 2022, 18, 126–135.
- 9. Wang, S.; Zhang, S.; Tang, F.; Shi, Y.; Sui, Y.; Fan, X.; Chen, J. Developing machine learning methods for automatic recognition of fishing vessel behaviour in the *Scomber japonicus* fisheries. *Front. Mar. Sci.* **2023**, *10*, 1085342. [CrossRef]
- 10. Wang, S.; Zhang, S.; Zhu, W.; Sun, Y.; Yang, Y.; Sui, J.; Shen, L.; Shen, J. Application of an electronic monitoring system for video target detection in tuna longline fishing based on YOLOV5 deep learning model. *J. Dalian Ocean Univ.* **2021**, *36*, 842–850.
- 11. Wang, S.; Sun, Y.; Zhang, S.; Sui, J.; Zhu, W.; Yang, S.; Fan, W. Automatic mapping of thermal diagram based on satellite AIS offshore ship position. *Fish. Inf. Strategy* **2021**, *36*, 45–53.
- 12. Pei, K.; Zhang, S.; Fan, W.; Hou, J.; Tang, X.; Zhu, W. Extraction method of stow net fishing intensity distribution in Zhejiang Province. *J. Fish. China* **2020**, *44*, 1913–1925.
- 13. Zhang, S.; Fan, W.; Zhang, H.; Yang, S.; Shen, J.; Zou, G. Text extraction from electronic monitoring video of ocean fishing vessels. *Fish. Inf. Strategy* **2020**, *35*, 141–146.
- 14. Tang, X.; Zhang, S.; Fan, W.; Pei, K. Fishing type identification of gill net and trawl net based on deep learning. *Mar. Fish.* **2020**, 42, 233–244.
- 15. Wang, S.; Zhang, S.; Dai, Y.; Wang, Y.; Sui, J.; Zhu, W. Research on calculating fishing depth of krill by sonar data. *South China Fish. Sci.* **2021**, *17*, 91–97.
- 16. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
- Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
- Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
- 19. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
- 20. Cui, C.; Gao, T.; Wei, S.; Du, Y.; Guo, R.; Dong, S.; Lu, B.; Zhou, Y.; Lv, X.; Liu, Q. PP-LCNet: A lightweight CPU convolutional neural network. *arXiv* 2021, arXiv:2109.15099.
- Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
- Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
- 23. Jiang, Y.; Tan, Z.; Wang, J.; Sun, X.; Lin, M.; Li, H. GiraffeDet: A heavy-neck paradigm for object detection. *arXiv* 2022, arXiv:2202.04256.
- Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856.

- 25. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
- 26. Du, K.L. Clustering: A neural network approach. Neural Netw. 2010, 23, 89–107. [CrossRef] [PubMed]
- 27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.