Optimization and Implementation of Garbage Classification Algorithm based on YOLOv8

Xiaolong Guo, Mingge Sun, Xingke Zhang, Shilei Cui, Yinggang Li

School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China

Abstract

In response to the challenges posed by the harsh working conditions and complex manual labor involved in garbage sorting, intelligent devices are employed to replace manual labor for garbage sorting. The current study focuses on three key improvements in the field of garbage classification based on the YOLOv8 object detection algorithm-SPPCSPC, EALAN, and C2fC3Ghost. Firstly, the SPPCSPC module has been enhanced to improve the model's ability to perceive garbage at multiple scales, thereby enhancing the accuracy of garbage detection. Secondly, the improved EALAN technology has been introduced to strengthen the extraction of deep features, further optimizing the extraction of garbage features. Finally, by replace Bottleneck with C3Ghost in C2f, the improved model is able to maintain performance at a low computational costs, resulting in enhanced recognition effectiveness. Experimental results demonstrate that the improved model achieves a 3.5% increase in Precision (P), 2.0% increase in Recall (R), and 3.7% increase in mean Average Precision (mAP). The enhanced garbage classification detection model not only improves detection accuracy but also ensures high operational efficiency.

Keywords

YOLOv8; Garbage Classification; Object Detection; mAP.

1. Introduction

With rapid urbanization, the amount of waste generated is also increasing, posing a huge challenge to the environment and society. In this context, intelligent waste classification technology is particularly important, not only to improve the efficiency of waste disposal, reduce the burden on the environment, and promote the recycling of resources. In recent years, deep learning techniques have achieved remarkable success in the field of computer vision, providing new perspectives for research in the field of waste classification[1, 2].

Jinhao Fan et al. proposed a Yolo_ES target detection algorithm for the real-time and accuracy requirements in waste sorting, replacing Yolov4's backbone network by EfficientNe to realize a lightweight algorithm. And the MaxPool layer in the SPP module is replaced using SoftPool to retain more fine-grained feature information. The algorithm improves the mAP by 4.25% and compresses the model size by 75.45% with respect to the pre-improvement period, which results in better performance in terms of both accuracy and real-time performance[3]. Zeng et al. proposed a method called Multi-scale Convolutional Neural Network (MSCNN) for the problem of hyperspectral image garbage detection, which focuses on classifying HIS (Hue, Intensity, Saturation) data pixels and generating binarized garbage segmentation maps. The method shows excellent performance in large area litter detection tasks[4]. Aiming at the problem of recyclable garbage, Yuewen Zhang et al. developed an intelligent sorting system for recyclable garbage based on machine vision. The robot's localization and grasping is ensured by Zhang Zhengyou's hand-eye calibration, and YOLOv4 and

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0014

KCF algorithms are used to make the robot's grasping success rate reach 97.6% and the sorting success rate reach 96.4%[5]. Li Hongbo et al. applied the One-stage target detection algorithm to solve the plastic pollution problem, resulting in a comprehensive average accuracy of 87.3%, which provides a practical and effective means for rapid detection of plastic waste from complex environments[6].

Despite the results of the above studies, it is still more difficult to recognize spam in complex scenarios. To this end, this study aims to improve the YOLOv8 target detection algorithm by combining key optimization strategies such as SPPCSPC, EALAN, and C2fC3Ghost in order to further enhance the performance of the waste classification system.

2. YOLOv8 Algorithm

In the field of deep learning, for image classification, detection essentially belongs to a large amount of data to do linear regression. The training part of deep learning is building the algorithmic framework and finding the weight parameters for each layer in the algorithmic framework. The validation and testing part of deep learning is to run the image on the trained algorithmic framework and compare the objects in the actual image to judge whether the algorithmic framework is good or bad. The validation part is usually used in the training process, and the reverse gradient is updated according to the validation results during each training to make the model training better. The testing part is usually used to evaluate the model after the training is completed.

YOLO (You Only Look One) is one of the most advanced algorithms in the field of deep learning for doing target classification, detection, segmentation, and posing, etc. the YOLO algorithm was first proposed by Joseph Redmon et al. in 2015[7]. It transforms the target detection problem into a single neural network that is capable of simultaneously predicting multiple bounding boxes and their corresponding category probabilities in an image. With the development of deep learning, the YOLO algorithm has evolved to its eighth generation.YOLOv8 (You Only Look One-level version 8) is to be released by Ultralytics, Inc. in 2023[8].

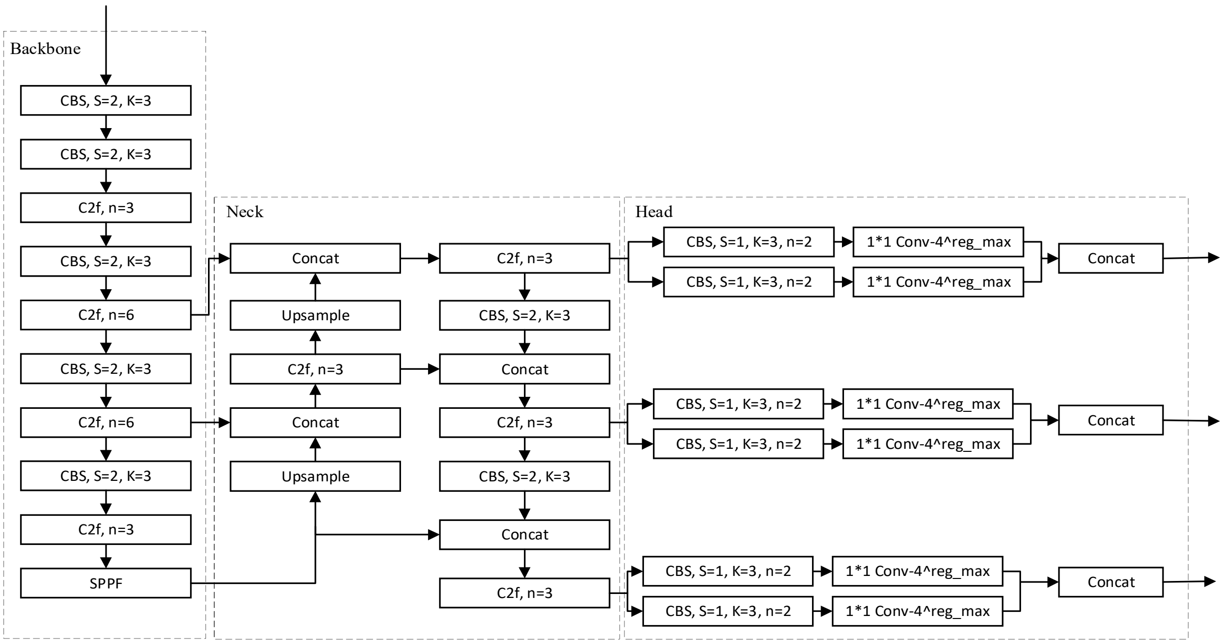

Fig. 1 YOLOv8n algorithmic framework

Currently, YOLOv8 shows great potential in waste sorting tasks with its high efficiency and accuracy. Compared to traditional waste classification methods, the YOLOv8-based detection system has more powerful target detection capabilities, and is able to recognize and classify waste items in a variety

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0014

of complex scenarios. There are five versions of YOLOv8, namely YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. As n, s, m, l, and x are incremented, the size and computation of the model gradually increase, and the speed of inference gradually slows down, but the corresponding accuracy also gradually increases. Since the algorithm is to be subsequently deployed on a computer and the detection speed is required to be fast, YOLOv8n with a smaller model is chosen.The algorithmic framework of YOLOv8n is shown in Fig. 1.The leftmost dashed box of YOLOv8 algorithmic framework is the Backbone backbone extracted feature part, the middle dashed box is the Neck feature fusion part, and the right-hand dashed box is the Head detection part[9].

YOLOv8n starts training with the two CBS modules of the backbone, CBS is the integration of convolutional, batch normalization and activation layers into a single module with a convolutional kernel K of size 3×3 and a step size S of 2. The first two layers are mainly for expanding the channels and narrowing the feature map. Then enter the core module C2f, extract the deep feature information, n is the number of repetitions for each Bottleneck module, C2f structure shown on the right side of Fig. 2, the picture enters the C2f, followed by channel segmentation, after segmentation, half of the channel information is directly to the Concat (according to the number of channels spliced) and the other half of the channel after multiple Bottleneck splicing. where Bottleneck is performing a residual splicing to splice the information in the forward propagation to the backward position. After the backbone network has performed four C2f extractions of features, feature fusion is performed using the SPPF module, as shown on the left side of Fig. 2, where the SPPF has a similar structure to the C2f, with the difference being that the Bottleneck portion is swapped out for the MP max-pooling, which in turn reduces the size of the output space.

The intermediate Neck network structure is first up-sampled (Upsample module) and the image is zoomed in using nearest neighbor interpolation to recover the details of the objects in the image. Then the feature extraction is continued using C2f and the cross-scale fusion connection is performed. The three output sections on the right side are Head decoupling heads, which separate classification and detection, and use a form of cross-entropy-like to optimize the left and right position probabilities closest to the label, thus allowing the network to quickly focus on the distribution in the region adjacent to the target position. When YOLOv8n carries out the detection, the picture goes through the trunk feature extraction, Neck's feature fusion, and in the HEAD decoupling head part will carry the classification information and the location information, and finally the three decoupling heads predict the large, medium, small, and large targets in the picture respectively.

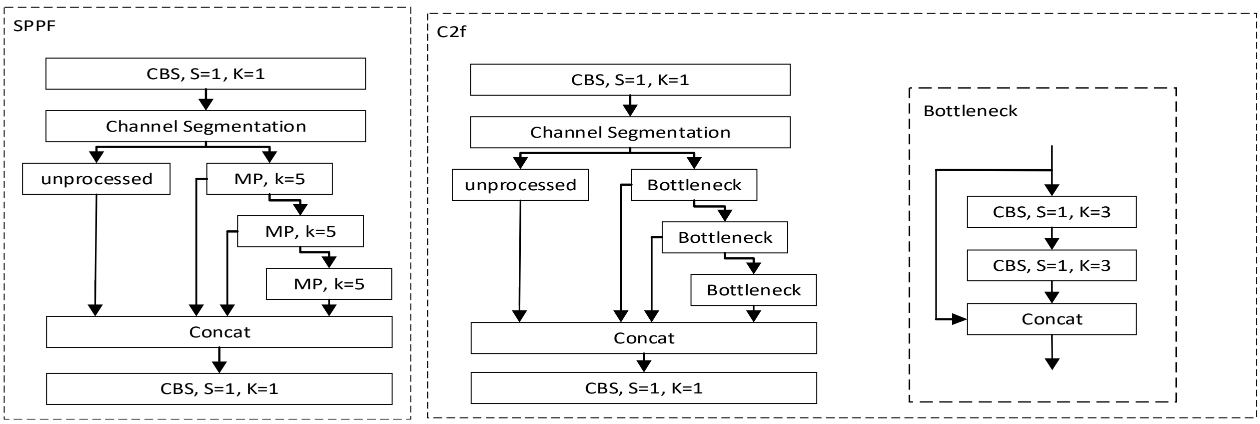

Fig. 2 Structure of SPPF and C2F modules

3. Improved YOLOv8n Algorithm

By delving into three key optimizations of the YOLOv8n algorithm: the SPPCSPC, EALAN and C2fC3Ghost modules. It is expected to maintain the model performance at low computational cost and achieve more superior waste classification recognition. The structure of the improved network is shown in Fig. 3, where the dashed box shows the position of the improvement, and the backbone is

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0014

modified to be a strengthened extraction network of ELAN and SPPCSPC, and two of them, C2f, are replaced by C2fC3Ghost. this makes the model in the parameter count rise slightly while the feature information is strengthened to be extracted.

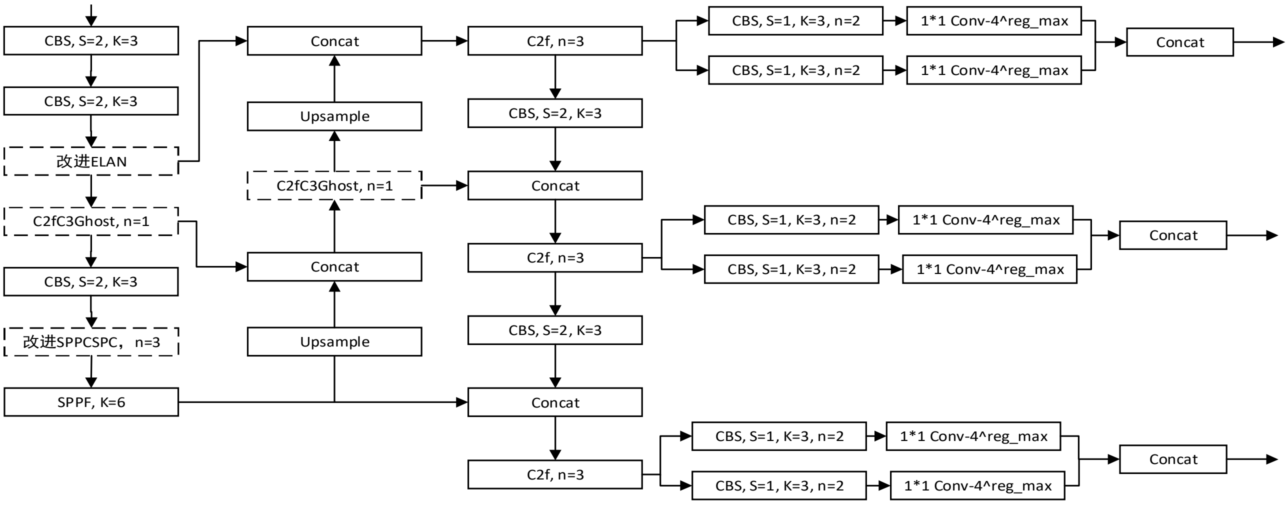

Fig. 3 Structure of the YOLOv8n improved network

3.1 SPPCSPC Improvements

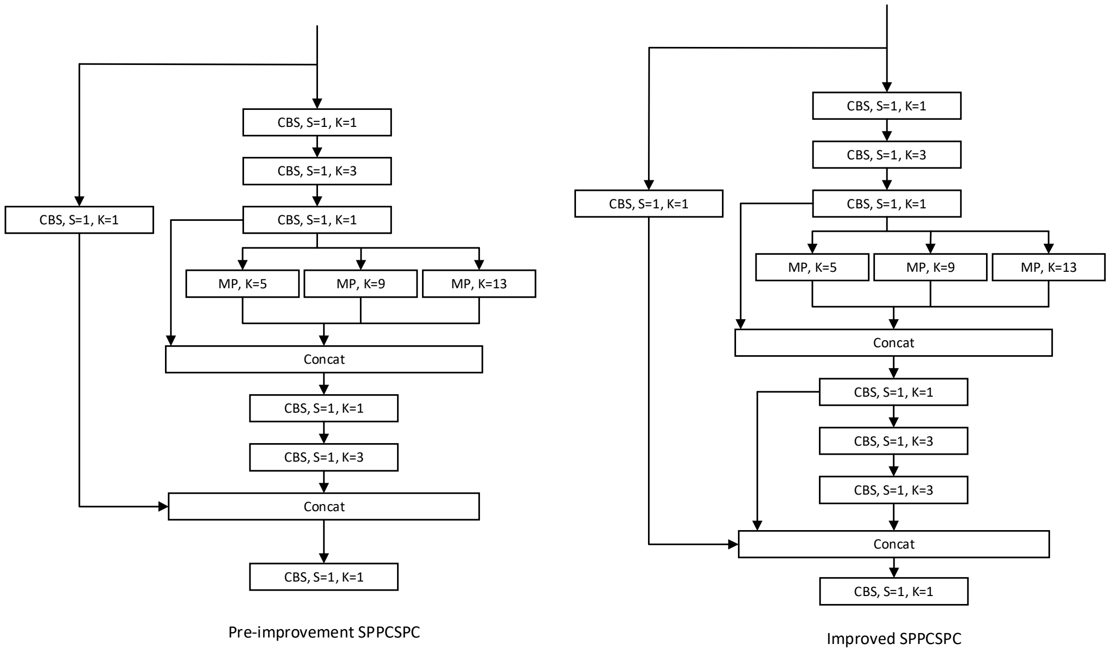

Fig. 4 Comparison before and after improvement of SPPCSPC

The YOLOv8n trunk section performs a large number of C2f extraction features, and there is only one SPPF in the trunk for feature fusion, so the last C2f in the trunk, is replaced with an improved SPPCSPC module for enhanced feature fusion. SPPCSPC is a feature fusion module in YOLOv7[10], as shown on the left side of Fig. 4, the image information enters into the network structure of SPPCSPC, and three CBS modules (convolutional layer, batch normalization, and activation layer) are performed first, with convolution kernel sizes of 1 × 1, 3 × 3, and 1 × 1, and the step size of all of them is one. The feature maps are then reduced in size and dimensionality by three different MP max-pooling, the convolution kernels of the three max-pooling are 5×5, 9×9 and 13×13 respectively.Next the feature maps of the three max-pooling and pooling upper layers are spliced by channel using Concat. After the splicing is completed, the feature map is then subjected to two CBS modules to extract features, and again the feature map that just entered the SPPCSPC is subjected to CBS to extract features and then spliced according to the channel to make a structure similar to the residuals, and finally the features are extracted in the CBS and the output channel of the SPPCSPC is adjusted.

The improved SPPCSPC module is shown on the right side of Fig. 4. The SPPCSPC module has fewer features extracted by the CBS convolution after the three maximum pooling, so a CBS module is added with a convolution kernel of 3 × 3 and a step size of 1, and the information from the first CBS is spliced to the end to produce a residual structure, which makes the model more conducive to extracting feature information.

3.2 EALAN Improvements

The EALAN module is a multi-branch stacking module in YOLOv7, as shown in the dashed box in Fig. 5, where the features extracted from each layer of CBS are spliced together at the end, and then a single CBS adjustment channel is performed for output. Multi-branch stacking module was heavily used in YOLOv7, which extracted features better, but the number of parameters in the whole network was greatly increased.The EALAN module was removed in YOLOv8, and since the backbone of YOLOv8 extracted features weakly, the EALAN module was improved to replace a part of the backbone in YOLOv8. The improved EALAN is shown in Fig. 5.

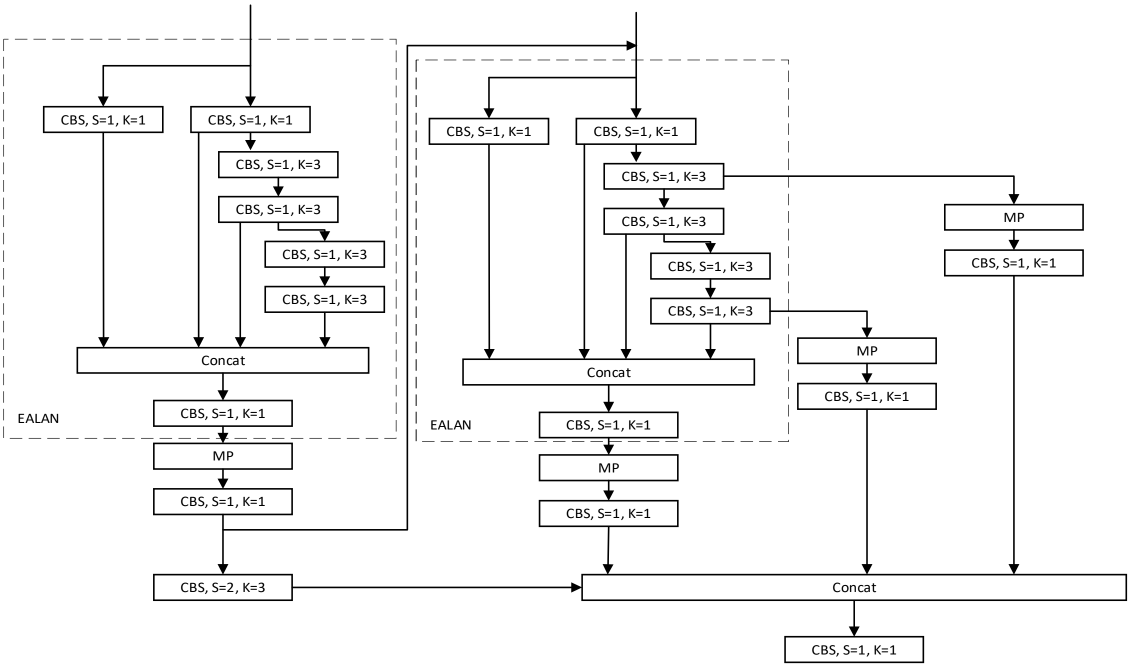

Fig. 5 Structure of the improved EALAN module

Multiple use of the EALAN module leads to an increase in the number of parameters and an increase in the complexity of the model and the number of channels. Therefore, two EALAN modules are taken, the first EALAN module is maximally pooled once at the end to reduce the feature dimensions, and two positions are elicited in the second EALAN module for maximally pooling, and the key information in the maximally pooled is concatenated at the end according to the channels.

3.3 C2fC3Ghost Module

The C2fC3Ghost module is to replace Bottleneck in C2f with C3Ghost. The structure of C2fC3Ghost is shown in Fig. 6, in which the left side of the figure shows the structure of C2fC3Ghost and the right side of the figure shows the structure of C3Ghost.

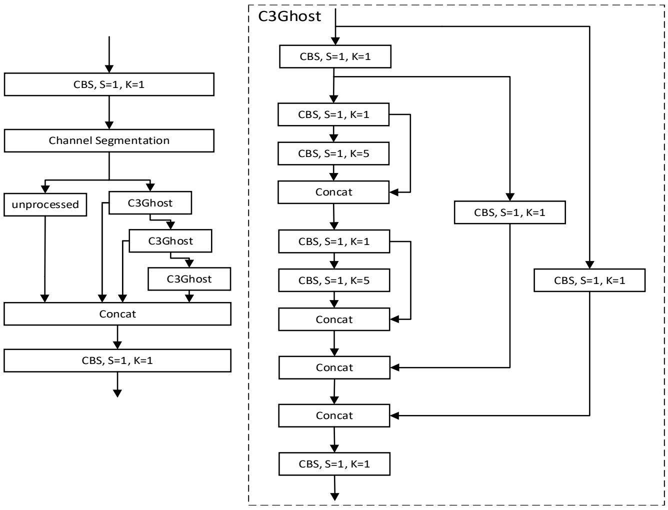

Fig. 6 Structure of C2fC3Ghost

GhostNet was proposed by Huawei Noah's Ark in 2020[11], which is mainly aimed at reducing the number of parameters of the model without loss of accuracy, in which the C3Ghost structure is to add GhostNet on the basis of the C3 module, and replace the internal structure of C3 with a small convolutional kernel and a large convolutional kernel for feature extraction. Similarly, the Bottleneck in C2f is replaced by C3Ghost, which not only ensures the residual structure of the original algorithm, but also slightly reduces the number of parameters in the model.

4. Experiment and Result Analysis

4.1 Experimental Platform

The operating system for the experiment is Windows 11 Professional, the processor is 11th Gen Intel(R) Core(TM) i5-11400H @ 2.70GHz 2.69 GHz, the running memory is 16GB, the GPU model is RTX 3060 with 6GB of video memory size, and the experiment is on Pytorch 1.8.2 deep learning framework, Cuda 11.3 architecture and the Python version is 3.9. Parameters for training: batch_size is set to 16, epochs is set to 300, initial and final learning rates are both 0.01, the size of the input image is automatically scaled to 416×416, no pre-training weights are used, and the other parameters are default values.

4.2 Experimental Data Set

Fig. 7 Partial image of the spam dataset

Before training the model in YOLO, the corresponding dataset images are needed, and since the detected targets will be in different positions on each image, the YOLO algorithm needs to label each class of targets in the images first when training. Therefore, a 6-classified waste dataset from Roboflow open source website was selected, which includes biodegradable waste, cardboard, glass,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0014

metal, paper and plastic in the dataset. The total number of images is 8745. Roboflow, which was created by a California-based company, provides a dataset that has all of the training targets labeled. However, the quality of the labeling on that site also varies, and the dataset needs to be preprocessed to remove some of the mislabeled or duplicate images. After preprocessing 7057 images are retained, as shown in Fig. 7 for some of the garbage images, and then the dataset is randomly divided according to 8:1:1, and finally 5645 images are training set images, 706 images are validation set images, and 706 images are test set images.

4.3 Evaluation Indicators

The number of parameters, arithmetic power, precision (precision, P), recall (recall,R) and mean average precision (mAP) are selected as the evaluation metrics of the model. Where mAP is calculated from precision P and recall,R. $P = \frac{TP}{TP + FP}$, $R = \frac{TP}{TP + FN}$, $AP = \int_0^1 PdR$, $mAP = \frac{1}{n}\sum_{i}^{n} AP_i$.

Where TP is the number of positive samples being judged correctly, FP indicates the number of misdetected samples, FN is the number of missed samples, AP is the area of the curve composed of accuracy P and R with respect to the axes; mAP is the average of all APs, and i in mAP indicates the current category. When mAP is higher, it means the model is trained better.

4.4 Ablation Experiments

Both the original YOLOv8n and the improved YOLOv8n model keep the same hyperparameters during the training process, and the mAP0.5 curve shown in Fig. 8 is plotted according to the log files during the training process, and the trained weight files are put into the test set for testing, and the mAP0.5 of the test set and the training process are compared, and the mAP0.5 of the pre-improved YOLOv8 network is not much different from the mAP0.5 of the training process, although the test set is higher than the the training process, the values are slightly higher, but basically the difference is not significant.

In terms of the number of parameters and arithmetic power the values are as small as possible, and although the improved model goes up slightly, the impact will not be too great for the average computer hardware when the model is deployed. The main thing is to identify all kinds of garbage accurately.

After the improvement, the trend of the improved curve in the training set is faster than the trend before the improvement, as shown in Figure 7 (mAP of the model in training). And the mAP0.5 in the test set increased by 3.7% as shown in Table 1 (results of the model in the test set) ablation experiment. Thus it can be seen that the improved model is higher than the original model in terms of convergence speed and localization accuracy.

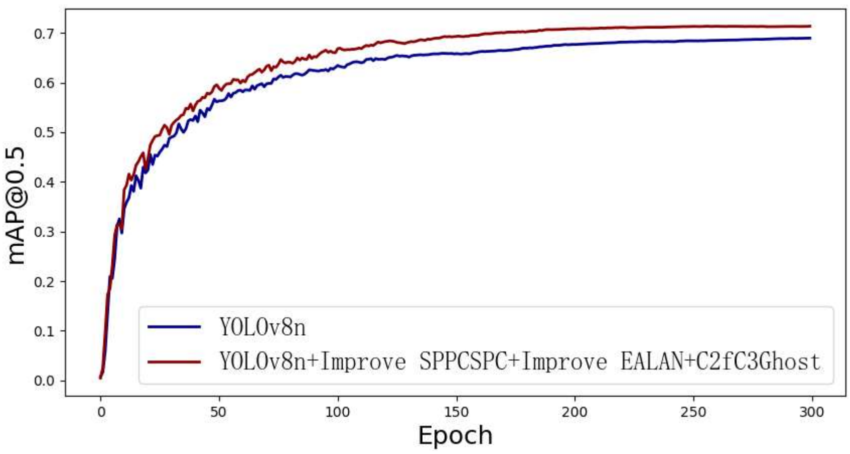

Fig. 8 mAP0.5 curves of original YOLOv8n algorithm and improved YOLOv8n algorithm

Table 1. Ablation experiments

algorithmic model	quantity of participants /M	arithmetic power /GFLOPs	P/%	R/%	mAP50/%
YOLOv8n	3.0	8.1	77.3	64.5	72.2
YOLOv8n+ Improvement of SPPCSPC	5.0	9.7	79.5	64.9	73.5
YOLOv8n + Improved SPPCSPC + Improved EALAN	5.2	14.4	81.9	66.0	75.1
YOLOv8n+ Improved SPPCSPC+ Improved EALAN+ C2fC3Ghost	5.0	13.7	80.8	66.5	75.9

4.5 Comparative Experiments

In order to further explore the advancement of the algorithms, the improved algorithm is compared with other algorithms, keeping the dataset and hyperparameters unchanged, and then comparing with some current mainstream algorithms, such as SSD, YOLOv7, and adding the unimproved SPPCSPC, EALAN, and C3Ghost to the same position of YOLOv8n for comparison. The number of parameters, arithmetic power, P, R, and mAP are still taken as evaluation indexes, and the comparison results are shown in Table 2.

From Table 2, it can be seen that the improved algorithm compared to YOLOv8n+SPPCSPC+ EALAN+ C3Ghost, YOLOv7, and SSD has an improvement in precision P, recall R, and mAP, and there is a slight rise in the number of parameters and arithmetic, but it does not affect the speed of model inference too much. Thus the algorithm proposed in this study performs better and has higher accuracy for waste classification detection.

Table 2. Experimental comparison of different algorithms

algorithmic model	quantity of participants/M	arithmetic power /GFLOPs	P/%	R/%	mAP50/%
SSD	24.4	30.7	63.5	43.2	57.5
YOLOv7-tiny	6.0	13.2	80.2	62.9	71.0
YOLOv7	36.5	103.2	78.4	64.9	73.1
YOLOv8n+SPPCSPC+EALAN+ C3Ghost	4.5	12.8	78.2	65.2	73.7
YOLOv8n+Improve SPPCSPC+Improve EALAN+C2fC3Ghost	5.0	13.7	80.8	66.5	75.9

5. Conclusion

Based on the YOLOv8 network, the garbage recognition algorithm is improved so that the garbage recognition achieves a better effect. Firstly, the feeling field in YOLOv8n algorithm is made larger by improving SPPCSPC, which strengthens the effect of recognizing large, medium and small garbage; secondly, the EALAN module is improved so that the model backbone has stronger ability to extract target features; finally, Bottleneck in C2f is replaced by C3Ghost, which makes the model

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0014

to improve the classification precision while the number of parameters does not rise. The improved model achieves significant improvement in precision, recall and mean average precision.

References

[1] Qi Liu, Gang Yin, Ying Wang, et al. Deep learning-based algorithm design for water surface floating object recognition[J]. Journal of Jilin Institute of Chemical Technology, 2022,39(7):28-33.

[2] Siqian Yu, Qirong Zhao, Jiachen Lin, et al. Detection of walnut shell defects based on deep learning[J]. Journal of Jilin Institute of Chemical Technology, 2022,39(9):80-85.

[3] Jinhao Fan, Lizhi Cui.A target detection model for garbage classification based on Yolo_ES[J]. Electronic Measurement Technology,2023,46(1): 160-166.

[4] ZENG D, ZHANG S, CHEN F, et al. Multi-scale CNN based garbage detection of airborne hyperspectral data[J]. IEEE Access,2019, 7: 104514-104527.

[5] Yuewen Zhang, Songheng Li, Wei Zhang, et al. Design of intelligent sorting system for recyclable waste based on machine vision[J]. Laboratory Research and Exploration,2022,41(7):98-103,107.

[6] Hongbo Li, Xianggang Liao, Li Chen. Automatic plastic identification system based on machine learning One-stage target detection algorithm[J]. Plastics Science and Technology,2020,48(12):86-89.

[7] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. IEEE Trans. Pattrern Anal, 2018, 15: 2236-1131.

[8] Ruijin Liu, Zhangming He. A fast target detection method for satellite remote sensing images based on YOLOv8[J]. Space Control Technology and Application,2023,49(5):89-97.

[9] Qiang Han. Research on improved YOLOv8 algorithm for small target detection [D]. Jilin:Jilin University, 2023.

[10] Xiangming Qi, Rui Chai, Yimeng Gao. Reconstructed SPPCSPC with optimized downsampling for small target detection algorithm[J]. Computer Engineering and Applications,2023,59(20):158-166.

[11] Junyi Huangfu, Qiao Meng, Lingchen Meng, et al. YOLOv5 traffic target detection based on GhostNet and attention mechanism[J]. Computer System Applications,2023,32(4):149-160.