Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/17/20230920

Research on identification of floating garbage using improved YOLO v7 algorithm

Weixuan Li[1,3,†] , Haotian Zhu[2,†]

[1]School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin 300387, China
[2]School of Engineering and Informatics, University of Sussex, Brighton, BN19RJ,UK

[3]2010920516@tiangong.edu.cn
[†]All authors contributed equally.

Abstract. The floating garbage is becoming more and more serious, but little research has addressed recognition of these floating garbage. Intelligent target recognition of floating garbage using deep learning techniques is therefore essential. The YOLOv7 algorithm has strong ability of extracting target features and is significantly faster than its previous version at the same accuracy. The technique is provide based on the YOLOv7 algorithm for identifying floating garbage in this paper, as a result, develop and implement a target monitoring function for floating garbage identification. Specifically, the combination of YOLO v7 and SE attention mechanism was used to improve ability of target sensing. The training process was optimized by using EIOU loss, resulting in a significant improvement in the efficiency of the final model compared to the normal YOLOV7 algorithm model, significant improvements of 20% and 25% in the model metrics mAP_0.5:0.95 and mAP_0.5, respectively.

Keyword: floating garbage, deep learning, target detection, YOLOv7.

1. Introduction

Water is the source of life and an important resource for human survival. However, with the accelerated urbanisation process, waste from various human activities such as industrial production and residential life is increasingly entering water bodies, leading to an increasingly serious problem of water pollution. Water surface litter not only affects the aesthetics of water bodies, but also has an adverse effect on aquatic organisms and even causes a series of environmental problems [1-3]. Therefore, the effective identification and treatment of surface litter is of great practical importance. Traditional inland river litter cleaning often relies on manual salvage, which is often inefficient and requires safety and other considerations, so it is not the best option for river cleaning. However, current unmanned cleaning boats also do not handle litter well, due to their current recognition efficiency and accuracy [4,5].

Therefore, in this study, using the conventional yolov7 technique, and the SE concern mechanism is added and the EIOU-loss is adopted as the new target frame regression loss function, and the PyQt interface program is designed on this basis, so as to truly realize the operation of identifying realistic water body litter and greatly improve the efficiency of water body litter treatment.

2. Method

This section introduces the main features of the YOLOv7 algorithm, gives the construction principles of the fusion of YOLOv7 and SE attention mechanisms, and where the advantages of using EIOU Loss over CIOU Loss lie.

2.1. Overview of the YOLOv7 algorithm

The YOLOv7 technique can be utilised in real-time systems and is a typical one-stage target identification algorithm based on deep neural networks for object recognition and localization [2,3]. In order to comply with the Backbone's requirements for input size, the input module scales the input image to a predetermined size. The CBS convolutional layer, the E-ELAN convolutional layer, and the MPConv convolutional layer make up the backbone network. This efficient layer aggregation network learns a range of characteristics by directing the computational modules of different feature groups, improving the network's learning capability without erasing the original gradient paths.

It creates the upper and lower branches before using the Constant operation to combine the features the upper and lower branches brought up to enhance the network's ability to extract data. The head network first uses the SPP pyramid structure to make the head network suitable for multi-scale input; then the aggregated feature pyramid network structure is used to pass the bottom information along the bottom-up path to the top to achieve the fusion of different covariance features; finally, the number of channels is adjusted by the REPcon structure for features of different scales [4]. Three components make up the YOLOv7 loss function: the localization function, the confidence loss, and the classification loss. The confidence loss and classification loss employ BCELoss binary cross-entropy loss, whereas the localization loss uses CIoU loss. The network architecture is shown in Figure 1.

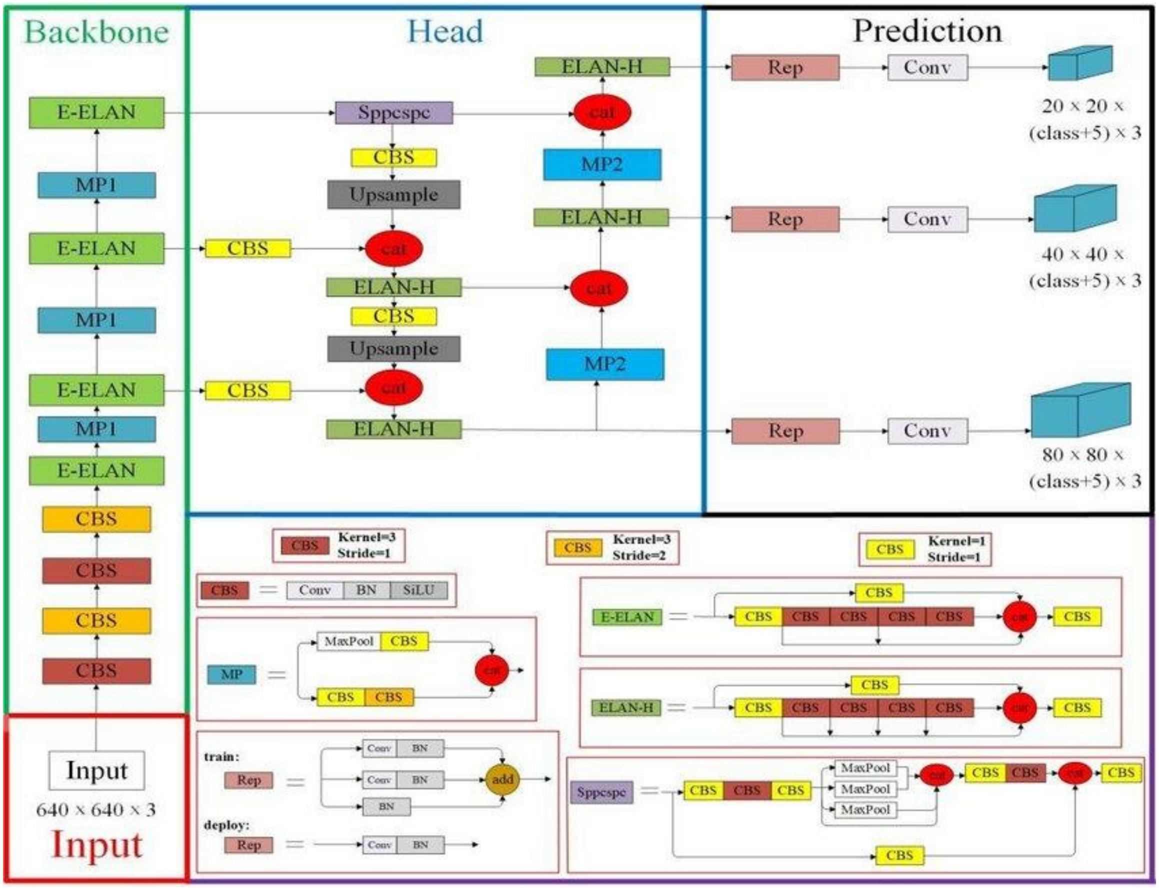

Figure 1. YOLOv7 algorithm structure.

DOI: 10.54254/2755-2721/17/20230920

2.2. Introduction to the SE attention mechanism

Squeeze and excitation are the two primary activities of SE attention mechanisms (Squeeze-and-Excitation Networks), which add attention mechanisms to the channel dimension [6].

By autonomously learning the importance between different channels in a neural network's feature map, and then setting different weight values to each feature based on this importance, so that it focuses on the feature channel with the greater weight. In this way the efficiency of the channels that are useful for the task at hand is improved [7,8].

Figure 2 shows that each channel in the feature map has the same amount of relevance when the SE attention mechanism is not present (see Figure 2 left C below). Following SENet, different colours signify different weights, giving each feature channel a varied level of priority and enabling the neural network to concentrate on select channels with high weight values.

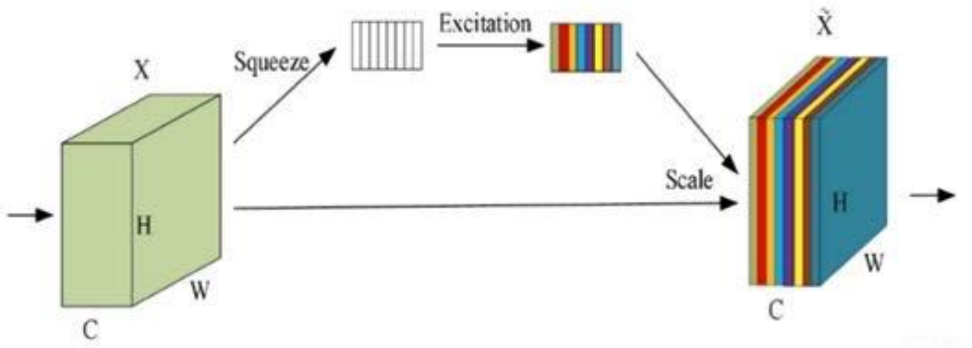

Figure 2. Structure of the SE attention mechanism.

Inserting the attention mechanism into the YOLOv7 algorithm structure:

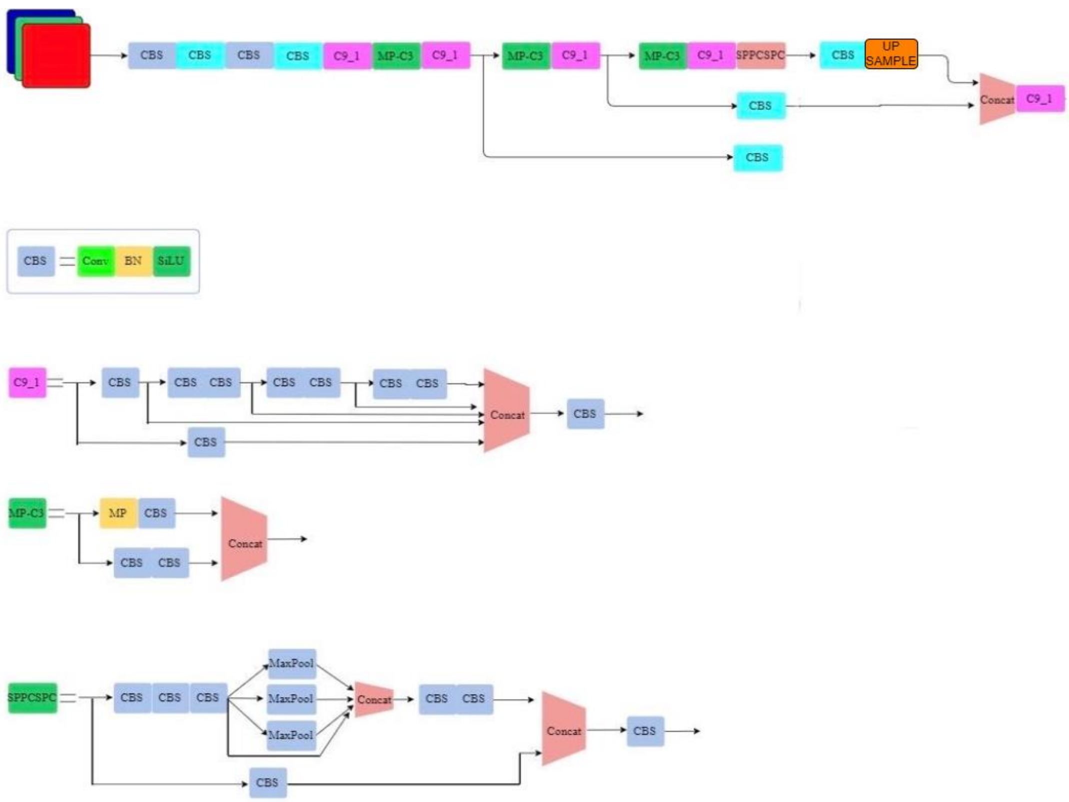

Figure 3. Structure of the YOLOv7 algorithm with the addition of the SE attention mechanism.

DOI: 10.54254/2755-2721/17/20230920

The SE attention mechanism is added to the YOLOv7 algorithm structure, as seen in Figure 3, to enhance the performance of the model. The SE attention method concentrates on the feature interactions between channels, amplifies the key features, and suppresses the unimportant features by adaptively recalibrating the channel feature responses. The YOLOv7 algorithm with the SE attention mechanism is implemented in this structure by substituting the first four C9_1s without the attention mechanism with the last channel of the C9_1 composite convolutional layer.

The rationale for adding the SE attention mechanism to the convolutional layer is that it can effectively capture the relationships between spatial features and thus improve the quality of the feature representation. The inclusion of the SE attention mechanism enhances feature representation, allows the model to better focus on important features and improves model performance in terms of recognition accuracy and generalisation, as well as making the model more interpretable by adaptively adjusting the channel response.

2.3. Advantages of EIOU Loss

CIOU loss considers the border box regression's aspect ratio, centroid distance, and overlapping area [9]. The model occasionally is unable to legitimately optimise the degree of similarity since, in its formulation. Instead of the difference between the actual width and height and its confidence level, v represents the variance in aspect ratio.

Based on the CIOU penalization term, the EIOU calculates the target and anchor boxes' length and breadth by dividing the aspect ratio influences [10]. The three components of the loss function are the overlap loss, centre distance loss, and width-height loss.

As a result, by utilising EIOU, the aspect ratio loss term can be split into the difference between the expected width and height and the smallest outer box width and height, hastening convergence and improving the regression's accuracy.

3. Analysis of experiments

This chapter focuses on the analysis and reflection of the experimental results using ablation experiments through the presentation of the data set.

3.1. Construction and production of the dataset

The training process mainly includes the construction of the dataset and the application of migration learning. The dataset is made up of a training set and a test set that are both drawn from the network and contain information about floating trash on the water's surface. Lablimage is used to manually calibrate the floating trash in the training set, which consists of 2700 training images and 300 test images.

The size of the images was 640×640, and six images were trained in each batch, with 100 learning rounds. A total of 11 categories of litter were classified, namely can, leaf, branch, grass, bottle, milk_box, plastic_bag, glass_bottle, bad_fruit, ball. The surface litter dataset is shown in Figure 4.

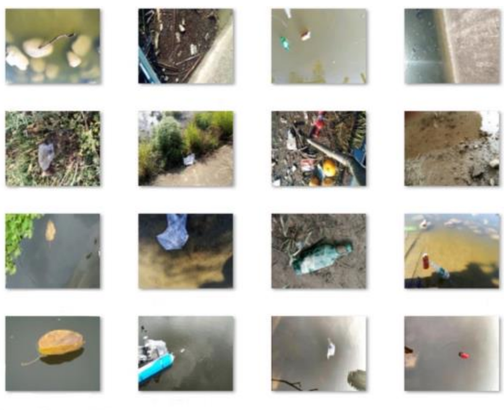

Figure 4. Data set of floating litter on the water surface.

DOI: 10.54254/2755-2721/17/20230920

3.2. Ablation experiments

In the subsequent trials, we concentrate on measuring the algorithm's efficiency under various models to assess whether the inclusion of the SE attention mechanism and adjusting the various loss functions have increased the model's effectiveness.

Table 1. F1 values and mAPs for different models.

Method	Indicators (F1 values)	mAP_0.5:0.95	mAP_0.5
YOLO V7 + CIOU loss	0.60	0.404	0.6067
YOLO V7 + EIOU loss	0.67	0.4528	0.6916
YOLO V7 + SENet-CIOU loss	0.75	0.5376	0.797
YOLO V7 + SENet-EIOU loss	0.83	0.6033	0.8536

The table 1 illustrates that the average surface litter recognition accuracy mAP_0.5 was 60.67% when the YOLOV7 algorithm was not improved by EIOU loss and the SE attention mechanism was added. While the average surface litter recognition accuracy mAP_0.5 improved by 24.69 percentage points after EIOU loss and SE attention mechanism were added in turn. This demonstrates that the EIOU loss and SE attention mechanisms have a significant impact in the enhancement of surface litter recognition skills.

- Comparison of the effects of adding the SENet algorithm:

The comparison of the confusion matrix shows that the overall prediction probability of the model improves by 26% with the addition of SE attention processing compared to no SENet (Figure 5-6). The five categories of milk_box, plastic_box, glass_bottle, bad_fruit, and ball achieved a 100% probability of recognition in the experimental test set. For milk_box, glass_bottle, bad_fruit, the most significant improvement was achieved, with glass_bottle achieving a 50% improvement compared to no SENet.

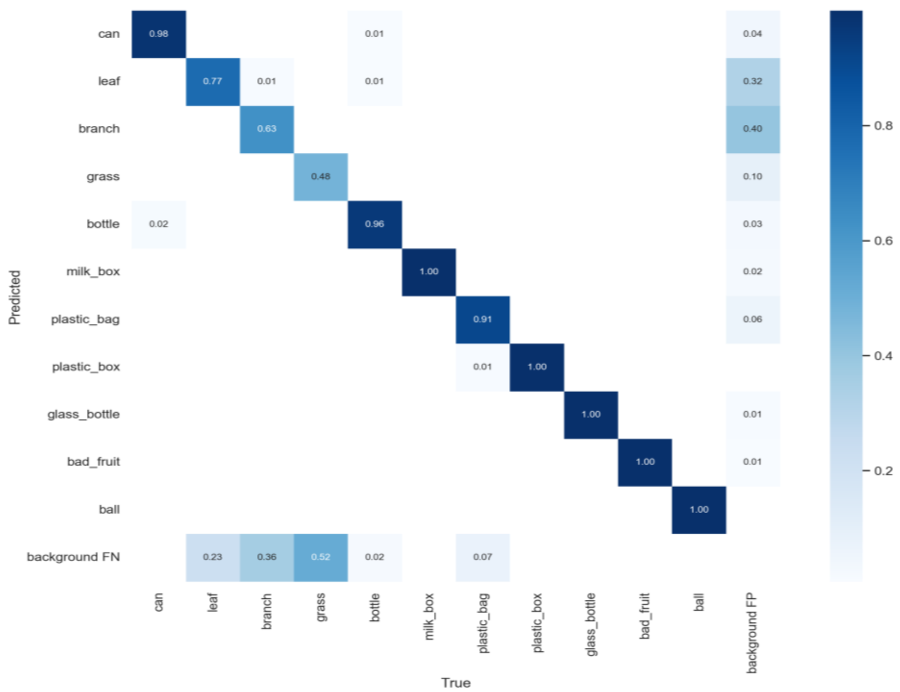

Figure 5. Confusion matrix under EIOU + SE Attention.

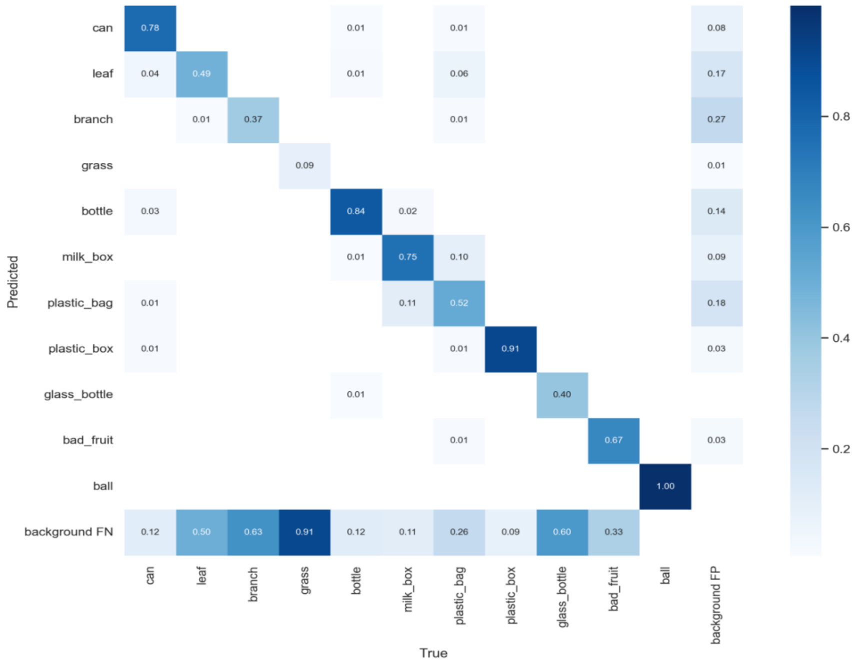

Figure 6. Confusion matrix under EIOU + no SE Attention.

The comparison of the F1 values of the different categories with and without Senet (Figure 7-8). Result shows that the overall F1 values are significantly improved with SEnet, with the confidence interval of 0.2-0.8 for milk_box, plastic_box, bad_fruit, ball, bottle reaching above 0.9, However, leaf, branch, and grass did not improve significantly for SEnet. The reason for this is that the three categories of leaf, branch and grass are more similar and can be easily classified into one category, and these three categories are more difficult to classify than the other categories.

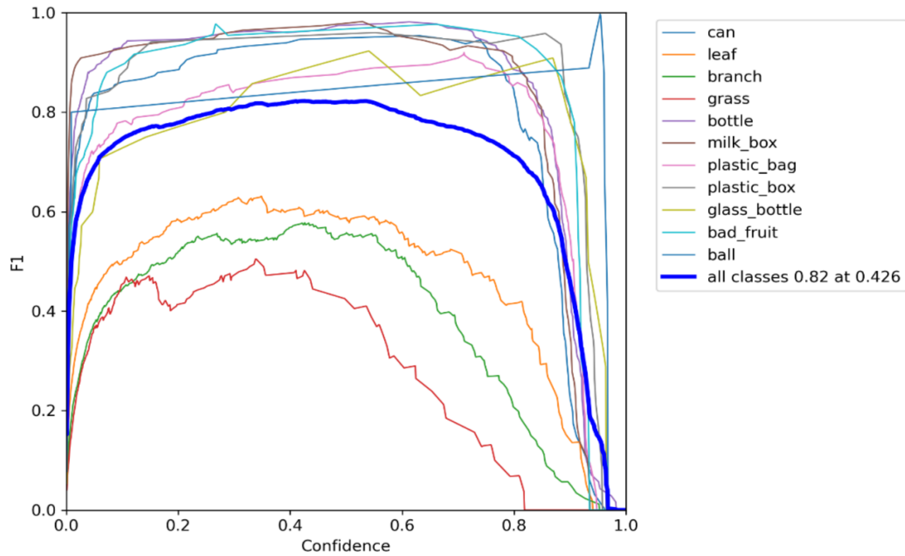

Figure 7. F1 values under EIOU + SEAttention.

DOI: 10.54254/2755-2721/17/20230920

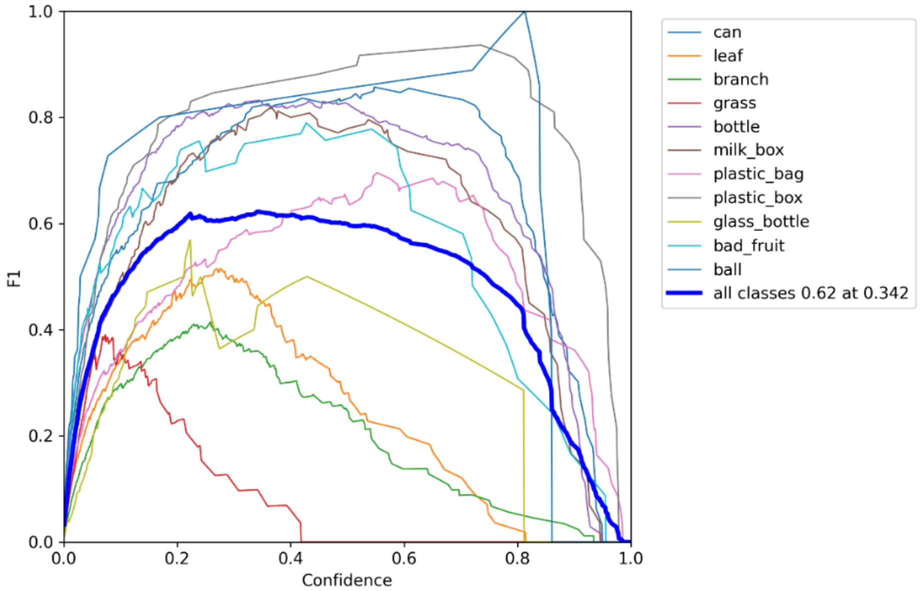

Figure 8. F1 values under EIOU + no SE Attention.

- Comparison of the effects of using EIOU loss and CIOU loss:

From the Figure 9-10, it can be seen that using EIOU as the loss function improves the overall recognition probability by 11.5% compared to CIOU, with glass_bottle and grass improving significantly by 30% compared to CIOU.

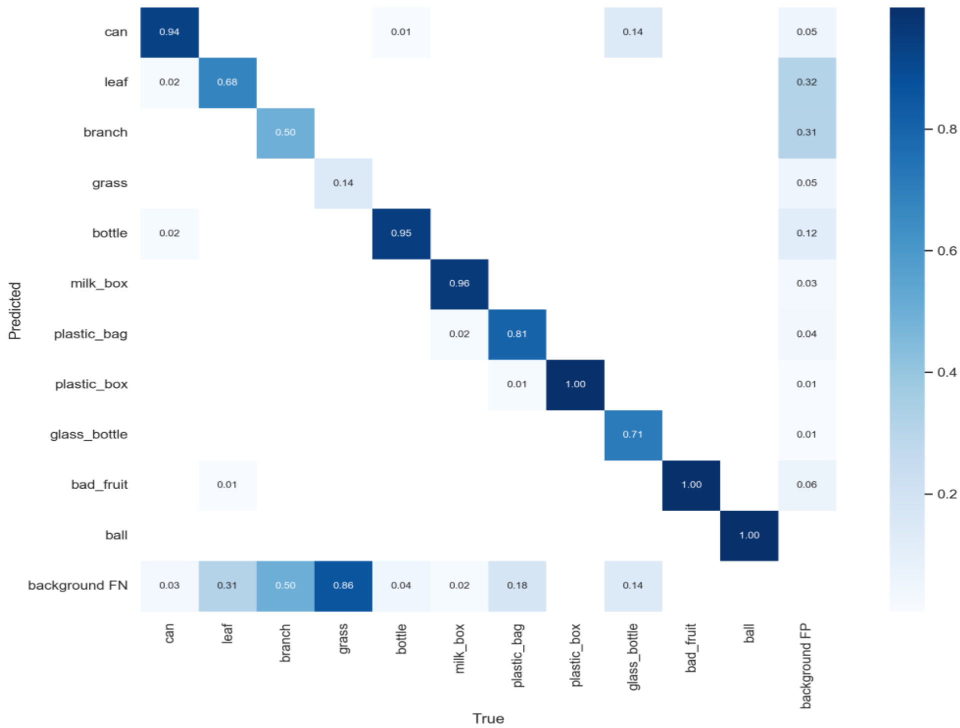

Figure 9. Confusion matrix under CIOU + SE Attention.

DOI: 10.54254/2755-2721/17/20230920

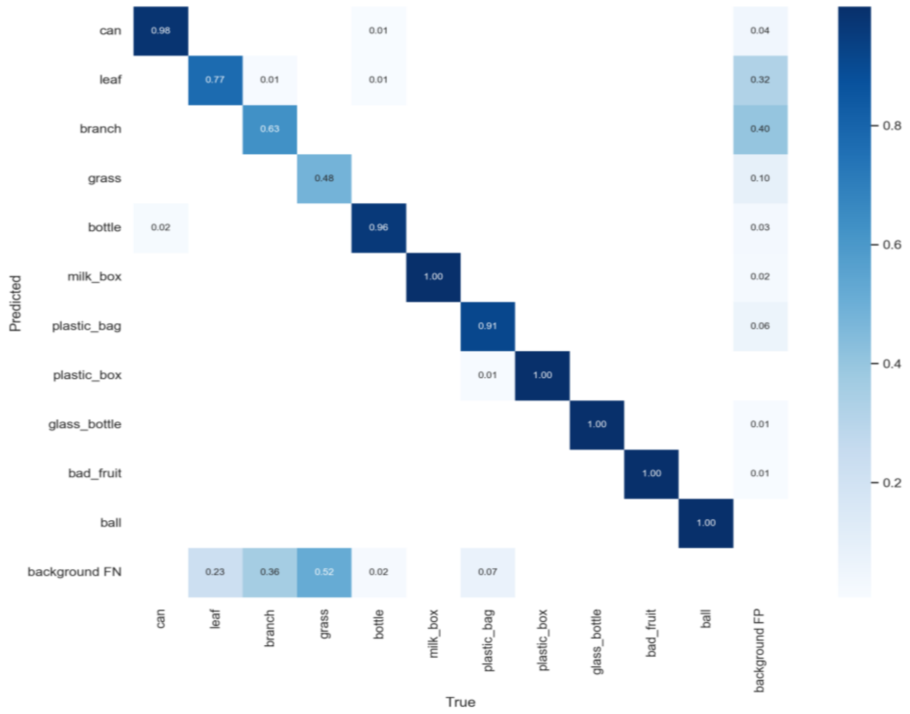

Figure 10. Confusion matrix under EIOU + SE Attention.

3.3. Analysis on the results of the running

From Figure 11, we can see that all objects in the image were identified and labelled with probability values above 80% for the majority of predictions, achieving the results of identifying regular types of litter on the water surface. For litter with a low recognition rate, the addition of the SE attention mechanism resulted in an increase of around 30% in the recognition rate.

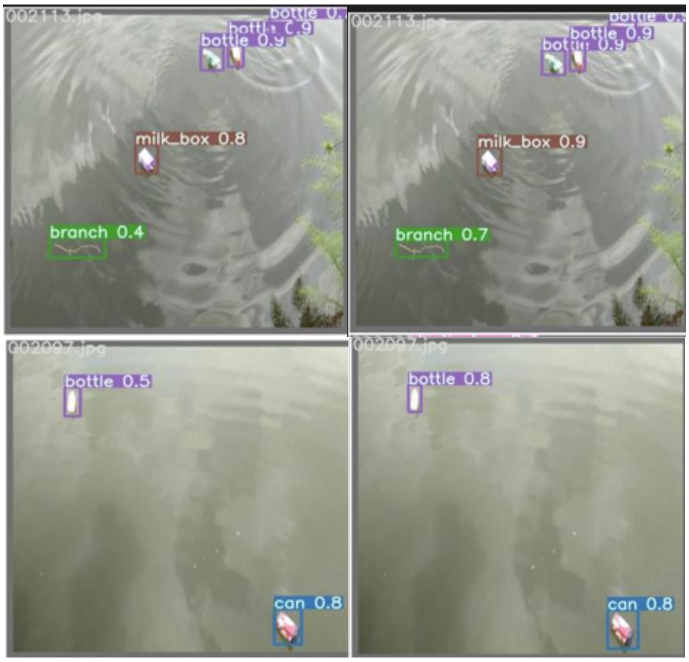

Figure 11. Presentation of running results.

DOI: 10.54254/2755-2721/17/20230920

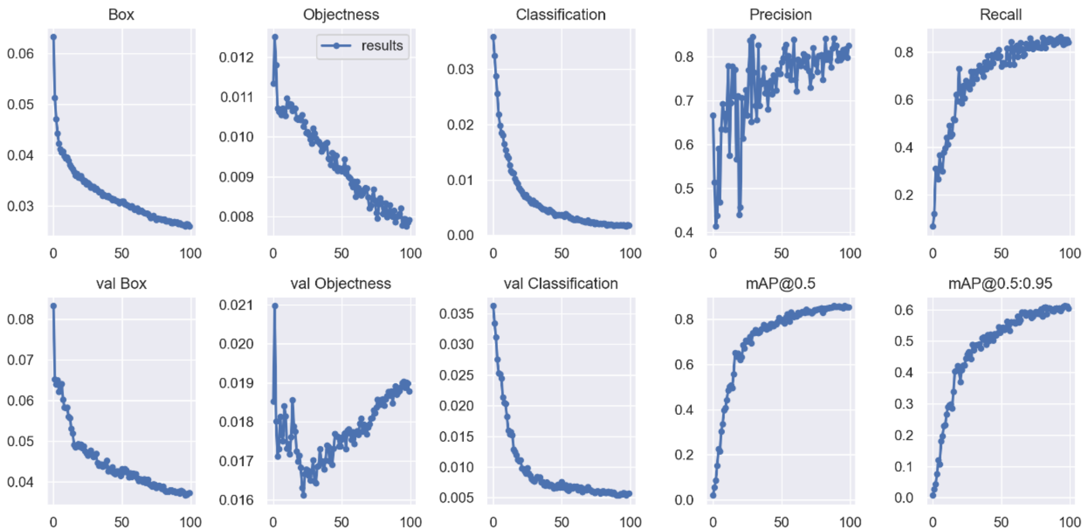

Figure 12. Performance of the model.

As can be seen from the Figure 12, the Box (GIoU) loss function has a mean value, the smaller the box the more accurate the detection. The overall trend is that the value of the function gradually decreases as the training increases, reaching a value of 0.026 at 100 training sessions. The focus of the subsequent experiments is to evaluate the algorithm's performance under various models to see if the addition of the SE attention mechanism and changing the various loss functions have improved the model's performance. Classification (presumably the mean value of the classification loss), the smaller the classification, the more accurate it is, and the overall trend is that the value of the function gradually decreases as the training increases, reaching the lowest value of 0.0017 when the number of training rounds reaches 100 in this experiment.

Precision (accuracy, positive classes found/all positive classes found), the overall trend is to increase with training, reaching a maximum value of 0.8447 at around 29 training rounds, probably due to oscillation during training.

3.4. PyQt interface application demonstration

The Figure 13 shows the PyQt interface program based on the algorithm model. With this simpler and easier to understand interface, more non-technical people can operate the water waste recognition system more easily and can actually apply it to their work activities.

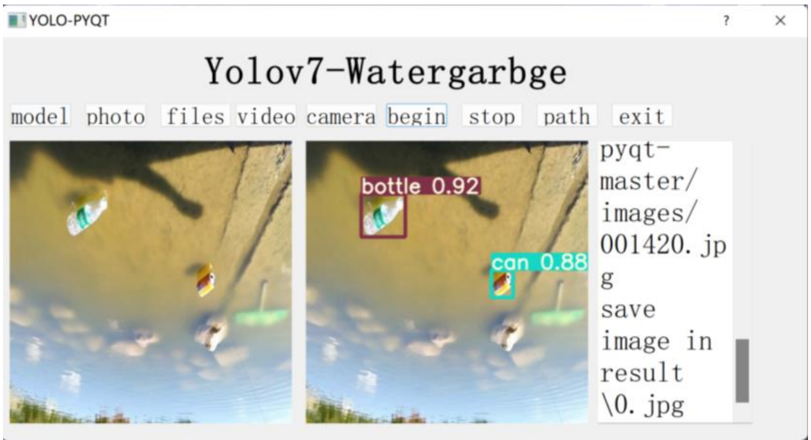

Figure 13. PyQt interface program.

DOI: 10.54254/2755-2721/17/20230920

4. Conclusion

Based on the YOLOv7 algorithm, this study provides an in-depth study of the problem of waterborne litter recognition. High accuracy waterborne litter detection was made possible by using the SE attention mechanism and the EIOU loss function in the algorithm. The experimental results show that the model's total prediction probability increases significantly with the addition of the SE attention processing mechanism, reaching 42.7%. Meanwhile, the use of the EIOU loss function improved the overall recognition probability by 11.5% compared to the CIOU loss function. This study also designed a PyQt-based interface program to enable non-technical personnel to easily operate the water-based litter recognition system.

This study has achieved some results in water-based litter recognition, but there is still room for improvement. To improve the generalisation of the model, the dataset needs to be expanded, especially for categories with low recognition rates; the algorithm needs to be optimised, such as introducing attention mechanisms and loss functions; and the algorithm needs to be applied to more scenarios, such as real-time monitoring by drones. Future prospects include the use of advanced network structures and algorithms, the combination of sensors and drones to build a perfect monitoring system, and the use of big data and cloud computing to achieve real-time monitoring, intelligent early warning and accurate management. Water surface litter identification technology has great potential in the field of environmental protection and is worthy of in-depth research and development.

References

[1] Tang Wei, Gao Han, Liu Siyang. Design and implementation of a small waterside intelligent rubbish cleaning robot system based on Raspberry Pi. 2019, Sci. Tech. Eng. **19(34)**: 239-247.

[2] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv:2207. 02696. https: / / arxiv. org / abs/ 2207. 02696.

[3] Zhang X P, Xu Z Y, Qu S,et al. Recognition algorithm of marine ship based on improved YOLOv5 deep learning. 2022, *J. Dalian Ocean Univ.*, **37(5)**:866-872.

[4] Zhang Dajie, YOLOv7 Network Architecture In-depth Analysis https: //blog. csdn. net/zqwwwm/article/details/125901507.

[5] Du Peng, Shen Ming, Su Tonghua. Deep learning and object detection. 2020, *Beijing: Electronic Industry Press*.

[6] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning.2021, *Neurocomput.* **452**: 48-62.

[7] Brauwers G, Frasincar F. A general survey on attention mechanisms in deep learning. 2021, *IEEE Trans. Knowl. Data Enginee*.,1-11.

[8] Kong Yisi, Hu Xiaofeng, Zhu Feng, Tao Jiuyang. Exploring the attention mechanism in battlefield situational awareness. 2017, *J. Sys. Simul.*, **29(10)**: 2233-2241.

[9] Zheng Z H,Wang P, Liu W, et al. Distance-IoU loss:faster and better learning for bounding box regression. 2020, *AAAI Conf. Artif. Intell.*, **34(7)**: 12993-13000.

[10] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression *arXiv: 2101.08158v1*.