M. Chen et al. (Eds.)

doi:10.3233/ATDE230549

Water Floating Garbage Detection Algorithm Based on Improved YOLOv7-Tiny

Xuan DU [a,1] and Zihao XIE [a]
[a] School of Mechanics and Power, China Three Gorges University, Hubei Key Laboratory of Design and Maintenance of Hydroelectric Machinery Equipment, China Three Gorges University, Hubei 443002 China

Abstract. This paper presents a SC-YOLOv7_tiny model based on YOLOv7-tiny, combined with a UAV platform to detect floating garbage in water areas, addressing the low efficiency and high cost of traditional water area floating garbage detection. Firstly, a skip connection-based feature pyramid network (SCFPN) is introduced to detect small objects by combining features of different scales. Secondly, to improve the CIOU loss function, the weight of the angle penalty term θ is used to penalize differences in length and width scales, thus mitigating the problems of feature distribution imbalance, scale mismatch, and difficulty in detecting small objects when combined with the SCFPN feature pyramid. Finally, the SimAM attention mechanism is introduced into the SCFPN feature pyramid module to facilitate the model’s understanding of contextual information and enhance its feature expression capabilities. To verify the effectiveness of the algorithm, this paper uses a public dataset on Universe for testing. Experimental results show that the parameters of the SC-YOLOv7-tiny model are significantly reduced while detection accuracy is improved.

Keywords. UAV, floating garbage, small target, deep learning

1. Introduction

Lakes and streams improve and beautify the urban environment. However, with the acceleration of urbanization and population growth, the water environment is facing increasingly serious pollution problems. Timely and effective monitoring and cleaning of floating garbage in water areas is crucial to improving the city's image.

The UAV-based visual detection technology for floating garbage in water involves image processing, deep learning and other disciplines. Traditional feature-based methods mainly divide the image into multiple regions according to the brightness, color [1] and texture [2] of the pixels in the image, and determine which regions contain garbage by setting a threshold. For example, literature [3] realized the effective segmentation and extraction of floating objects on the water surface by improving the background update method of the traditional Gaussian mixture model algorithm. Using deep learning to detect floating garbage in waters mainly uses deep learning models

[1] Xuan DU, Corresponding author, School of Mechanics and Power, China Three Gorges University, Hubei Key Laboratory of Design and Maintenance of Hydroelectric Machinery Equipment, China Three Gorges University, Hubei 443002 China; E-mail: 1468179488@qq.com.

such as convolutional neural networks, such as YOLO [4], SSD [5] and Faster RCNN [6], to extract features from images. For small targets, the detection accuracy is low, and problems such as missed detection and false detection are prone to occur. Literature [7] uses the improved YOLOV5 algorithm to classify and identify floating garbage datasets from the perspective of drones, but the overall detection accuracy still needs to be improved. Literature [8] solves the problem of small-scale objects and domain drift by retaining high-resolution feature maps while filtering out low-resolution feature maps. However, the algorithm has many model parameters, and the calculation cost increases. Although the YOLOv7 model performs well in common task scenarios such as pedestrian [9,10] and vehicle [11] detection, there are some problems in directly using the original network for the detection of floating garbage in water. When the loss function is combined with the SCFPN feature pyramid, it will lead to an unbalanced feature distribution, resulting in inaccurate small object detection results. In addition, the original YOLOv7-tiny network did not introduce an attention mechanism to pay more attention to small objects. Therefore, this article improves on the basis of YOLOv7-tiny.

2. Model Structure Improvement of YOLOv7-tiny Network

2.1. Improvement of Feature Pyramid Structure

Due to the small size of floating garbage targets in water environments from a drone's perspective and the low resolution of small target objects, they are difficult to detect. In the feature fusion process of the original YOLOv7-tiny, the deeper feature map is upsampled and fused with the shallow feature map, resulting in the loss of details of the small target.

To address the aforementioned issues, this paper adds a layer of multi-branch superposition module ELAN between UP and Concat, so that the second layer output of the backbone network passes through the CBL module and is directly spliced with the superposition module ELAN, and the fusion feature map is obtained. The connection method adopts skip connection, transfers the bottom feature directly to the top layer, and realizes the feature fusion through the overlay stitching operation of ELAN module, effectively avoiding the loss of small details of the target. The improved feature pyramid network SCFPN is shown in figure 1.

Figure 1. Schematic diagram of the network model after improving the feature pyramid.

2.2. Improvement of Loss Function CIOU

The original YOLOv7-tiny network uses CIOU as the regression loss function. The formula for calculating CIOU is:

$$CIOU = IOU - \frac{\rho^2}{c^2} - \alpha \cdot v \tag{1}$$

$$\alpha = \frac{v}{1 - IOU + v} \tag{2}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{3}$$

In order to solve the problem of uneven feature distribution and scale mismatch, we introduce an aspect ratio penalty term, and the penalty is calculated as follows:

$$\_CIOU_{loss} = IOU - \frac{\rho^2}{c^2} - \alpha \cdot v - \theta \cdot \frac{\arctan\left(\frac{w_{gt}}{h_{gt}}\right) - \arctan\left(\frac{w}{h}\right)}{\pi^2} \tag{4}$$

In the formula, ρ represents the Euclidean distance between the center point of the predicted boundary frame and the center point of the ground true boundary frame ; c represents the diagonal distance between the minimum enclosing area of the predicted and ground truth bounding boxes; α is a weight function; v measures the similarity of aspect ratios; w^{gt}, w, h^{gt}, h indicate the width and height of the real and predicted boxes.θ represents the weight of the angle penalty term, with a value of 0.25.

2.3. Using SimAM Attention Mechanism

Due to the smaller size of small objects, more details are needed for accurate detection. However, obtaining high-resolution feature maps requires increasing the depth and complexity of the network, which increases the computational and memory overhead. To solve this problem, the SimAM [12] attention mechanism is introduced into the SCFPN feature pyramid module, which enables SCFPN to select and fuse the most representative features from multiple scales and levels, thus providing a more diversified feature representation.

3. Experimental Verification

3.1. Dataset

In order to evaluate the detection performance of the network model in different scenarios, a small floating garbage dataset is constructed. First, floating garbage datasets from two different scenarios are combined to expand the sample size. Next, re-label the unqualified label box using the LabelImg tool. Finally, the training set, verification set and test set are divided by 8:1:1 to complete the small floating garbage data set.

3.2. Experimental Environment Settings

The experimental environment of this paper is Windows 11 operating system,CPUi5-12490F and RTX-3060 graphics card. Complete the environment setup using Python 3.10, PyTorch 1.12, and CUDA 11.3.

3.3. Analysis of Experimental Results

Both the YOLOv7-tiny model and the improved YOLOv7-tiny model were trained using the same dataset and configuration parameters. During training, the evaluation loss is saved for each epoch. Based on the saved loss file, we plot the model evaluation loss val_loss comparison curve shown in figure 2.

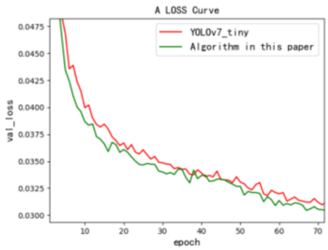

Figure 2. Comparison of evaluation loss before and after improvement.

Where the horizontal and vertical coordinates denote the number of network iterations and evaluation loss, respectively. It can be seen from the figure that the improved model in this paper converges faster and has a smaller loss value than the baseline model. This shows that the improved model has better convergence performance in the evaluation process .

3.3.1. Comparative Experiment

To verify the reliability of the improved network, we compared the results of the SC-YOLOv7-tiny model with those of the SSD, Faster-RCNN, and original YOLOv7-tiny algorithms. The experimental results are shown in table 1.

Table 1. Performance comparison of different algorithm models on datasets.

Network Type	Input Size	Recall Rate	mAP@0.5	Model parameter size/MB
Faster-RCNN	640*640	86.14%	83.34%	108MB
SSD	640*640	63.84%	86.41%	90.6MB
YOLOv7-tiny	640*640	85.91%	90.74%	23.1MB
SC-YOLOv7-tiny	640*640	87.85%	91.42%	21.8MB

Experiments show that compared with Faster-RCNN and SDD models, the parameter number of the improved SC-YOLOv7-tiny model is reduced by about 4 times, and mAP@0.5 is significantly increased. There were improvements in mAP@0.5 and recall rates and a 5.6% reduction in the number of model references compared to the baseline model. The improved method can effectively improve the accuracy of network detection and reduce the number of model parameters.

3.3.2. Ablation Experiment

Through ablation experiments, the optimization effect of each improved module is fully verified. We use mAP@0.5 and model size to evaluate the improvement. The experimental comparison results are shown in table 2.

Table 2. Ablation experiments.

Model	Attention Mechanism	Feature Pyramid	Loss Function	mAP@0.5	Model parameter size/M
YOLOv7-tiny				90.74%	23.1M
YOLOv7-tiny+SCFPN		√		91.27%	21.7M
YOLOv7-tiny +SCFPN+_CIOU		√	√	91.39%	21.7M
SC-YOLOv7-tiny	√	√	√	91.42%	21.8M

The experiment shows that the parameters of the model can be reduced effectively and the detection accuracy can be improved by increasing the depth of the network intermediate layer. The improved SC-YOLOv7-tiny model integrates SCFPN, improved CIOU and SimAM to promote each other and is more suitable for small target detection scenarios.

3.4. Visual Inspection Result Analysis

Floating garbage in two different situations was selected from the test set to verify the feasibility of the model. In scenario 1, the amount of garbage in a single frame image is small; in scenario 2, the amount of garbage in a single frame image is large and dense. The detection effect is shown in figure 3.

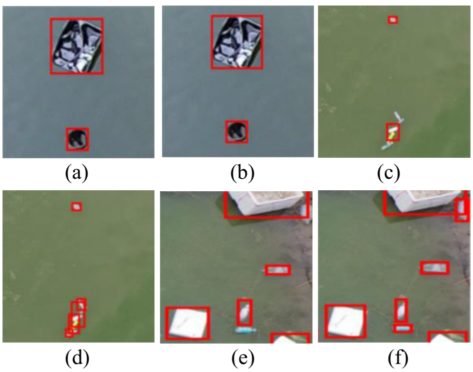

Figure 3. (a) Represents the detection results of the original model under scenario 1. (b) Represents the detection results after improving the model under scenario 1. (c) and (e) represent the detection results of the original model under scenario 2. (d) and (f) represent improved model detection results under scenario 2.

Both YOLOv7-tiny and SC-YOLOv7-tiny networks can accurately detect garbage when there are fewer target objects in a single frame. But as the amount of garbage increased, the YOLOv7-tiny model began to miss detection. The SC-YOLOv7-tiny network has advantages in intensive target detection tasks, enabling more accurate detection of small targets, avoiding additional missed detection and secondary detection costs. Although the overall detection accuracy is only improved by 0.68%, this makes SC-YOLOv7-tiny more valuable in practical applications.

4. Conclusion

In this paper, by modifying the feature pyramid, introducing attention mechanism and modifying loss function, the model can obtain the optimal detection effect on small target and dense target tasks. Compared with baseline model and classical target detection model, SC-YOLOv7-tiny has higher detection accuracy and real-time performance. This algorithm provides important technical support for water area monitoring and has a wide application prospect.

References

[1] Sun T. Research on background segmentation algorithm of moving image in speed skating video. Journal of Computational Methods in Sciences and Engineering. 2019; 19(4): 1037-1053.
[2] Xue P. Foreground and background segmentation based on super piexel-level feature representation. Journal of Xi'an University of Science and Technology. 2017; (05):731-735.
[3] Niu PW. Research on intelligent monitoring technology of floating objects on water surface based on background difference. Nanjing University of Posts and Telecommunications, 2020.
[4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; pp. 779-788.
[5] Berg AC, Fu CY, Szegedy C, et al. SSD: Single shot multibox detector. 10.1007/978-3-319-46448-0_2. 2015.
[6] Ding X, Li Q, Cheng Y, et al. Local keypoint-based faster R-CNN. Applied Intelligence. 2020; 50: 3007-3022.
[7] Li DX, Yan ZG, Sun JY. Research on classification and detection technology of floating garbage in river based on UAV vision. Metal Mine. 2021; (9):7.
[8] Chen RF, Peng Y, Li ZW. Target detection method for floating objects on water surface based on continuous unsupervised domain adaptation strategy. Systems Engineering and Electronic Technology. 2023.
[9] Wang H, Jin L, He Y, et al. Detector–tracker integration framework for autonomous vehicles pedestrian tracking. Remote Sensing. 2023; 15(8): 2088.
[10] Patel D, Patel S, Patel M. Application of image-to-image translation in improving pedestrian detection. arXiv preprint arXiv:2209.03625, 2022.
[11] Zhang Y, Sun Y, Wang Z, et al. YOLOv7-RAR for urban vehicle detection. Sensors. 2023; 23(4): 1801.
[12] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks. International conference on machine learning. PMLR, 2021; pp. 11863-11874.