Optimization and Evaluation of YOLO Model on Small Dataset Based on Fine Tuning

Zhengqun Sun *

School of Software Technology, Dalian University of Technology, Dalian, 116620, China

* Corresponding Author Email: 1569901499@mail.dlut.edu.cn

Abstract. In the current era of rapid development of artificial intelligence, object detection is one of the core technologies of computer vision. The degree of accuracy is an important criterion to judge the quality of models. Aimed to improve the degree of accuracy of You Only Look Once (YOLO) models, this experiment adopted the transfer learning strategy, through backpropagation and gradient descent, fine-tuning the model parameters under the condition of a small learning rate. This experiment trained data on a small dataset to enhance the efficiency of learning. The result showed that the degree of accuracy of each model has slightly improved. For example, the mAP50 of the YOLOv8l model has been raised from 0.8625 to 0.8798, which is a 2% rise. Although the increase was relatively small, the result is in line with expectations, which verifies the feasibility of this method, and this is a highly generalized method to improve the accuracy of the model's object detection.

Keywords: Computer vision, Target detection, Artificial intelligence.

1. Introduction

Artificial intelligence is a topic widely discussed in the academic, business and other sectors of society today. Among them, computer vision is one of the most widely applied areas. Object detection is one of the core technologies of computer vision. Accuracy, robustness and other indicators have always been the key metrics for evaluating models. In order to improve the accuracy of models, many methods have been developed.

Researchers from both domestic and international sources have been committed to improving these key indicators of models. For example, Yan et al., aiming to address the issues of high cost of the detector and the inability to handle local weighting, introduced GLIF, a novel Global-Local Interaction Fusion network, which is good at distinguishing in a complex background. It improved the accuracy and the mAP increased by 2.7% to 11.3%, respectively [1]. Moreover, Abenezer Girma et al. proposed an adaptive point network (ADP-Net) in order to enable small-scale features to be accurately captured by the model. This framework can dynamically adjust the receptive field through an adaptive attention mechanism, significantly improving the accuracy and robustness of object detection for small objects. mAP has been improved by 28.93% compared to the YOLO model with the best performance [2]. There is another study aimed at solving the problem of target detection for pedestrians in obstructed conditions, proposing a feature dynamic weighting model for occluded person re-identification (FDWM), which improved the accuracy of pedestrians under different occlusions. On the Occluded-Duck dataset, the mAP is increased by 2.8% [3]. It can be seen that there have been numerous studies on improving the accuracy of target detection and recognition in various specific situations and classes.

However, these methods mentioned above lack generalization ability. They can't be applied to object detection in other classes. But this study used a small dataset as an example, through fine-tuning, enabling the YOLOv8l and a series of YOLOv8 models to learn the features on any dataset. As a result, this is a highly generalized method for improving the accuracy of models.

2. Method

2.1. Model Introduction

In 2023, the Ultralytics company released the YOLOv8 series of models [4]. This experiment included some models of this series such as YOLOv8n, YOLOv8s, YOLOv8m and YOLOv8x. Through increasing the depth and width of the network, they gradually shifted from emphasizing speed to emphasizing accuracy, with the increasing of the computational load and parameter quantity. And compared to YOLOv5 models, this series of models' accuracy has improved significantly on average [5]. Taking the most lightweight and fastest YOLOv8n model as an example, its structure is designed as "Backbone – Neck – Head" [6]. Backbone network is used to extract image features, and the C2f module within it can be optimized [7], and the neck is used for fusion enhancement, and the head can predict the target detection bounding box.

2.2. Fine Tuning

This experiment conducted multiple sets of control experiments. Through transfer learning, firstly the accuracy of some original pre-trained models such as YOLOv8n, YOLOv8m, is evaluated by assessing their mAP50 and other metrics. Then, keeping the network architecture of the model unchanged, this paper utilized its initial parameters and fine-tuned parameters with a small learning rate, through backpropagation and gradient descent on the training set. After fine-tuning, this experiment calculated various indicators, such as mAP50 of the model for evaluating the accuracy of the model, and calculated the improvement percentage of these indicators of the model. At last, these indicators can assess the effect that fine-tuning has on the model.

3. Experiment

3.1. Dataset

In order to test the accuracy of the model, this experiment used the coco128 dataset, including 128 images, 80 common classes with wide coverage [8]. Among them, there are classes such as person, car, cup, bowl and so on, whose resolutions are mostly in the range of 640*420 to 640*480. Since the original coco128 dataset has not been divided to training set and validation set – there is only a training set, manual division was carried out, and the ratio of training set to validation set is approximately 10:3. Training and validating on a small dataset improved the experimental efficiency. Moreover, the classes of this dataset are numerous, which can eliminate possible errors specific class causing.

3.2. Setup of Experimental Environment

This experiment was based on the computer equipped with NVIDIA RTX 4060 graphics card as the hardware environment. And the software environment was based on Python 3.8 programming language and PyTorch 2.0 deep learning framework to set up YOLOv8 running environment. To ensure the reproducibility of the experiment, the following steps to configure the environment should be strictly followed. Firstly, the official source code of YOLOv8 can be downloaded from GitHub repository, and the Ultralytics library and its associated dependent packages can be installed through pip tool. Then this experiment configured CUDA 11.7 and cuDNN 8.4 to enable GPU acceleration, and ultimately verified the "torch.cuda.is_available()" returned "Ture" to make sure that the model utilized GPU computing power during the reasoning process.

3.3. Evaluation Methodology

This experiment tested the target detection accuracy of models through calculating mAP50, mAP50-95, precision (P) of the model [9], and calculated the improvement percentage of these indicators after fine-tuning to assess the ability of fine-tuning to improve the accuracy of the model.

$$mAP50 = \frac{\sum AP@0.5}{N} \quad (1)$$

$$P = \frac{TP}{TP+FP} \quad (2)$$

In the above formula, N is the total number of datasets. TP represents the positive class that is predicted to be true. FP represents the false positive class that is predicted as positive.

4. Results

4.1. Model Validation

After setting up the experimental environment, the picture taken in the dormitory was uploaded. As shown in Fig. 1, there are a large number of items with various classes. After model testing, as shown in Fig. 2, it can be seen that the model identified some classes like book, bottle and so on, but there are still so many items that had not been identified. Since the YOLOv8 model was mainly trained using the COCO dataset, and the COCO dataset doesn't include classes such as pen, headset, rubber and so on, incorrect detection or failure to identify happened.

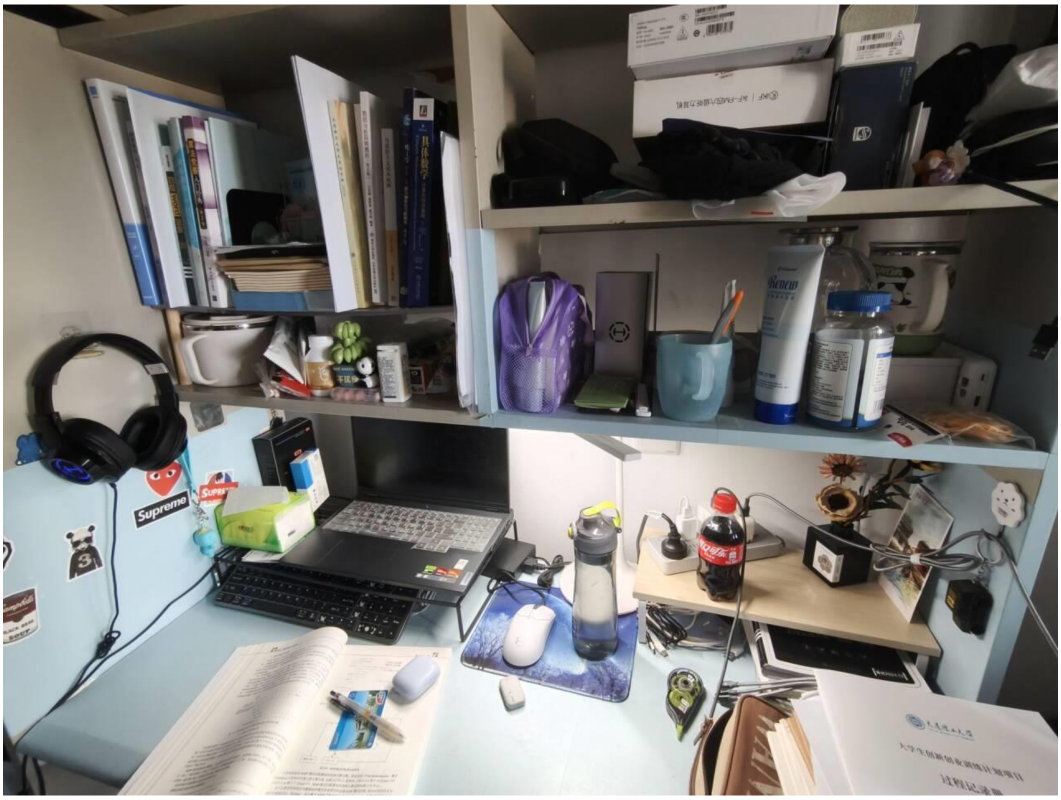

Fig. 1 Dormitory verification set

(Picture credit: Original)

Fig. 2 Testing result of dormitory

(Picture credit: Original)

In the process of calculating the accuracy, the model identified the targets in the COCO128 dataset. From the comparison of Fig. 3 and Fig. 4, it can be concluded that the model can identify targets and calculate accuracy through the COCO128 dataset.

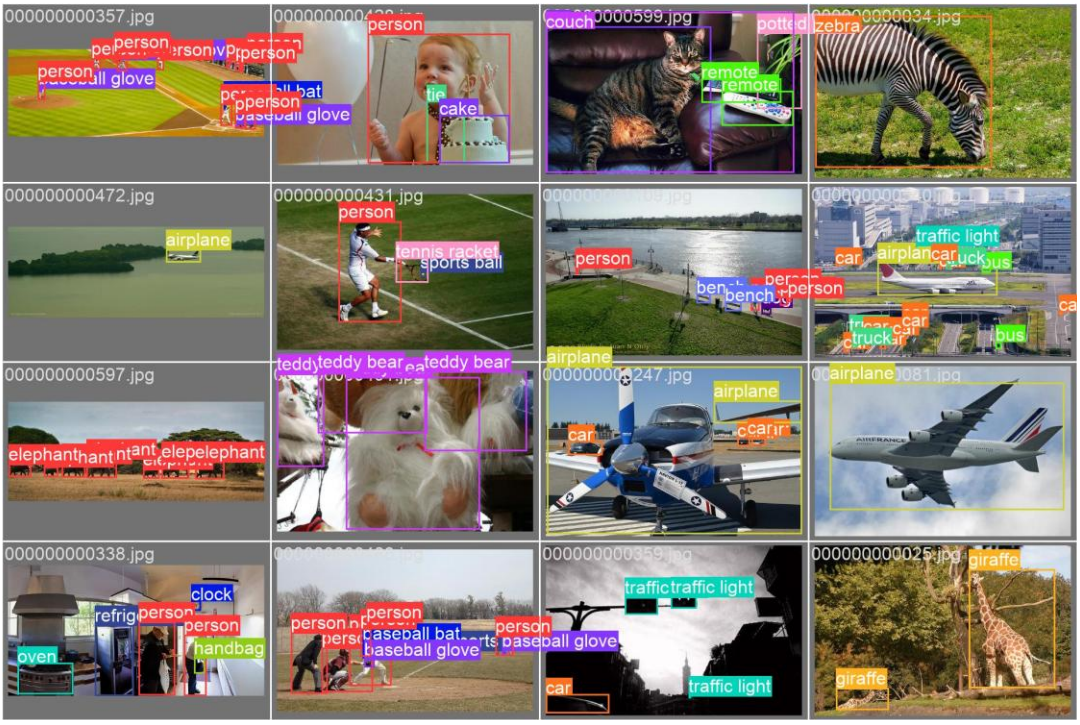

Fig. 3 Validation set

(Data from: COCO128 dataset)

Fig. 4 Testing result

(Data from: COCO128 dataset)

4.2. Model Accuracy Results

As seen in Fig. 5, the gray curves represent the PR curves of each class the model identified on the validation set, and the blue curve represents the PR curve of all classes [10]. The precision of the original YOLOv8l model decreases as the recall increases. And as shown in Fig. 6, after fine-tuning, the blue PR curve of the model moves slightly upwards, and the precision is higher, which demonstrates the effect of fine-tuning clearly.

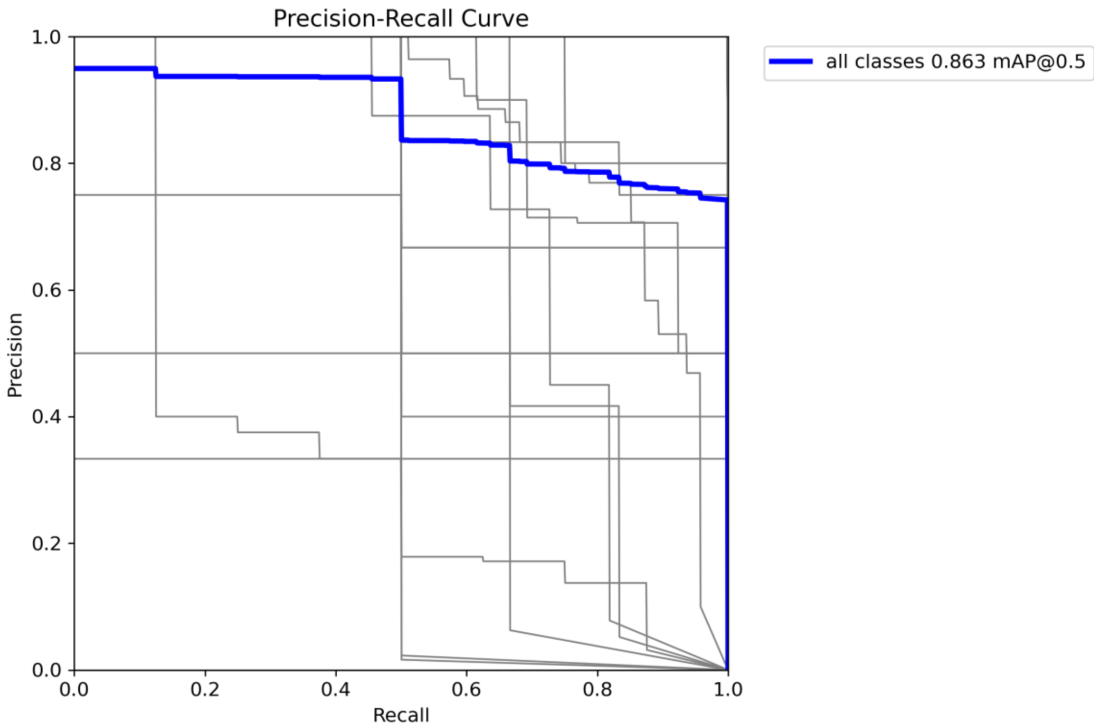

Fig. 5 Original YOLOv8l model PR_curve

(Data from: COCO128 dataset)

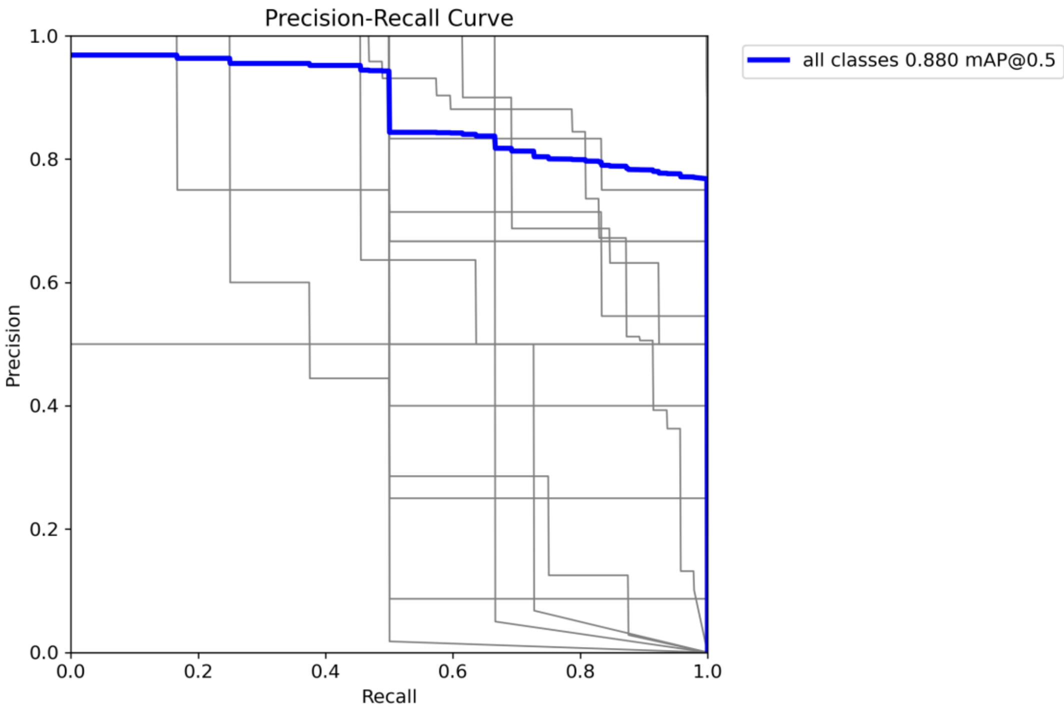

Fig. 6 Fine-tuned YOLOv8l model PR_curve

(Data from: COCO128 dataset)

As shown in Table. 1, the mAP50 of the original YOLOv8l model is 0.8625, mAP50-95 is 0.7153. After fine-tuning training on only small datasets, the mAP50 increases by 2%, and the mAP50-95 increases by 1.9%, and the precision increases by 2.6%.

Table 1. Assessment of the improvement of the YOLOv8l model's accuracy

Assessment Indicators	The Original YOLOv8l Model	The Fine-tuned YOLOv8l Model	Improvement Magnitude
mAP50	0.8625	0.8798	2.0%
mAP50-95	0.7153	0.7290	1.9%
Precision	0.6696	0.6870	2.6%

To demonstrate the generality of the fine-tuning, as seen in Table. 2, fine-tuning method led to a small improvement in accuracy on most models.

Table 2. The assessment of mAP50

	YOLOv8n	YOLOv8s	YOLOv8m	YOLOv8x
Original mAP50	0.7035	0.8309	0.8875	0.8847
Fine-tuned mAP50	0.7071	0.8418	0.8931	0.8843
Improvement Magnitude	0.5%	1.3%	0.6%	0.0%

5. Discussion

This study can be applied to areas regarding targets detection such as intelligent monitoring, industrial quality inspection and so on. In these areas, this method can improve the accuracy when the model identifies the target, which can help to improve safety performance and quality inspection accuracy without a large dataset.

Moreover, through fine-tuning, the model can learn features on only small dataset, which suggests that this method can enable the model learn more features on large dataset and the improvement is even greater.

However, fine-tuning is not a method targeting certain categories or specific scenarios after all. Additionally, there are limitations to the small dataset, resulting in a limited increase range. In future studies, a larger dataset with more classes can be used for fine-tuning, and research on improvement of a specific class can be continued.

6. Conclusion

In order to make the computer vision technology applied more effectively in reality, improving the accuracy of models' target detection is an important study. By fine-tuning the parameters of the YOLOv8l model and other models on small dataset COCO128 dataset, this study slightly improved indicators related to the accuracy of target detection like mAP50, mAP50-95, precision and so on. Among them, on YOLOv8l, the mAP50 increased by 2.0%, and the precision increased by 2.6%. The experiment verified that transfer learning is also effective in the case of a small sample size, offering a feasible program for model optimization in resource-constrained environments. Though the increase was relatively small, the result obtained of the efficient experiment on a small dataset is reasonable, which verified that parameter fine-tuning plays a role in improving the accuracy of the model.

References

[1] Yan X, Zhang Y, Chen H, et al. GLIF Global–Local interaction fusion based UAV small target detection domain shift suppression network. Infrared Physics and Technology, 2026, 152: 106206.

[2] Girma A, Homaifar A, Mahmoud M N. ADP-Net: Adaptive point network with multi-scale attention mechanism for small object detection. Neurocomputing, 2026, 659: 131381.

[3] Wang Y. Research on Object Re-identification and Tracking in Multi-camera Scenarios. Hangzhou: Hangzhou Dianzi University, 2025.

[4] Varghese R, S. M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In: 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai, India, 2024: 1-6.

[5] Mao S, Wang W. A Review of YOLO Series Object Detection Algorithms Based on Deep Learning. Journal of Yan'an University (Natural Science Edition), 2024, 43(02): 88-95.

[6] Zhou H, Jin S, Zhou L, et al. Classification and recognition of camellia oleifera fruit in the field based on transfer learning and YOLOv8n. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(20): 159-166.

[7] Wang C, Guo X, Ma C, et al. A Flat Peach Bagged Fruits Recognition Approach Based on an Improved YOLOv8n Convolutional Neural Network. Horticulturae, 2025, 11(11): 1394.

[8] Kim G, Jung H, Lee S. Spatial reasoning for few-shot object detection. Pattern Recognition, 2021, 120: 108118.

[9] Xie W, Bai X, Liu M, et al. Transformer-based multi-scale feature fusion for real-time CT bone metastasis detection. Bone, 2026, 203: 117729.

[10] Alruwaili M, Siddiqi M H, Atta M N, et al. Deep learning and ubiquitous systems for disabled people detection using YOLO models. Computers in Human Behavior, 2024, 154: 108150.