electronics

Article

Multi-Site and Multi-Scale Unbalanced Ship Detection Based on CenterNet

Feihu Zhang * and Xujia Hou

School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China; hxj1363947894@mail.nwpu.edu.cn
* Correspondence: feihu.zhang@nwpu.edu.cn

Abstract: Object detection plays an essential role in the computer vision domain, especially the machine learning-based approach, which has developed rapidly in the past decades. However, the development of convolutional neural networks in the marine field is relatively slow, such as in ship classification and tracking. In this paper, ship detection is considered as a central point classification and regression task but discards the non-maximum suppression operation. We first improved the deep layer aggregation network to enhance the feature extraction capability of tiny targets, then reduced the number of parameters through the lightweight convolution module, and finally employed a unique activation function to enhance the nonlinearity of the model. By doing this, the improved network not only suits unbalanced sample ratios in classifying, but is more robust in scenarios where both the number and resolution of samples are unstable. Experimental results demonstrate that the proposed approach obtains outstanding performance and especially suits tiny object detection compared with current advanced methods. Furthermore, in contrast to the original CenterNet framework, the mAP of the proposed approach increased by 5.6%.

Keywords: deep learning; object detection; anchor-free; neural network; artificial intelligence

check for updates

Citation: Zhang, F.; Hou, X. Multi-Site and Multi-Scale Unbalanced Ship Detection Based on CenterNet. *Electronics* **2022**, *11*, 1713. https://doi.org/10.3390/electronics11111713

Academic Editor: George A. Papakostas

Received: 8 April 2022
Accepted: 21 May 2022
Published: 27 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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/).

1. Introduction

Recently, vision-based object detection has attracted significant interest from both industrial and research domains, due to the progress of deep learning techniques [1–4]. The purpose of object detection requires the computer to automatically identify the object from the original images concerning its category, position, and confidence. The developments of object detection have also achieved significant milestones in various domains such as autonomous driving [5,6], robot vision [7], video surveillance [8,9], and medical imaging [10]. However, in contrast to the scenarios mentioned above, deep learning has still been underappreciated in ocean exploitation [11].

Transportation departments can significantly facilitate the workforce supervising ships because they can dispatch and provide early warning according to artificial intelligence technologies. In addition, in the case of electromagnetic signal interference, ship detection can also assist in completing the perception of the surrounding environment. Hence, more and more scholars and research institutions have paid attention to artificial intelligence for ship detection [12–14]. Current mainstream methods could be mainly divided into traditional machine learning for feature modeling and deep learning for end-to-end solving. As the former method is mainly applicable to limited scenarios, more and more research focuses have turned to the latter. However, deep learning-based methods have the following issues: First and foremost, the efficiency of deep learning depends heavily on the quantity and quality of the training dataset [15], which is a fatal issue for sparse scenes. The final performance will be biased if the ship dataset is not comprehensive enough. Second, the size of the ship varies largely. For example, the length and height between a fishing boat and an aircraft carrier could lead to the solution to tolerate significant changes in network

characteristics. Compared with the existing target detection algorithm, the proposed anchor-free method considers the object detection task as a regression task for the central point and gets rid of the limitation of NMS [16].

Although various technologies have been developed, no methodology is perfect in practice. Especially in marine environments, types of ships have been collected with unbalanced ratios for further training, depending on the availability of online resources. Meanwhile, both the number and resolution of samples are also unstable, which increases the burden of training and seriously affects the effect of detection. Our previous work introduced an anchor-free method, however, only with simple experiments and without both qualitative and quantitative analysis.

This paper proposes an improved CenterNet approach for ship detection, where a public dataset is also configured by collecting public and legal photos. After standardizing the dataset and counting its characteristics distribution, the proposed network structure and convolution modules have configured fewer parameters. As the improved network does not depend on anchor setting, the robustness is thus guaranteed. Meanwhile, by comparing the effects of different activation functions, the flexibility is also exhibited, as the proposed approach could be modularly applied in other fields with minor modifications. Compared to our previous work [17], the proposed approach has been evaluated with rigorous analysis in the experiment, which illustrates that the anchor-free method has powerful competitiveness in evaluation indicators. The main contributions of this paper are as follows:

- A large ship detection dataset with 9800 images was open to the public, which contains different kinds of ships with multi-site and multi-scale characteristics;
- Different from traditional anchor-based methods, ship detection is considered an object center point classification and regression task. Although the NMS operation is avoided, its performance is still guaranteed;
- Both the advantages and disadvantages are analyzed to guide the specific scenarios in the future.

This paper is organized as follows: Section 2 introduces the related work, concerning traditional feature-based methods and deep learning-based methods for ship detection. Section 3 presents the improved CenterNet framework. Section 4 analyzes the performance of the proposed approach in contrast to the widely utilized methods, and Section 5 concludes the paper and discusses possible directions for future work.

2. Related Work

Before neural networks were widely used, the feature extraction method was essential for many scientific and industrial interests. However, the synthesis of feature extraction remains a major challenge, and research has consistently shown that it lacks precision in most scenarios. Currently, new methods based on deep learning have replaced the former. Therefore, the work in ship detection is mainly accomplished by adjusting and improving the methods mentioned above.

2.1. Traditional Ship Detection Algorithm

Many literary studies have been reported involving feature-based ship detection. Matsumoto [18] used Histograms of Oriented Gradient(HOG) and Support Vector Machines(SVM) for ship classification. Since the feature extraction process uses the sliding window to select possible targets, accuracy and speed are limited in practice. Shi [19] proposed a ship detection method based on a visual attention model, where a wavelet transforms and extracts the image's low-frequency and high-frequency features. However, this method is prone to omission or error detection in high cluttered scenarios. Furthermore, by considering that the camera is fixed at the dock, Zhu [20] detected moving objects by using the difference between adjacent video frames and then extracted and classified their features. However, although it avoids unnecessary calculations, such scenarios are limited regarding stationary ships. Jin [21] utilized high-resolution images and Harris corner detec-

tor to detect the sharp part of the ship, which detects not only the ship but also the ship's direction. However, relying solely on the above methods is far from enough. In summary, using traditional methods, ship detection technologies are feasible in mathematical derivation but not satisfactory in efficiency and accuracy. Moreover, the application scenarios of these methods are limited, mainly affected by light and noise conditions. Moreover, the robustness of hand-designed features is relatively low and cannot be applied to multiple scenes.

2.2. Deep Learning-Based Ship Detection Algorithm

Concerning deep learning-based ship detection approaches, Wang [22] proposed an improved YOLOv3 [23] end-to-end ship detection system by introducing the CFE [24] module and modifying the loss function. As a result, its detection accuracy has been significantly improved compared with the traditional algorithms. Chen [25] addressed the poor training dataset problem , where an improved Generative Adversarial Networks(GAN) was proposed to generate new samples, and then YOLOv2 [26] was utilized for detection. Zhao [27] divided the detection task into two parts—detection, and recognition—which have been successfully deployed in embedded devices. Zou [12] used the MobilenetV2 [28] network to extract ship features. The network is pre-trained on the coco dataset and then fine-tuned. Later, faster and higher precision detection results were obtained by using Faster R-CNN [29] as the comparison group. Aiming at the problems of missing detection in small-scale ship images, Yu [13] modified the feature network of YOLOv3 by adding aspect ratio into the loss function and finally obtained a higher detection result. Liu [14] combined YOLOv5 and GhostNet [30] to refine image features given the uneven distribution of horizontal and vertical components of ships to achieve good results.

However, questions have been raised about the manual configuration of the prolonged use of deep learning methods inevitably relying on anchors. Previous studies have suffered from several conceptual, comprehensive issues and methodological weaknesses regarding hyperparameters. Some evidence suggests that learning-based strategy would significantly reduce the performance in unbalanced sample environments. Furthermore, the anchor-based method always employs NMS to remove repeated test results, which occupies a considerable part of the computing resources.

3. Ship Detection with Improved CenterNet

3.1. CenterNet

As aforementioned, anchors are utilized in the strategies for manual configuration. However, only a few could benefit from the training process. In other words, most background anchors lead to uneven training and thus drop the accuracy performance. Furthermore, the unbalanced distribution of anchors often leads to the model conduct statistics issues. Hence, anchor-free models are designed, called CenterNet, as shown in Figure 1.

CenterNet considers the detection task as a critical point estimation problem. First, the original images are scaled to a fixed size and sent to the backbone, which typically includes three types: DLA-34 [31], Hourglass-104 [32], and Resnet-18 [33]. Then, the feature map is obtained with convolution operation and is restored to the original size by upsampling afterward. Finally, three branches are jointly connected at the back-end of the network layer to predict categories, sizes, and center bias. The computational cost has been significantly dropped because the NMS operation has been abandoned.

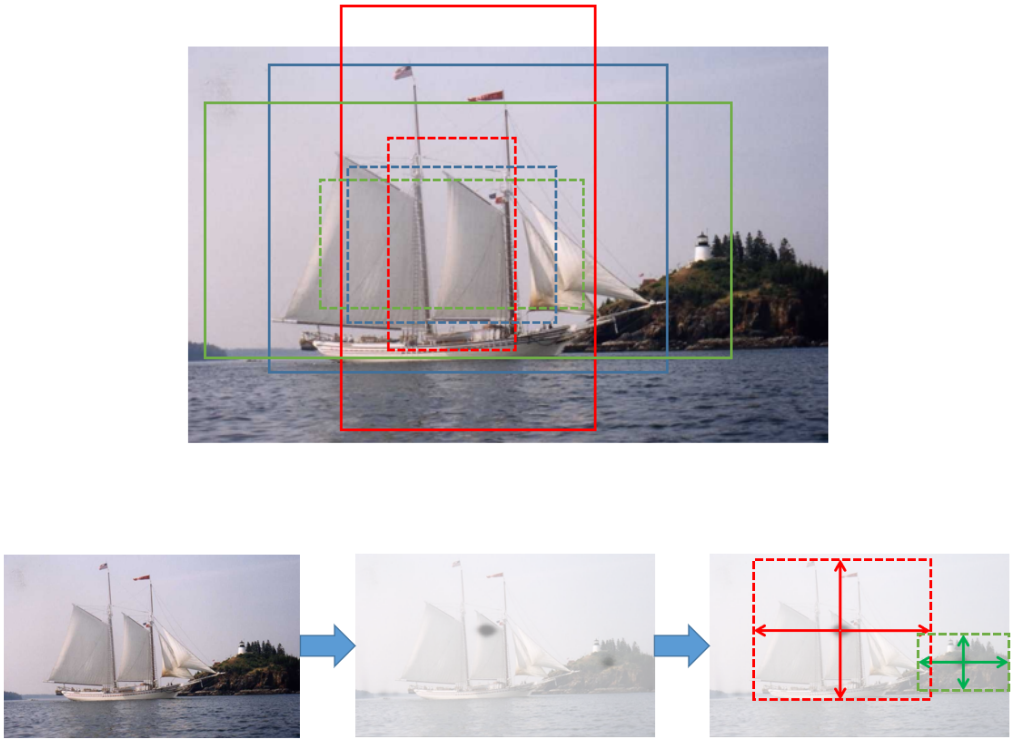


Figure 1. The method based on anchor generates a large number of anchors. Candidate boxes with IOU more than 0.7 are considered as positive samples, while the rest are considered as negative samples. After feature extraction, CenterNet uses 2D Gaussian distribution to generate a heat map, whose peak value is considered as the object, and then regress to object attributes according to the peak point

3.2. Proposed CenterNet

In contrast to earlier detection tasks, learning the statistical features and data distribution is necessary for CenterNet. Figure 2 exhibits the location and size distribution of the training dataset, which implies that large proportions are constructed of unbalanced multi-scale and multi-size samples. Most samples are within 0.1 of the image size and distributed in the center of the dataset. In addition, due to the structural characteristics of the ship itself, it can be seen that the statistical value in the horizontal direction is more significant than that in the vertical direction.

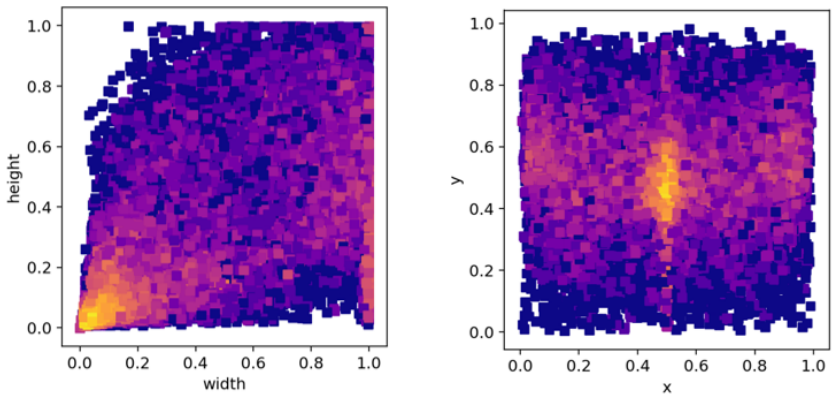


Figure 2. The sample distribution.

Considering the dataset's characteristics and practical issues such as convenient deployment, the lightest backbone, DLA, is adopted. Hence, a backbone extraction network is constructed based on DLA, and the detailed process of the proposed method is illustrated in Figure 3. The input image will first go through DLA for feature extraction. In DLA, downsampling will be carried out continuously. Therefore, 3×3 deformable convolution is connected to upsampling after the last bit of DLA. Finally, three branches output the heatmap, the center offset, and the target size. Heatmap is used to predict target categories, and center offset is used to correct position errors.

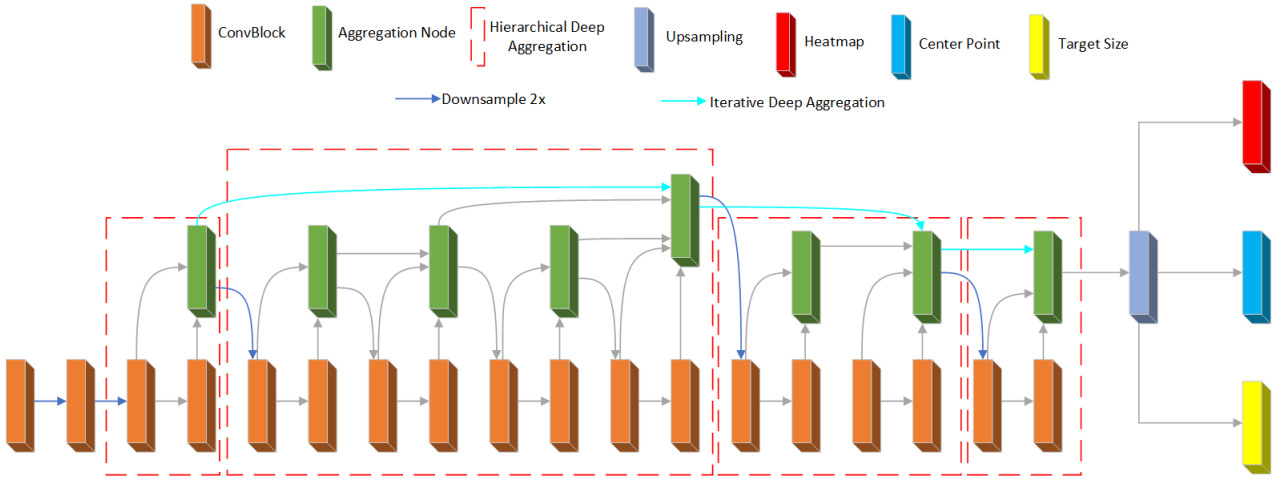


Figure 3. The flow chart of the proposed method

In the proposed DLA model, the receptive field and non-linearity representation are required to enhance feature extraction performance. Meanwhile, different blocks and modules merge spatial and semantic data by aggregating feature layers. Hence, both Hierarchical Deep Aggregation(HDA) and Iterative Deep Aggregation(IDA) are developed for existing and future networks. Here, HDA is used to fuse features from different modules, whereas IDA is connected across scales and resolutions. Figure 3 also exhibits that the proposed DLA could take advantage of both feature pyramid and dense connection. According to the primary structure, the IDA model is utilized to aggregate scales and extract resolution. Then, representations from various groups are aggregated by HDA through tree-like structures. Finally, after four down-sampling steps, the IDA fuses different levels, whereas HDA is stacked on the upper layer for feature extraction.

In this paper, due to the imbalance of data distribution, we focus on strengthening the second layer feature extraction of HDA. After extracting shallow features such as texture in the first layer, the combination of HDA and IDA in the second layer is used to fuse different size features. Then, the last two HDA layers carry out deep feature extraction and fusion. Hence, the multi-scale and multi-size ship samples could be obtained with standard formats. Meanwhile, to decline the number of network parameters and enhance detection speed, the convolution layer varies from two 3×3 convolutions to $1 \times 1 + 3 \times 3 + 1 \times 1$ convolution, where the first convolution layer can effectively reduce dimension. Finally, the last convolution layer is employed to fuse the information of other convolution layers. Furthermore, the proposed model also utilizes a non-monotonic neural activation function, as follows (Ref. [34]):

$$f(x) = x * \tanh(\log(1 + e^x)) \tag{1}$$

The activation function plays an essential role in the trial, introducing nonlinear factors to neurons and making the network flexible. Various activation functions are shown in Figure 4. Sigmoid is widely employed, however, with the gradient disappearance problem. Tanh has a faster convergence rate but with a similar issue. ReLU could alleviate gradient disappearance, as the derivative is constant on the positive half axis. Moreover, leaky ReLu

also has a minimal gradient on the negative half axis compared to the original ReLU. Mish is a self-regularized non-monotone activation function whose curve is smooth everywhere, allowing helpful information to penetrate the neural network. However, it is not as fast as the other activation functions due to the complexity of the calculation.

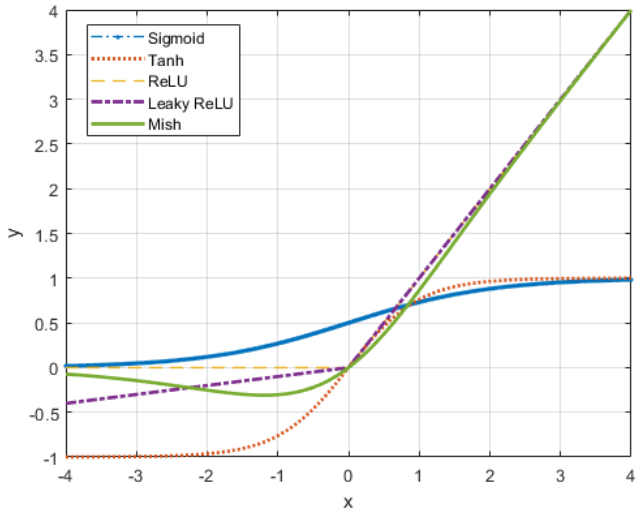


Figure 4. Mish - ReLU comparison.

4. Experiment

4.1. Environment Configuration

The experimental environment in this paper is Ubuntu 18.04LTS, CUDA 10.0, Pytorch 1.3.1, cuDNN 7.6.5, and NVIDIA RTX 2080 Ti GPU.

4.2. Experimental Data

Due to unbalanced samples' scales and sizes issues, obtaining plenty of training data with annotations is quite challenging. Hence, we combined the targets in a significant dataset with specific tuning. The implemented dataset originated from 9800 tagged images that include 6 categories: Linear, Container Ship, Bulk Carrier, Sailboat, Island Reef, and Other Ship, while the total numbers of each type are exhibited in Figure 5. The dataset can be downloaded from https://drive.google.com/file/d/1XwXfFYmrdTzutelblsDRCWXLhQ0MyfP-/view?usp=sharing (accessed on 10 December 2021)

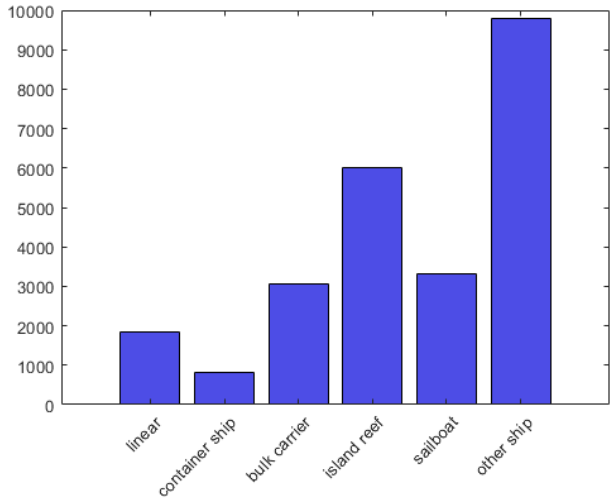


Figure 5. Total number of each categories in the experiment.

Each frame is configured with an associated XML file containing the filename, image size, type of object, and location during the experiment. Furthermore, the test set, verification set, and training set are also divided with the ratio 1:1:8. Figure 6 presents the original images of the training set.

Figure 6. The dataset images.

4.3. Loss Function

In training, corresponding to the three outputs of the network, there are three loss functions: (1) heatmap loss, (2) the center point offset loss, and (3) the target's size loss. Regarding heatmap loss, the loss function is configured with Focal Loss [35] function as follows:

$$L_k = -\frac{1}{N}\sum_{xyc}\begin{cases} (1-Y_p)^{\alpha}\log(Y_p), & \text{if } Y=1 \\ (1-Y)^{\beta}(Y_p)^{\alpha} & \\ \log(1-Y_p), & \text{otherwise} \end{cases} \tag{2}$$

where Y illustrates the actual position of the object point, and Y_p describes the predicted position of the object point. α and β are the hyper-parameters of focal loss, where N is the total number of critical points. Once the weight of Y_p is equal to 1, it could be efficiently detected. Otherwise, the proposed network could not fully calculate the center point. Hence, the corresponding parameter should be raised. Otherwise, the parameters from nearby points around the actual center adjusted $(Y_p)^{\alpha}$ to 0. Since the final feature map only takes 1/4 of the information in contrast to the original one, the center point offset loss functions are introduced to train the basis:

$$L_{off} = \frac{1}{N}\sum_{p}\left|O_p - \left(\frac{P}{R} - P_p\right)\right| \tag{3}$$

Here P defines the center point of the target box, O_p is the prediction of the basis, and R represents the downsample ratio, which is configured as 4 in the experimentation. P_p denotes the estimated P/R, so $(P/R - P_p)$ is the absolute offsets. The loss function concerning the target box is calculated as follows:

$$L_{size} = \frac{1}{N}\sum_{k=1}^{N}\left|S_p - S\right| \tag{4}$$

For class c targets (experiment $c = 1, 2, \ldots 6$), the real target box coordinates of the k_{th} target are $(x_1^k, y_1^k, x_2^k, y_2^k)$, and the estimated target box coordinates are $(\hat{x}_1^k, \hat{y}_1^k, \hat{x}_2^k, \hat{y}_2^k)$, then the real size of the target box is $S = (x_2^k - x_1^k, y_2^k - y_1^k)$, and the prediction of the target box is $S_p = (\hat{x}_2^k - \hat{x}_1^k, \hat{y}_2^k - \hat{y}_1^k)$. The total loss is with different weights assigned, and the formula is as follows:

$$L_{det} = L_k + \omega_{off}L_{off} + \omega_{size}L_{size} \tag{5}$$

where ω denotes the corresponding weight, both ω_{off} and ω_{size} are considered as 1 and 0.1, respectively.

4.4. Network Parameters and Evaluation Indicators

Operating parameters in the training process are exhibited in Table 1, which illustrates that Batch_size is configured as 16 images. Hence, both the training speed and the gradient are enhanced. Down_ratio and Lr represent the down-sampling and learning rates, respectively. When the epoch increases, the learning rate is towards 0.1. Then, the network could accelerate the convergence speed at the beginning phase. The change of loss is shown in Figure 7.

Table 1. Parameters for training

Parameters	Value
Batch_size	16
Down_ratio	4
flip	0.5
lr	1.25×10^{-4}
lr_step	60,90
Reg_loss	L1
Val_intervals	5
epoch	150

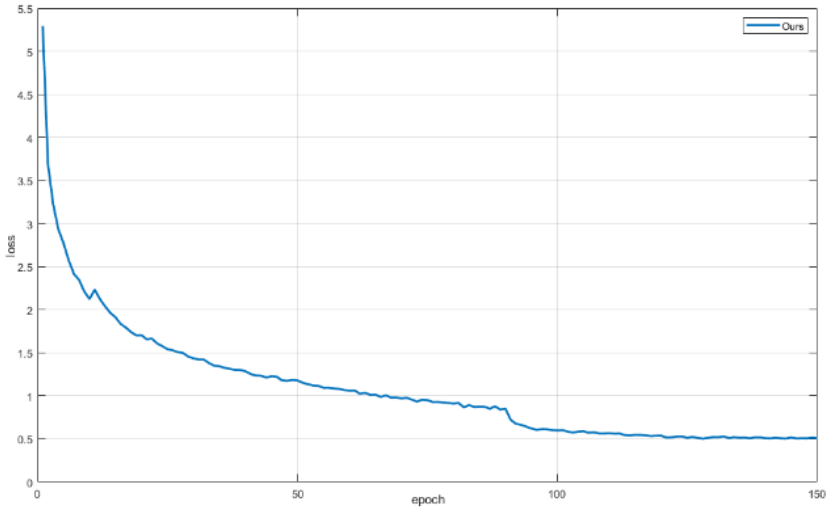


Figure 7. Loss changes during training.

The mAP (mean average precision) is also employed as a third-party operator for qualities and quantitative investigation. The final indicators and corresponding performances are summarized in Table 2. Experimental results demonstrate that the proposed approach has better performance than the state-of-the-art approaches.

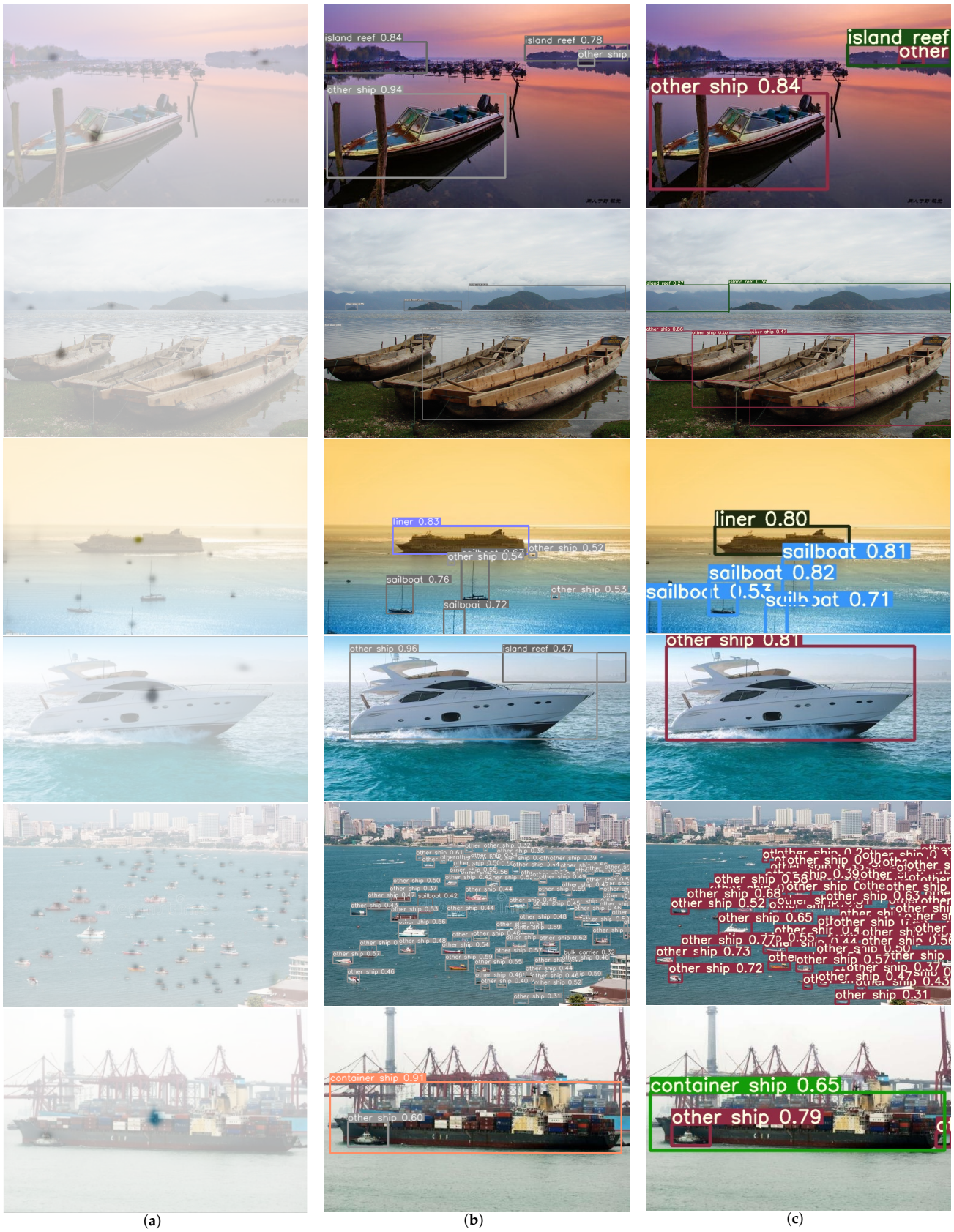
Table 2. Performance comparison.

Model	Backbone	mAP	AP_{50}	AP_{75}
DSSD [36]	ResNet-101	44.1	66.3	44.8
DeNet [37]	ResNet-101	45.8	68.4	46.0
RetinaNet800 [38]	ResNet-101	47.2	72.9	53.4
CenterNet	DLA-34	51.0	75.7	55.7
CornerNet	Hourglass-104	51.2	72.8	56.9
MaskRCNN [39]	ResNeXt-101	51.8	77.5	54.4
ExtremeNet [40]	Hourglass-104	52.1	76.8	57.3
Cascade R-CNN [41]	ResNet-101	54.9	77.1	57.6
PANet [42]	ResNeXt-101	55.7	74.9	59.3
YOLOv5	CSP	56.2	76.7	**61.2**
Ours	DLA-34 *	**56.6**	**77.8**	59.7

DLA34 * represents the improved DLA in this paper. Bold showed the best performance in AP75.

4.5. Discussion

As shown in Figure 8, the first column is the output of the branch heatmap, where each local maximum is considered a target. The size branch regresses the size position of the object through these points. The second column is the detection results of the proposed method, and the last column is the results of other approaches (YOLOv5, for example). The downsampling rate of the proposed method is only 4, which ensures that the complete image information is retained in convolution. In addition, the modified network structure can integrate in-depth and shallow features to have a comprehensive understanding of global information. Thus, the detection effect of both long-range and small targets is more promising, which can be illustrated in island reefs in columns 1, 2, and 4. Meanwhile, by improving the DLA module, the proposed strategy gives additional attention to tiny objects, such as the "other ship" to the right in line 2 and the "other ship" to the upper left in line 5. However, an anchor-free method also has drawbacks: in scenarios of targets overlapping, the targets may be combined due to heatmap branch output, resulting in reduced confidence and even missed detection, as can be seen in columns 2 and 6.



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

Figure 8. Results analysis. (**a**) heatmap branch output (**b**) the proposed method detection results (**c**) YOLOv5 detection results

5. Conclusions

This paper proposes a ship detection method that addresses imbalance distribution issues regarding category, scale, and quantity. The proposed model is within the framework of the anchor-free object detection; specifically, by redesigning the DLA network structure, the feature extraction ability of small targets is enhanced. The number of parameters is significantly reduced by improving the convolution module, and the real-time performance is guaranteed. Experimental results demonstrate that the proposed method has higher accuracy than anchor-based methods, especially in small target scenarios. In addition, the mAP improves by 5.6% compared to the original model.

In the future, a more compact and efficient network architecture will be employed to solve this problem. Furthermore, given the lack of datasets in this field, weakly supervised and few-shot learning should be further considered.

Author Contributions: Funding acquisition, F.Z.; supervision, F.Z.; writing—original draft, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC), grant number 52171322.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible https://drive.google.com/file/d/1XwXfFYmrdTzutelblsDRCWXLhQ0MyfP-/view?usp=sharing (accessed on 10 December 2021).

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

Abbreviations

The following abbreviations are used in this manuscript:

HOG	Histograms of Oriented Gradient
SVM	Support Vector Machines
GAN	Generative Adversarial Network
YOLO	You Only Look Once
SSD	Single Shot MultiBox Detector
RPN	Region Proposal Network
NMS	Non-Maximum Suppression
HDA	Hierarchical Deep Aggregation
IDA	Iterative Deep Aggregation

References

1. Raghunandan, A.; Mohana; Raghav, P.; Aradhya, H.V.R. Object Detection Algorithms for Video Surveillance Applications. In Proceedings of the 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 3–5 April 2018; pp. 0563–0568. [CrossRef]
2. Wahyutama, A.B.; Hwang, M. YOLO-Based Object Detection for Separate Collection of Recyclables and Capacity Monitoring of Trash Bins. *Electronics* **2022**, *11*, 1323. [CrossRef]
3. Mane, S.; Mangale, S. Moving Object Detection and Tracking Using Convolutional Neural Networks. In Proceedings of the 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 14–15 June 2018; pp. 1809–1813. [CrossRef]
4. Ajmera, F.; Meshram, S.; Nemade, S.; Gaikwad, V. Survey on Object Detection in Aerial Imagery. In Proceedings of the 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 4–6 February 2021; pp. 1050–1055. [CrossRef]
5. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot.* **2020**, *37*, 362–386. [CrossRef]
6. Liu, J. Survey of the Image Recognition Based on Deep Learning Network for Autonomous Driving Car. In Proceedings of the 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT), Shenyang, China, 13–15 November 2020; pp. 1–6. [CrossRef]
7. Ruiz-del Solar, J.; Loncomilla, P.; Soto, N. A survey on deep learning methods for robot vision. *arXiv* **2018**, arXiv:1803.10862.

8. Chen, J.; Li, K.; Deng, Q.; Li, K.; Philip, S.Y. Distributed deep learning model for intelligent video surveillance systems with edge computing. *IEEE Trans. Ind. Inform.* **2019**, 1. [CrossRef]
9. Gautam, A.; Singh, S. Trends in Video Object Tracking in Surveillance: A Survey. In Proceedings of the 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 12–14 December 2019; pp. 729–733. [CrossRef]
10. Sitaula, C.; Shahi, T.B.; Aryal, S.; Marzbanrad, F. Fusion of multi-scale bag of deep visual words features of chest X-ray images to detect COVID-19 infection. *Sci. Rep.* **2021**, *11*, 23914. [CrossRef] [PubMed]
11. Mittal, S.; Srivastava, S.; Jayanth, J.P. A Survey of Deep Learning Techniques for Underwater Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–15. [CrossRef] [PubMed]
12. Zou, Y.; Zhao, L.; Qin, S.; Pan, M.; Li, Z. Ship target detection and identification based on SSD_MobilenetV2. In Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; pp. 1676–1680.
13. Yu, H.; Li, Y.; Zhang, D. An Improved YOLO v3 Small-Scale Ship Target Detection Algorithm. In Proceedings of the 2021 6th International Conference on Smart Grid and Electrical Automation (ICSGEA), Kunming, China, 29–30 May 2021; pp. 560–563. [CrossRef]
14. Ting, L.; Baijun, Z.; Yongsheng, Z.; Shun, Y. Ship Detection Algorithm based on Improved YOLO V5. In Proceedings of the 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China, 15–17 July 2021; pp. 483–487. [CrossRef]
15. Shahi, T.B.; Sitaula, C.; Neupane, A.; Guo, W. Fruit classification using attention-based MobileNetV2 for industrial applications. *PLoS ONE* **2022**, *17*, e0264586. [CrossRef] [PubMed]
16. Zhang, C. DGANet: Dynamic Gradient Adjustment Anchor-Free Object Detection in Optical Remote Sensing Images. *Remote Sens.* **2021**, *13*, 1642.
17. Hou, X.; Zhang, F. The Improved CenterNet for Ship Detection in Scale-Varying Images. In Proceedings of the 2021 3rd International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 8–11 November 2021; pp. 1–5. [CrossRef]
18. Matsumoto, Y. Ship Image Recognition using HOG. *J. Jpn. Inst. Navig.* **2013**, *129*, 105–112. [CrossRef]
19. Shi, G.; Suo, J. Ship Targets Detection Based on Visual Attention. In Proceedings of the 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Qingdao, China, 14–16 September 2018; pp. 1–4. [CrossRef]
20. Qiuyu, Z.; Yilong, J.; Bo, C. Design and implementation of video-based detection system for wharf ship. In Proceedings of the IET International Conference on Smart and Sustainable City 2013 (ICSSC 2013), Shanghai, China, 19–20 August 2013; pp. 493–496. [CrossRef]
21. Jin, B.; Cong, Y.; Zhou, W.; Wang, G. A new method for detection of ship docked in harbor in high resolution remote sensing image. In Proceedings of the 2014 IEEE International Conference on Progress in Informatics and Computing, Shanghai, China, 16–18 May 2014; pp. 341–344. [CrossRef]
22. Wang, Y.; Ning, X.; Leng, B.; Fu, H. Ship Detection Based on Deep Learning. In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2019; pp. 275–279. [CrossRef]
23. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
24. Zhao, Q.; Sheng, T.; Wang, Y.; Ni, F.; Cai, L. CFENet: An accurate and efficient single-shot object detector for autonomous driving. *arXiv* **2018**, arXiv:1806.09790.
25. Chen, Z.; Chen, D.; Zhang, Y.; Cheng, X.; Zhang, M.; Wu, C. Deep learning for autonomous ship-oriented small ship detection. *Saf. Sci.* **2020**, *130*, 104812. [CrossRef]
26. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
27. Zhao, H.; Zhang, W.; Sun, H.; Xue, B. Embedded Deep Learning for Ship Detection and Recognition. *Future Internet* **2019**, *11*, 53. [CrossRef]
28. 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, 18–23 June 2018; pp. 4510–4520.
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef] [PubMed]
30. 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, 14–19 June 2020; pp. 1580–1589.
31. Yu, F.; Wang, D.; Shelhamer, E.; Darrell, T. Deep layer aggregation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2403–2412.
32. Newell, A.; Yang, K.; Deng, J. Stacked hourglass networks for human pose estimation. In *Computer Vision—ECCV 2016, Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 483–499.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Misra, D. Mish: A self regularized non-monotonic activation function. *arXiv* **2019**, arXiv:1908.08681.

35. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
36. Fu, C.Y.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. Dssd: Deconvolutional single shot detector. *arXiv* **2017**, arXiv:1701.06659.
37. Tychsen-Smith, L.; Petersson, L. Denet: Scalable real-time object detection with directed sparse sampling. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 428–436.
38. Zhang, H.; Chang, H.; Ma, B.; Shan, S.; Chen, X. Cascade retinanet: Maintaining consistency for single-stage object detection. *arXiv* **2019**, arXiv:1907.06881.
39. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
40. Zhou, X.; Zhuo, J.; Krahenbuhl, P. Bottom-up object detection by grouping extreme and center points. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seoul, Korea, 27 October–2 November 2019; pp. 850–859.
41. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6154–6162.
42. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 9197–9206.