electronics

MDPI

Article

An Efficient Ship-Detection Algorithm Based on the Improved YOLOv5

Jia Wang [1], Qiaoruo Pan [1], Daohua Lu [1,2,*] and Yushuang Zhang [1]

[1] School of Mechanical Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China
[2] Marine Equipment and Technology Institute, Jiangsu University of Science and Technology, Zhenjiang 212003, China
* Correspondence: ludaohua_just@126.com

Abstract: Aiming to solve the problems of large-scale changes, the dense occlusion of ship targets, and a low detection accuracy caused by challenges in the localization and identification of small targets, this paper proposes a ship target-detection algorithm based on the improved YOLOv5s model. First, in the neck part, a weighted bidirectional feature pyramid network is used from top to bottom and from bottom to top to solve the problem of a large target scale variation. Second, the CNeB2 module is designed to enhance the correlation of coded spatial space, reduce interference from redundant information, and enhance the model's ability to distinguish dense targets. Finally, the Separated and Enhancement Attention Module attention mechanism is introduced to enhance the proposed model's ability to identify and locate small targets. The proposed model is verified by extensive experiments on the sea trial dataset. The experimental results show that compared to the YOLOv5 algorithm, the accuracy, recall rate, and mean average precision of the proposed algorithm are increased by 1.3%, 1.2%, and 2%, respectively; meanwhile, the average precision value of the proposed algorithm for the dense occlusion category is increased by 4.5%. In addition, the average precision value of the proposed algorithm for the small target category is increased by 5% compared to the original YOLOv5 algorithm. Moreover, the detection speed of the proposed algorithm is 66.23 f/s, which can meet the requirements for detection speed and ensure high detection accuracy and, thus, realize high-speed and high-precision ship detection.

Keywords: object detection; YOLOv5; small target; crowded detection; attention mechanism

Citation: Wang, J.; Pan, Q.; Lu, D.; Zhang, Y. An Efficient Ship-Detection Algorithm Based on the Improved YOLOv5. *Electronics* **2023**, *12*, 3600. https://doi.org/10.3390/electronics12173600

Academic Editor: Donghyeon Cho

Received: 27 July 2023
Revised: 22 August 2023
Accepted: 24 August 2023
Published: 25 August 2023

1. Introduction

With the development of artificial intelligence and other emerging technologies, the research on intelligent and unmanned boats has been highly valued. As a developing country, China has also experienced an inevitable trend of using intelligent and unmanned boats in the development of national strategies. Intelligent boats denote a new type of boat, with their capabilities of autonomous environmental perception, autonomous planning, and autonomous navigation having a broad application scope in both military and civilian fields. Intelligent boats can complete military tasks in different types of dangerous waters and have an important role in various marine-related tasks, including surface cleaning, maritime transport, maritime patrol, and maritime monitoring. However, with the increase in the demand for intelligent vessels to complete target-detection tasks in different water types, vision-detection methods have been constantly evolving.

Traditional vision-detection methods are mostly based on machine learning, where features are extracted by image algorithms, such as LBP (Local Binary Patterns) [1], and then fed to classifiers, such as SVM (Support Vector Machine) [2–4] and KNN (K–Nearest Neighbor) [5], to classify candidate targets. Traditional visual-detection methods have many drawbacks; for instance, machine-learning-based visual-detection methods rely greatly on the design of manual features and are susceptible to interference from the external environment. However, since the sea surface is a complex and volatile environment,

traditional target-detection methods are not robust enough and have an effective real-time performance in this environment. Therefore, to perform the maritime target-detection task efficiently and accurately, a more efficient, accurate, and robust method is necessary to complete image feature extraction and target recognition. Currently, deep-learning-based target-detection algorithms have been a hot spot and a challenging research area in the field of vision detection. Compared to the traditional machine-learning-based vision-detection methods, deep-learning-based target-detection methods can learn features from a large amount of data and achieve high accuracy in both prediction and classification. By continuously learning features and adjusting a model's parameters to handle nonlinear problems, self-adaptability can be achieved. Deep-learning-based target-detection algorithms mainly include one-stage algorithms, such as the YOLO (You Only Look Once) [6–8] series of algorithms, and two-stage detection algorithms, such as the Faster R-CNN (Region-CNN) [9]. It should be noted that the YOLO series of algorithms have been widely used due to their simple network structure and high real-time performance.

In recent years, target-detection algorithms have been widely studied in the fields of pedestrian detection [10–14], object tracking [15–17], face detection [18,19], stereo images [20–22], car detection [23–25], defect detection [26,27], semantic detection [28,29], and hyperspectral-anomaly detection [30,31]. However, there are certain limitations in their practical application, particularly due to the problems of poor small-target-detection performance and target occlusion. Namely, in small-target detection, small low-resolution targets with little visualization information make it challenging to extract features with discriminative power; the detection performance is highly susceptible to interference from environmental factors, which can easily cause items in the background to be identified as targets, resulting in false detection. To address the problems of low resolution and little visual information on small targets, Lim et al. [32] proposed a context-based target-detection method; this method can solve the problem of little visual information on small targets; but, its detection accuracy and speed are still low. Deng et al. [33] proposed an extended feature pyramid network (EFPN) with additional high-resolution pyramid layers; the EFPN has been designed particularly for small-target detection; but, the feature coupling still affects the detection performance of this method for small objects.

Occlusion includes the mutual occlusion between both the targets to be detected and the targets to be detected being occluded by interferers [34]. The first type of occlusion makes the target location difficult to determine because the features between targets are similar; meanwhile, the second type of occlusion causes a large loss of target information due to the occlusion caused by interfering objects and can easily lead to missed detection. Under the condition of self-obscuration between targets [35], the locating capability of a model and the sensitivity to the NMS threshold [36,37] can be enhanced to improve detection accuracy. Moreover, when targets are obscured by interferers, the missed-detection rate can be reduced by improving the ability of a model to learn target features.

Aiming to address the problems of the large-scale changes, density, and occlusion of ship targets, and a low detection accuracy caused by difficulties in locating and identifying small targets, the existing mainstream target-detection algorithms have been tested on sea test data. The results have shown that the YOLOv5 algorithm's mAP (mean average precision) and detection FPS are higher than those of the one-stage detection algorithms, such as the YOLOv4 and YOLOv7 algorithms, and two-stage detection algorithms, such as the Faster R-CNN algorithm. In view of that, the YOLOv5 algorithm [38–40] was used as a basic algorithm for ship target detection.

This paper introduces a ship target-detection algorithm based on the improved YOLOv5 algorithm. In the neck part, the BiFPN (Bidirectional Feature Pyramid Network) [41] is used as a weighted bidirectional feature pyramid network from top to bottom and from bottom to top. During the feature-fusion process, the importance of different features is learned by performing channel-stitching operations on the shallow and deep information; different weights are applied to features according to their importance to solving the problem of large changes in the target scale. The CNeB2 module is designed to enhance

the correlation of the coded airspace so that all pixels in a $k \times k$ square area centered on the center point of the spatial position in each spatial position are interrelated, thus allowing the extraction of independent information on different targets, reducing the interference of redundant information and strengthening the recognition and differentiation abilities to block dense targets. Further, the SEAM [42] (Separated and Enhancement Attention Module) attention mechanism is added before an image is input to the detector to focus on important information needed in the current task, which reduces the attention to other miscellaneous information and enhances the ability to identify and locate small targets.

2. Materials and Methods

2.1. YOLOv5 Model

The YOLOv5 algorithm is a one-stage target-detection algorithm that can perform three tasks: target detection, semantic segmentation, and image classification. The structures of the YOLOv5 network models are the same; the only difference relates to the depth and width of the network model. The structure of the YOLOv5 network model is shown in Figure 1, where it can be seen that this model consists of three main parts: backbone, neck, and head. The backbone performs the main feature extraction of an image using three modules, namely, the CBS module, consisting of the Conv module, a subsumed layer, and an activation function; the C3 module; and the SPPF module. In the neck part, multiple Upsample modules, C3 modules, CBS modules, and channel-stitching operations are used to fuse the feature maps of different receptor fields and enhance the multiscale feature extraction capability of the model. The head part consists of three detectors. In addition, adaptive anchor frames are used to perform target detection on feature maps of different scales.

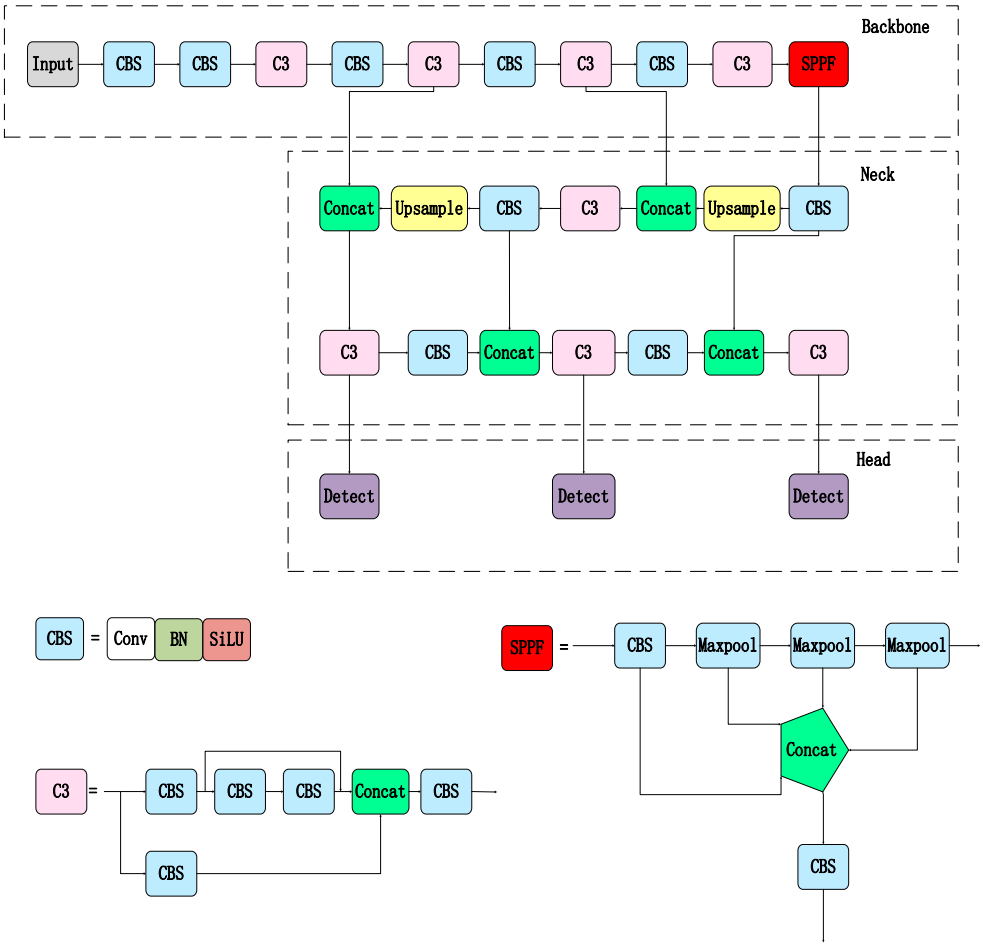


Figure 1. Illustration of the YOLOv5 network structure.

2.2. BiFPN Model

When processing the features extracted by the backbone network, the YOLOv5 algorithm adopts a separate top-down and bottom-up PANet (Path–Aggregation Network) for feature-fusion processing; meanwhile, the weighted bidirectional feature pyramid network (the BiFPN, see Figure 2) regards each top-down and bottom-up path as a feature network layer. Repeated operations can be carried out in the feature layer. The features extracted on different levels are fully integrated. In addition, the BiFPN learns the importance of different features through feature fusion and applies different weights to features according to their importance. In this way, all extracted features can be fully integrated while distinguishing different features, and the problem of the large-scale changes of ship targets, can be effectively solved.

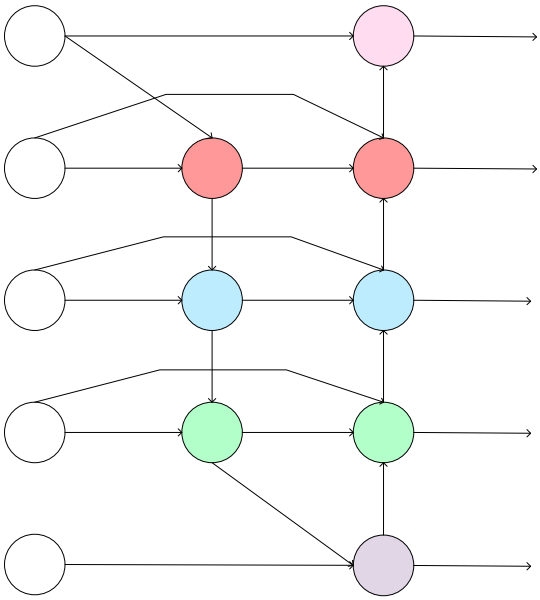


Figure 2. The simplified structure of the BiFPN.

2.3. CNeB2 Module

The CNeB2 module introduces the idea of convolutional modulation [43], based on the C3 module, and its structure is presented in Figure 3. For an input of $X \in \mathrm{RH} \times W \times \mathrm{C}$, where H represents the height of an input image, W is the width of the input image, and C is the number of channels in the image input, a simple depth convolution with a kernel size of $k \times k$ and the Hadamard product are used to calculate the output of Z as follows:

$$Z = A \odot V \tag{1}$$

$$A = DWConv_{k \times k}(W_1 X) \tag{2}$$

$$V = W_2 X \tag{3}$$

where W_1 and W_2 are the weight matrices of the two linear layers and $DWConv_{k \times k}$ denotes a depth convolution with a kernel size of $k \times k$.

The convolutional modulation operation allows all pixels within a $k \times k$ square region centered on the centroid of the spatial location in each spatial location to be mutually correlated. Information interaction between channels is achieved through linear layers and the output data of the depth convolution are used as weights to modulate the features after linear projection, where the output of each spatial location represents a weighted sum of all pixels within a square region. The CNeB2 module can enhance the coded space domain

correlation, reduce interference from redundant information, and extract independent information on different targets, solving the problem of mutual occlusion between targets.

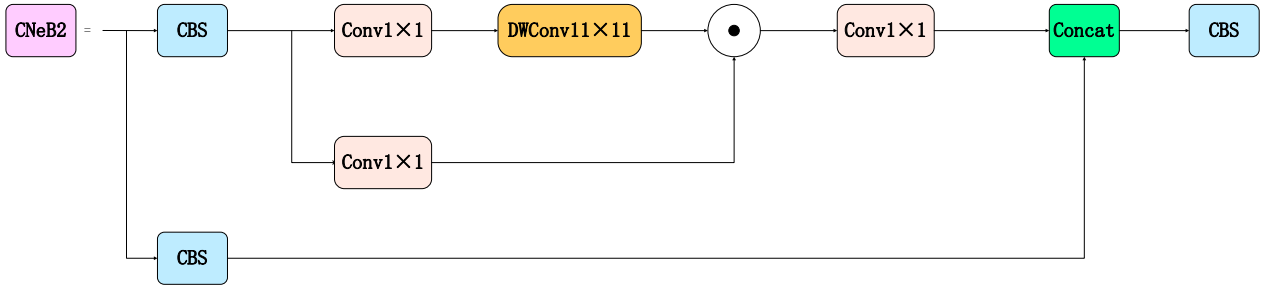


Figure 3. The block diagram of the CNeB2 module.

In the CNeB2 module, the data first pass through the CBS module in the first branch and then through two convolutions, one having a kernel size of 1×1 and the other having a deep convolution with a kernel size of 11×11. The data are then multiplied in the convolution block with a kernel size of 1×1, which converts the input to the output. Finally, this result of the channel-stitching operations is combined with the CBS module's output in the second branch through the convolution with a kernel size of 1×1.

2.4. SEAM

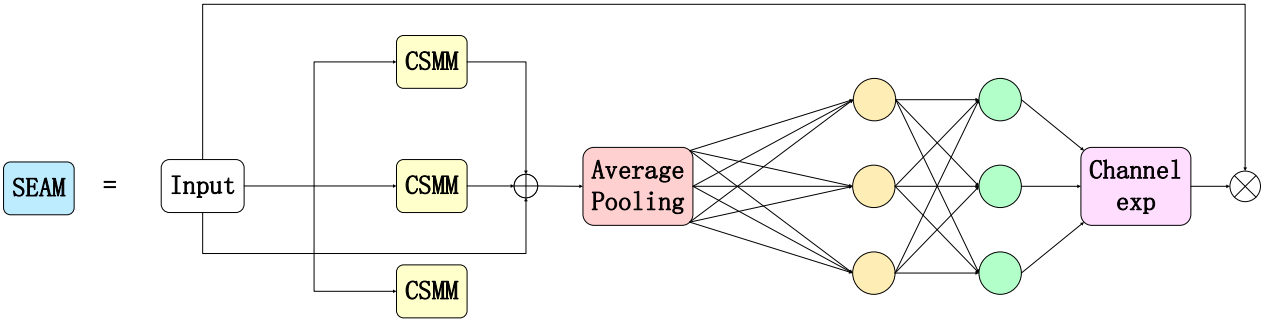
The SEAM is used to focus on important information needed for the task at hand and reduce the attention to redundant information, even filtering out distracting information to solve the problem of information overload and improve the accuracy and efficiency of task processing. As there are many small targets in the dataset used in this study, which are difficult to identify, the SEAM attention mechanism is used to improve the algorithm's ability to locate and identify small targets to enhance the algorithm's attention to small target features.

The structure of the SEAM attention mechanism is presented in Figure 4, where it can be seen that the first part of the SEAM is a depth-separable convolution with residual connections. The depth-separable convolution operates on the depth-by-depth level, where the convolution is separated on a channel-by-channel basis. Although the depth-separable convolution can learn the importance of different channels and reduce the number of parameters, it ignores the informational relationships between channels. Therefore, the output data of different depth convolutions are combined using the point-by-point (1×1) convolution and the channel information is fused using two fully connected layers to strengthen the connections between the channels. Then, the output of the fully connected layer is processed using an exponential function to enhance the locating capability of the module for small targets; finally, the output of the SEAM's output is multiplied by the original features.

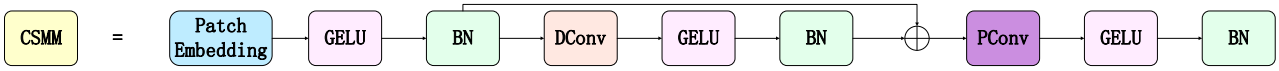
2.5. Proposed Improved Algorithm

The overall framework of the proposed improved YOLOv5 algorithm for maritime ship target detection is shown in Figure 5. As shown in Figure 5, the backbone part extracts the main image features using the CBS, C3, and SPPF modules. The BiFPN is employed in the neck part to combine the shallow semantic information extracted by the backbone part with the deep semantic information extracted by the neck part to achieve multiscale feature fusion. In the improved model, the C3 module in the first three layers of the neck part of the original YOLOv5 network is changed to a CNeB2 module to enhance the coded space relevance, reduce interference from redundant information, and enhance the recognition of dense targets for distinction in occlusion. The backbone part is mainly used to extract main feature information and, then, multiscale feature fusion is performed in the neck part. However, the up-down fusion of multiple scales may cause a certain redundancy in feature information, thus affecting the accuracy of the prediction results. The main function of the

CNeB2 module is to enhance the coding spatial correlation and reduce interference from redundant information, which can improve this problem. Therefore, this study replaces the C3 module in the neck part of the original YOLOv5 model with the CNeB2 module. Further, before an image is input into the detector, the SEAM attention mechanism is added. This mechanism focuses on important information needed for the current task among the massive input data, reduces the algorithm's attention to miscellaneous information, and even filters out interference information, thus improving both the positioning and recognition abilities of the proposed algorithm for small targets.



(**a**) SEAM structure



(**b**) CSMM structure

Figure 4. The block diagram of the SEAM attention mechanism.

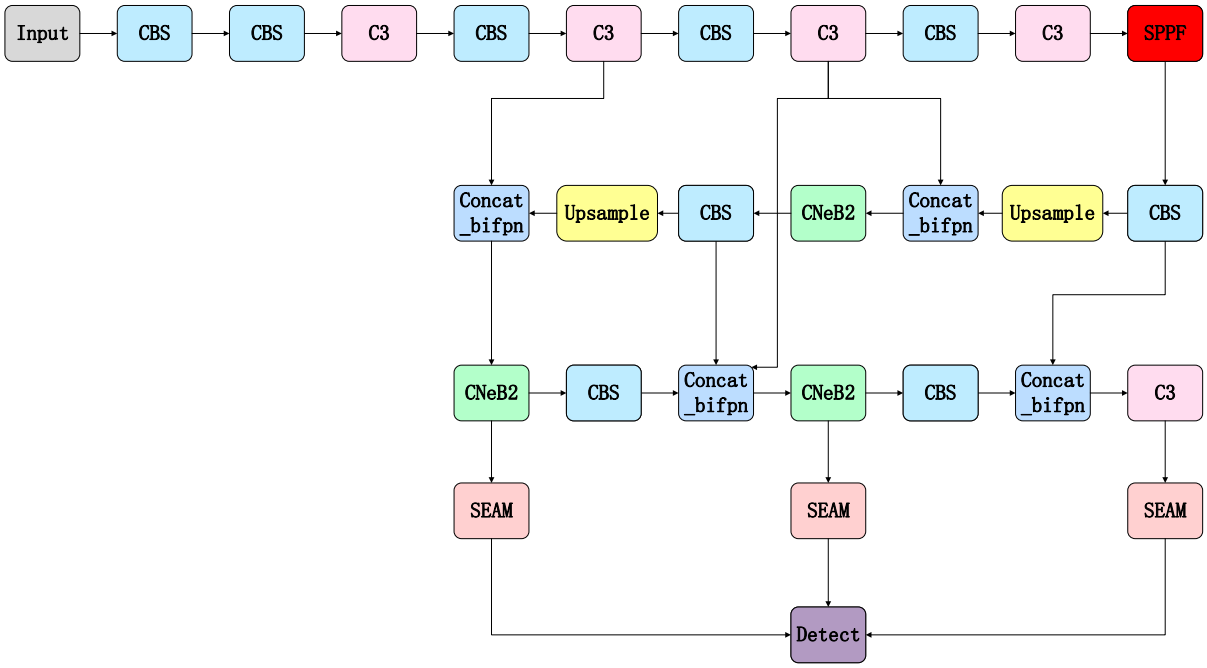


Figure 5. The block diagram of the proposed model.

3. Experiments

The proposed algorithm was tested on a sea trial dataset. The process of constructing the sea trial dataset used in the experiments, the experimental platform, the parameter settings, and the experimental methodology is described in the following section.

3.1. Experimental Dataset

The experimental dataset was obtained from the video data on marine vessels collected during the sea trial of the "Shipborne Unmanned Underwater Vehicle Retract System," which is a key, special project of "Deep–sea Key Technologies and Equipment" under the national key research and development plan led by Jiangsu University of Science and Technology. The dataset included 740 images obtained through frame extraction and deweighting; the VOC format annotation was performed on the acquired images through LabelImg. The marked images were enhanced with offline data to improve the robustness of the proposed algorithm. By performing a random combination of operations of flip transformation, brightness adjustment, Gaussian blur, and Gaussian noise addition, 2960 experimental images were obtained after offline data enhancement of the existing dataset. The data were divided into training and validation sets according to the ratio of 8:2. Further, the sea trial dataset was divided into five categories: unmanned surface vessel; kayak boat; the experimental mother ship; autonomous underwater vehicle; and other ships, which were labeled as USV, kayak boat, muchuan, AUV, and ship, respectively. The categories and their numbers of labels are shown in Figure 6.

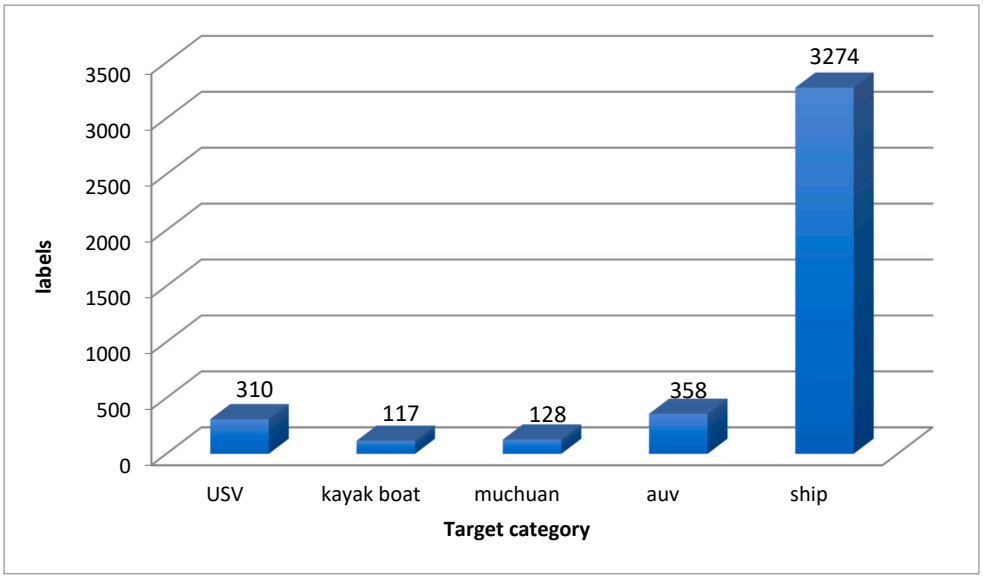


Figure 6. The number of target labels by category in the experimental dataset.

3.2. Experimental Platform and Parameter Settings

The operating system of the experimental platform was Windows 11; the deep learning framework was Pytorch1.8.1; the CUDA version was 11.1; the cuDNN version was 8.0.4; the GPU was RTX 3060 Ti; and the memory size was 32 GB. The 640 × 640 RGB images were used as an input; the batch size was set to four; and the maximum number of training epochs was set to 300. The initial learning rate was 0.01; the function-optimization method was SGD; the warmup epochs were 3.0; and the box loss gain and class loss gain were 0.05 and 0.5, respectively. The IoU training threshold was set to 0.2 and the anchor-multiple threshold was 4.0. Mosaic was not used for data enhancement.

3.3. Experimental Design

3.3.1. Evaluation Indexes

To verify the accuracy of the proposed model, the *mAP* and detection speed values obtained by precision and recall were used as evaluation metrics in this study. In this study, the *AP* (average precision) indicated the detection accuracy of a single category. The P–R curve denoted a curve drawn based on the precision and recall values, using them as the vertical and horizontal axes, respectively. The *AP* indicated the area under the P–R curve of a certain category and the *mAP* value represented the average of the area under the P–R

curve of all categories; the larger the *mAP* was, the higher the algorithm's accuracy was. The *mAP* was calculated by:

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{4}$$

where n denotes the total number of categories and i is the category index.

Further, the precision and recall values were, respectively, calculated by:

$$P = \frac{TP}{(TP + FP)} \tag{5}$$

$$R = \frac{TP}{(TP + FN)} R = \frac{TP}{(TP + FN)} \tag{6}$$

where *TP*, *TN*, *FP*, and *FN* are defined as follows:

TP (True Positive): the original sample represented a positive class, as well as the prediction result;

TN (True Negative): the original sample was a negative class, as well as the prediction result;

FP (False Positive): the original sample denoted a negative class but the prediction result was a positive class;

FN (False Negative): the original sample was a positive class but the prediction result was a negative class.

The detection speed *FPS* was calculated as follows:

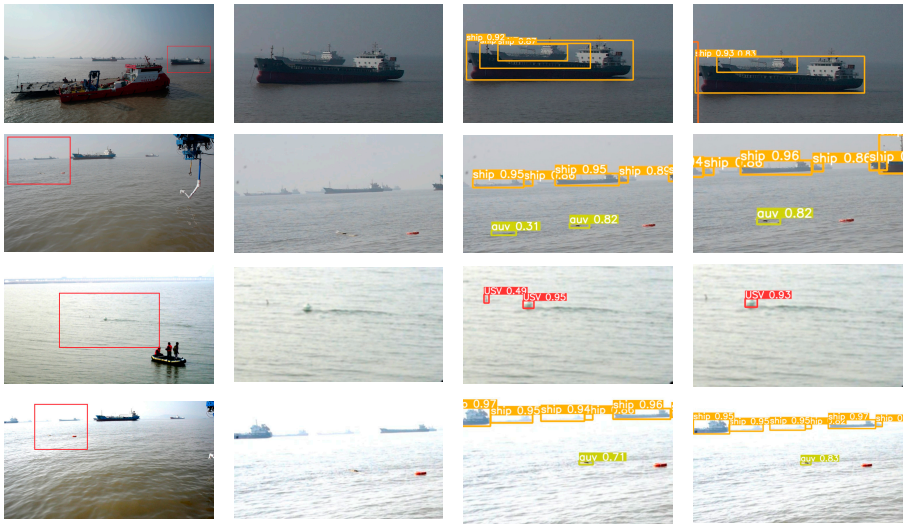
$$FPS = \frac{1000}{t} \tag{7}$$

where t is the inference time, expressed in ms.

3.3.2. Experimental Results

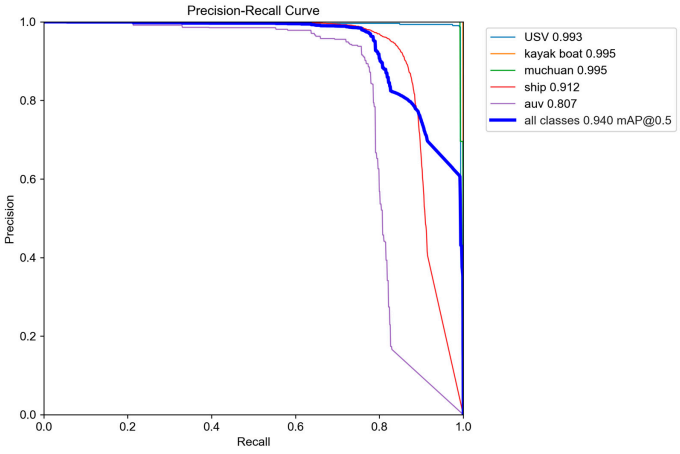
The trained YOLOv5 algorithm and the proposed improved algorithm were tested on the same dataset, in turn, and the test results are displayed in Figure 7. Since the ship class coincidence rate in the figure exceeded 50%, it was defined as an obscured target. As shown in the first image in Figure 7a, the ship category was a dense target. The ratio between the width and height of the boundary box for the AUV class in the figure and the width and height of the image was less than 0.1; it was defined as a small target. As shown in Figure 7, compared to the detection result of the original YOLOv5 algorithm, the detection box of the proposed algorithm fitted the target to be detected more closely and could detect dense and occluding targets better. Meanwhile, the detection accuracy of the proposed algorithm for small targets was higher than that of the original YOLOv5 algorithm, which indicated that the proposed algorithm reduced the false detection and missed detection rates of the original YOLOv5 algorithm and improved the detection of ships at sea.

The P–R curves of the original YOLOv5 algorithm and the proposed algorithm trained on the same sea trial dataset are shown in Figure 8; Figure 8a displays the P–R curve of the YOLOv5 algorithm and Figure 8b presents the P–R curve of the proposed algorithm. In Figure 8, it can be seen that the AP of the dense occlusion category (ship) was improved by 4.5%, the AP of the small target category (AUV) was improved by 5%, and the APs of the other categories were unchanged.

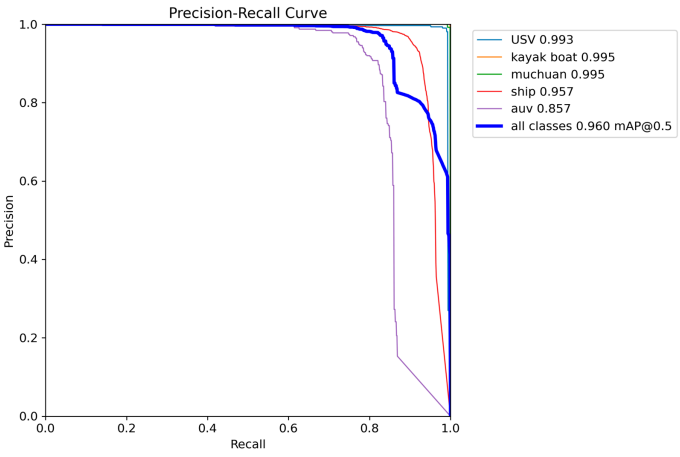


(**a**) (**b**) (**c**) (**d**)

Figure 7. Visualization of the detection results; (**a**) the original image; (**b**) the locally enlarged image; the corresponding local enlarging is marked by red boxes in the original image in (**a**); (**c**) the detection result of the original YOLOv5 algorithm; (**d**) the detection result of the proposed algorithm.



(**a**)



(**b**)

Figure 8. Comparison results of the P–R curves of the original YOLOv5 algorithm and the proposed algorithm; (**a**) displays the P–R curve of the YOLOv5 algorithm and (**b**) presents the P–R curve of the proposed algorithm.

The loss curve can reflect the dynamic trend of network training; so, the loss curve was selected in this study to measure the difference between the predicted results and the corresponding real values. The loss curve of the proposed algorithm is shown in Figure 9, where it can be seen that both training loss and validation loss converged and the two curves basically coincided, thus achieving a good fit.

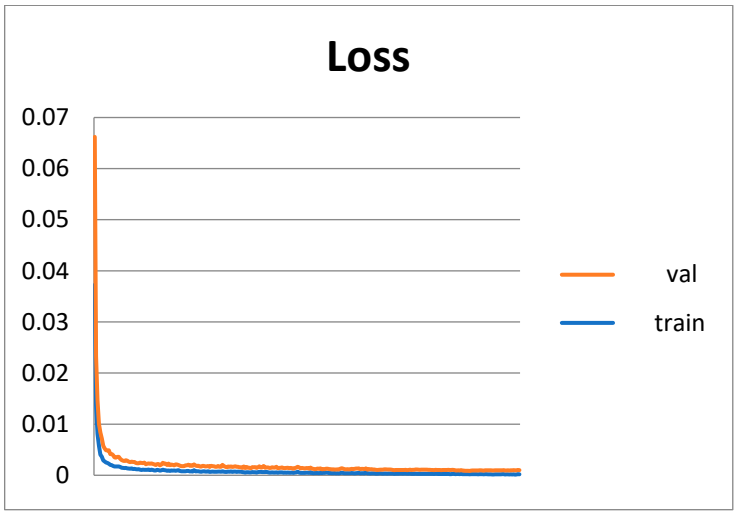


Figure 9. The loss curve of the improved proposed algorithm.

4. Discussion

4.1. Ablation Experiments

The effectiveness of the modules in the improved algorithm was verified through ablation experiments. The results of the ablation experiments are shown in Table 1.

Table 1. The results of the ablation experiments.

Algorithm	mAP (%)	FPS (Frame/s)
YOLOv5s	94.0	67.56
YOLOv5s + BiFPN	94.8	65.36
YOLOv5s + SEAM	94.3	66.67
YOLOv5s + CNeB2	94.4	72.46
YOLOv5s + BiFPN + SEAM	95	64.12
YOLOv5s + BiFPN + CNeB2	95.5	71.42
YOLOv5s + CNeB2 + SEAM	94.6	70.92
YOLOv5s + BiFPN + CNeB2 + SEAM	96	66.23

In Table 1, it can be seen that adding the BiFPN, SEAM attention mechanism, and CNeB2 module to the original YOLOv5 algorithm, separately, improved the mAP value by 0.8%, 0.4%, and 0.3%, respectively. However, adding the BiFPN and the SEAM attention mechanism reduced the detection speed by 2.20 f/s and 0.89 f/s, respectively. In contrast, the addition of the CNeB2 module to the original YOLOv5 algorithm increased the detection speed by 4.90 f/s. Further, the addition of two lots of the BiFPN, CNeB2 module, and SEAM attention mechanism to the original YOLOv5 algorithm, separately, improved the mAP by 1.0%, 1.5%, and 0.6%, respectively. However, adding two lots of the BiFPN reduced the speed by 3.44 f/s; meanwhile, adding two lots of the CNeB2 module and SEAM attention mechanism increased the speed by 3.86 f/s and 3.36 f/s, respectively. Finally, adding all of the modules to the original YOLOv5 model at the same time increased the mAP to 96%, which improved the overall mAP by 2%; the speed decreased by 1.33 f/s but still had a high real-time performance.

4.2. Comparative Experiments

On the sea trial dataset, the proposed Yolov5 model was compared with five mainstream algorithms, including the Faster R-CNN, SSD, YOLOv4, YOLOv5, and YOLOv7 algorithms, using the evaluation metrics presented in Section 3.3.1. The comparison results are shown in Table 2. To conduct a fair comparison, the training parameters of the YOLOv4 and YOLOv7 algorithms, such as input shape, epoch, and batch size, were the same.

Table 2. Comparison of the proposed algorithm and several mainstream algorithms.

Algorithm	USV (%)	Kayak Boat (%)	Muchuan (%)	Ship (%)	AUV (%)	mAP (%)	FPS (Frame/s)
Faster R-CNN	70.4	94.2	97.8	43.1	59.2	72.9	19.84
SSD	90.8	95.6	92.4	81.2	74.1	86.8	30.64
YOLOv4	88.2	99.8	100.0	85.4	79.6	90.6	28.11
YOLOv5	99.3	99.5	99.5	91.2	80.7	94.0	67.56
YOLOv7	95.8	99.5	97.5	91.2	79.4	92.7	61.35
Ours	99.3	99.5	99.5	95.7	85.7	96.0	66.23

As shown in Table 2, compared to the two-stage Faster R-CNN algorithm, the proposed algorithm improved the mAP and detection speed by 23.1% and 46.39 f/s, respectively; compared to the one-stage algorithms, SSD, YOLOv4, YOLOv5, and YOLOv7, the mAP of the proposed algorithm was improved by 9.2%, 5.4%, 2%, and 3.3%, respectively. Further, compared to the one-stage algorithms, SSD, YOLOv4, and YOLOv7, the detection speed of the proposed algorithm was improved by 35.59 f/s, 38.12 f/s, and 4.88 f/s, respectively. However, compared to the YOLOv5 algorithm, the detection speed of the proposed algorithm decreased by 1.33 f/s; but, it could still ensure high real-time performance. The APs of the proposed algorithm in all categories were the highest, except for the categories muchuan and kayak boat, for which the performances of the proposed algorithm were 0.5% and 0.3% below the maximum value, respectively.

The closer the value of precision was to one, the more targets the algorithm could detect. As shown in Table 3, compared to the two-stage Faster R-CNN algorithm, the proposed algorithm improved the total precision by 43.7%; for all types of targets, the precision values of the proposed algorithm were higher than those of the two-stage Faster R-CNN algorithm. Further, the precision values of the proposed algorithm for the kayak boat, muchuan, and AUV targets were 1%, 0.2%, and 0.5% lower than the maximum values of the one-stage algorithms, the SSD, YOLOv4, and YOLOv5 algorithms; but, the total precision values were improved by 0.9%, 4.3%, and 1.6%, respectively. Meanwhile, for the USV and ship, the proposed algorithm had the highest precision values among all algorithms.

Table 3. Comparison results of the precision of different algorithms.

Algorithm	USV (%)	Kayak Boat (%)	Muchuan (%)	Ship (%)	AUV (%)	All (%)
Faster R-CNN	41.4	51.9	59.2	59.7	60.2	54.5
SSD	92.9	100.0	100.0	96.1	97.3	97.3
YOLOv4	95.4	100.0	100.0	88.5	85.7	93.9
YOLOv5	98.8	99.1	100.0	91.3	94.0	96.6
YOLOv7	91.2	95.4	95.3	67.7	87.5	90.9
Ours	99.0	99.0	99.8	96.4	96.8	98.2

Finally, the closer the value of recall was to one, the higher the accuracy of the target detection of the algorithm was. As shown in Table 4, compared with the one-stage algorithms, the SSD and YOLOv7, the total recalls of the proposed algorithm increased by 31.6% and 14.9%, respectively. The results indicated that the recall values of the proposed algorithm for the USV and muchuan classes were 0.1% and 0.8% lower than those of the one-stage algorithm YOLOv5 and the two-stage algorithm Faster R-CNN; but, the total recall values were improved by 1.2% and 17.8%, respectively. Further, the recall values

of the proposed algorithm for the kayak boat, ship, and AUV were the highest among all algorithms.

Table 4. Comparison results of the recall of different algorithms.

Algorithm	USV (%)	Kayak Boat (%)	Muchuan (%)	Ship (%)	AUV (%)	All (%)
Faster R-CNN	74.0	97.6	100.0	45.3	58.4	75.0
SSD	62.8	92.6	64.0	50.9	35.6	61.2
YOLOv4	84.6	95.2	97.6	81.8	71.3	86.1
YOLOv5	99.0	100.0	99.2	85.6	74.1	91.6
YOLOv7	80.6	98.3	95.3	75.4	47.5	77.9
Ours	98.9	100.0	99.2	89.9	76.0	92.8

5. Conclusions

This paper proposes an improved YOLOv5-based ship target-detection algorithm to address the problems in the existing maritime ship target-detection methods. The proposed algorithm uses top-down and bottom-up weighted bidirectional feature pyramid networks for multiscale feature fusion in the neck part. In addition, the CNeB2 module and SEAM attention mechanism are used to optimize the model structure to solve the problems of obscured dense targets, which are difficult to distinguish and identify, and small targets that are difficult to locate. The performance of the proposed algorithm is verified by a large number of experiments on a sea trial dataset. The experimental results show that compared to the original YOLOv5 algorithm, the proposed algorithm can improve the accuracy, recall, and mAP values by 1.3%, 1.2%, and 2%, respectively; the AP of the occlusion intensive category (ship) is improved by 4.5%, and the AP of the small target category (AUV) is improved by 5%. The proposed algorithm can achieve better detection accuracy while meeting the detection speed requirements, thus effectively realizing high-speed and high-precision ship detection.

However, there is still room for improvement in the detection accuracy of the proposed algorithm for AUV small targets. In view of that, future research will aim to optimize the network structure of the proposed algorithm and enhance its feature-learning ability and recognition and locating capabilities for small targets, which will further improve its overall detection accuracy.

Author Contributions: Conceptualization, J.W. and Q.P.; methodology, J.W. and Q.P.; software, Q.P.; validation, Q.P.; formal analysis, J.W. and D.L.; investigation, Q.P. and Y.Z.; data curation, Q.P. and Y.Z.; writing—original draft preparation, Q.P.; writing—review and editing, J.W. and D.L.; visualization, Q.P.; supervision, J.W. and D.L.; project administration, J.W. and D.L.; funding acquisition, J.W. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Jiangsu Province Key Research and Development Program (No.BE2022062).

Data Availability Statement: Unfortunately, the data are not available due to project confidentiality.

Acknowledgments: The authors are thankful for the funding from Jiangsu Province Key Research and Development Program (No.BE2022062).

Conflicts of Interest: The authors declare that they have no conflict of interest to report regarding the present study.

References

1. Laurent, C.; Sébastien, M. On the detection of morphing attacks generated by GANs. In Proceedings of the 2022 International Conference of the Biometrics Special Interest Group, Darmstadt, Germany, 14–16 September 2022.
2. Gómez, J.K.C.; Puentes, Y.A.N.; Niño, D.D.C.; Acevedo, C.M.D. Detection of Pesticides in Water through an Electronic Tongue and Data Processing Methods. *Water* **2023**, *15*, 624. [CrossRef]
3. Xie, Z.; Du, S.; Lv, J.; Deng, Y.; Jia, S. A Hybrid Prognostics Deep Learning Model for Remaining Useful Life Prediction. *Electronics* **2021**, *10*, 39. [CrossRef]

4. Liu, S.; You, S.; Yin, H.; Lin, Z.; Liu, Y.; Cui, Y.; Yao, W.; Sundaresh, L. Data source authentication for wide–area synchrophasor measurements based on spatial signature extraction and quadratic kernel SVM. *Int. J. Electr. Power Energy Syst.* **2022**, *140*, 108083. [CrossRef]
5. Jason, H.; Lyons, D.M. Wall Detection Via IMU Data Classification In Autonomous Quadcopters. In Proceedings of the 7th International Conference on Control, Automation and Robotics, Singapore, 23–26 April 2021.
6. Yan, K.; Li, Q.; Li, H.; Wang, H.; Fang, Y.; Xing, L.; Yang, Y.; Bai, H.; Zhou, C. Deep learning–based substation remote construction management and AI automatic violation detection system. *IET Gener. Transm. Distrib.* **2022**, *9*, 16. [CrossRef]
7. Jesse, R.; Polina, D.; Josh, S.; Bryant, T.C.; Ross, M.; Tim, R.; Andrew, M.K. 7 Characterization of Feeder Cattle Behavior Using an Integrated Machine Vision Learning System. *J. Anim. Sci.* **2022**, *100*, 23–24.
8. Oguine, K.J.; Oguine, O.C.; Bisallah, H.I. YOLO v3: Visual and Real–Time Object Detection Model for Smart Surveillance Systems (3s). In Proceedings of the 5th Information Technology for Education and Development, Abuja, Nigeria, 6–8 September 2022.
9. Kabra, K.; Xiong, A.; Li, W.; Luo, M.; Lu, W.; Garcia, R.; Vijay, D.; Yu, J.; Tang, M.; Yu, T.; et al. Deep object detection for waterbird monitoring using aerial imagery. In Proceedings of the 21st IEEE International Conference on Machine Learning and Applications, Nassau, Bahamas, 12–15 December 2022.
10. Li, L.; Guo, X.; Wang, Y.; Ma, J.; Jiao, L.; Fang, L.; Xu, L. Region NMS–based deep network for gigapixel level pedestrian detection with two–step cropping. *Neurocomputing* **2022**, *468*, 482–491. [CrossRef]
11. Maji, D.; Nagori, S.; Mathew, M.; Poddar, D. YOLO–Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, New Orleans, LA, USA, 19–20 June 2022.
12. Gilroy, S.; Jones, E.; Glavin, M. Overcoming Occlusion in the Automotive Environment—A Review. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 23–35. [CrossRef]
13. Chen, N.; Li, M.; Yuan, H.; Su, X.; Li, Y. Survey of pedestrian detection with occlusion. *Complex Intell. Syst.* **2021**, *7*, 577–587.
14. Zhang, J.; Liu, C.; Wang, B.; Chen, C.; He, J.; Zhou, Y.; Li, J. An infrared pedestrian detection method based on segmentation and domain adaptation learning. *Comput. Electr. Eng.* **2022**, *99*, 107781. [CrossRef]
15. Alotaibi, M.F.; Omri, M.; Khalek, S.A.; Khalil, E.; Mansour, R.F. Computational Intelligence–Based Harmony Search Algorithm for Real–Time Object Detection and Tracking in Video Surveillance Systems. *Mathematics* **2022**, *10*, 733. [CrossRef]
16. Leira, F.S.; Helgesen, H.H.; Johansen, T.A.; Fossen, T.I. Object detection, recognition, and tracking from UAVs using a thermal camera. *J. Field Robot.* **2021**, *38*, 242–267. [CrossRef]
17. Ong, J.; Vo, B.; Kim, D.Y.; Nordholm, S. A Bayesian Filter for Multi–View 3D Multi–Object Tracking with Occlusion Handling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2246–2263. [CrossRef] [PubMed]
18. Zeng, D.; Veldhuis, R.; Spreeuwers, L. A survey of face recognition techniques under occlusion. *IET Biom.* **2021**, *10*, 581–606. [CrossRef]
19. Yang, T.; Wu, J.; Liu, L.; Chang, X.; Feng, G. VTD–Net: Depth Face Forgery Oriented Video Tampering Detection based on Convolutional Neural Network. In Proceedings of the 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020.
20. Shi, Y.; Guo, Y.; Mi, Z.; Li, X. Stereo CenterNet–based 3D object detection for autonomous driving. *Neurocomputing* **2022**, *471*, 219–229. [CrossRef]
21. Zhou, Z.; Du, L.; Ye, X.; Zou, Z.; Tan, X.; Zhang, L.; Xue, X.; Feng, J. SGM3D: Stereo Guided Monocular 3D Object Detection. *IEEE Robot. Autom. Lett.* **2022**, *7*, 10478–10485. [CrossRef]
22. Jiang, H.; Lu, Y.; Chen, S. Research on 3D Point Cloud Object Detection Algorithm for Autonomous Driving. *Math. Probl. Eng.* **2022**, *2022*, 8151805. [CrossRef]
23. Pillai, U.K.; Valles, D. An Initial Deep CNN Design Approach for Identification of Vehicle Color and Type for Amber and Silver Alerts. In Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 27–30 January 2021.
24. Lian, J.; Wang, D.; Zhu, S.; Wu, Y.; Li, C. Transformer–Based Attention Network for Vehicle Re–Identification. *Electronics* **2022**, *11*, 1016. [CrossRef]
25. Hu, R.; Ma, W.; Lin, W.; Chen, X.; Zhong, Z.; Zeng, C. Technology Topic Identification and Trend Prediction of New Energy Vehicle Using LDA Modeling. *Complexity* **2022**, *2022*, 9373911. [CrossRef]
26. Chen, X.; Lv, J.; Fang, Y.; Du, S. Online detection of surface defects based on improved YOLOV3. *Sensors* **2022**, *22*, 817. [CrossRef]
27. Dmitriev, S.F.; Malikov, V.; Ishkov, A.; Voinash, S.; Kalimullin, M.; Marat, S.; Liliya, N. Ultra–Compact Eddy Current Transducer for Corrosion Defect Search in Steel Pipes. *Mater. Sci. Forum* **2022**, *1049*, 282–288. [CrossRef]
28. Ma, W.; Gong, C.; Xu, S.; Zhang, X. Multi–scale spatial context–based semantic edge detection. *Inf. Fusion* **2020**, *64*, 238–251. [CrossRef]
29. Li, J.; Wang, P.; Ni, C.; Rong, W. Loop Closure Detection Based on Image Semantic Segmentation in Indoor Environment. *Math. Probl. Eng.* **2022**, *2022*, 7765479. [CrossRef]
30. Sheng, L.; Min, Z.; Xi, C.; Liang, W.; Xu, M.; Wang, H. Hyperspectral Anomaly Detection via Dual Dictionaries Construction Guided by Two–Stage Complementary Decision. *Remote Sens.* **2022**, *14*, 1784.
31. Lin, S.; Zhang, M.; Cheng, X.; Zhou, K.; Zhao, S. Hyperspectral Anomaly Detection via Sparse Representation and Collaborative Representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 946–961. [CrossRef]

32. Lim, J.; Astrid, M.; Yoon, H.; Lee, S. Small Object Detection using Context and Attention. In Proceedings of the 2021 International Conference on Artificial Intelligence in Information and Communication, Jeju Island, Republic of Korea, 13–16 April 2021.
33. Deng, C.; Wang, M.; Liu, L.; Liu, Y. Extended Feature Pyramid Network for Small Object Detection. *IEEE Trans. Multimed.* **2020**, *24*, 1968–1979. [CrossRef]
34. Li, W.; Wei, Y.; Lyu, S.; Chang, M.C. Simultaneous multi–person tracking and activity recognition based on cohesive cluster search. *Comput. Vis. Image Underst.* **2022**, *214*, 214. [CrossRef]
35. Wang, Z.; Xie, Q.; Wei, M.; Long, K.; Wang, J. Multi–feature Fusion VoteNet for 3D Object Detection. *ACM Trans. Multimed. Comput. Commun. Appl.* **2022**, *18*, 1–17. [CrossRef]
36. Zhora, G. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
37. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise–IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
38. Feng, H.; Jie, S.; Hang, M.; Wang, R.; Fang, F.; Zhang, G. A novel framework on intelligent detection for module defects of PV plant combining the visible and infrared images. *Sol. Energy* **2022**, *236*, 406–416.
39. Zhou, Q.; Liu, H.; Qiu, Y.; Zheng, W. Object Detection for Construction Waste Based on an Improved YOLOv5 Model. *Sustainability* **2023**, *15*, 681. [CrossRef]
40. Jubayer, F.; Soeb, J.; Mojumder, A.; Paul, M.; Barua, P.; Kayshar, S.; Akter, S.S.; Rahman, M.; Islam, A. Detection of mold on the food surface using YOLOv5. *Curr. Res. Food Sci.* **2021**, *4*, 724–728. [CrossRef] [PubMed]
41. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.
42. Yu, Z.; Huang, H.; Chen, W.; Su, Y.; Liu, Y.; Wang, X. YOLO–FaceV2: A Scale and Occlusion Aware Face Detector. *arXiv* **2022**, arXiv:2208.02019.
43. Hou, Q.; Lu, C.; Cheng, M.; Feng, J. Conv2Former: A Simple Transformer–Style ConvNet for Visual Recognition. *arXiv* **2022**, arXiv:2211.11943.