

# Research on Improved Method based on YOLOV5s Target Detection Model

Xiuhuan Dong \*, Shixin Li, Liming Zhou

College of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

\* Corresponding author: Xiuhuan Dong (Email: 0422211032@tute.edu.cn)

**Abstract:** Aiming at the problem of low detection accuracy of small targets, an object detection method based on average pooling improved YOLOV5s model is proposed. The algorithm introduces the Squeeze Excitation attention module and the Efficient Intersection Over Union loss function to comprehensively improve the detection calculation efficiency and accurate deployment ability. With the development of deep learning technology, which is of great significance to improve the detection accuracy and detection rate. YOLO greatly improves detection performance, three times faster than retinanet and 2 times faster than faster-rcnn. YOLO has strong generalization ability, can be applied to different application scenarios, and is also easy to deploy. The steel surface defect public dataset was selected for verification. The results show that the improved YOLOV5s model is better than the original YOLOV5s model, the test average accuracy mAP can reach 81.8%, the average accuracy mAP of the model is increased by 7.4%, and the overall performance of the improved model is better than other conventional models.

**Keywords:** Small Targets; Squeeze Excitation Attention Module; Defect Detection.

## 1. Introduction

One-stage object detection algorithms, mainly YOLO series, SSD, etc. Its core goal is to return the category and location of the target through a network. YOLO (You Only Look Once, YOLO) series algorithms have good comprehensive performance [1] [2]. Yolo uses a convolutional network to extract features, and then uses a fully connected layer to obtain predicted values. The network structure refers to the GooLeNet model and contains 24 convolutional layers and 2 fully connected layers [3]. The output vector of YOLO includes not only the category of the target. Also include the coordinates of the bounding box and the confidence level of the prediction. Then, locally unique prediction boxes are obtained by non-maximum suppression. In the field of deep learning-based object detection, rectangles are used to label their position and size [4]. At the same time, convolutional neural networks are used to build machine learning models, which are driven by a large amount of labeled data. On top of the 20 convolutional layers obtained by pre-training, 4 convolutional layers and 2 fully connected layers are randomly initialized [5].

On the basis of the original YOLO backbone model, on the one hand, the SE attention mechanism is introduced [10]. The SE module was originally proposed by Jie Hu et al. in 2017 with up to 20,000 citations, aiming to improve the efficiency of channel-to-channel information transmission in convolutional neural networks (CNNs). SE (Squeeze and Excitation, SE) attention mechanism learns an adaptive channel weight model [11]. It focuses on more useful channel information, with high detection efficiency and optimal ablation. On the other hand [12], EIOU is introduced, and the aspect ratio is replaced by the width and height difference value on the basis of CIOU, and Focal Loss solves the problem of difficult sample imbalance [15].

This paper finds that the YOLOV5s framework still has room for improvement in terms of speed and accuracy. There is a problem of low detection accuracy for object detection methods. An improved object detection method based on average pooling YOLOV5s is proposed. The improved YOLOV5s model comprehensively improves detection performance and deployment capabilities. The main improvements are as follows:

(1) The SE attention mechanism is introduced, and the improved YOLOV5s model effectively improves the feature extraction ability of the model.

(2) EIOU Loss is introduced, and the improved YOLOV5s model effectively improves the accuracy of model object detection, accelerates the calculation speed, and ensures that the object detection model obtains more important feature information.

## 2. Method

As shown in Fig. 1, The original YOLO model introduced the SE attention mechanism. SE is a mechanism for assigning weight parameters with the goal of assisting the model in capturing important information. The attention mechanism of mobile network is introduced, the model complexity is low, and the model detection effect is greatly improved. The IOU (Intersection Over Union Loss, IOU) loss of the original YOLO model directly uses the IOU value between the predicted bounding box and the real bounding box as the loss [18], but the gradient of the loss function is small and cannot accurately reflect the coincidence of the two boxes. Therefore, it is proposed to introduce EIOU (Efficient Intersection Over Union Loss, EIOU), replace the aspect ratio with the width and height difference value on the basis of CIOU, Focal Loss solves the problem of unbalanced difficult samples, and greatly improves the accuracy of detecting surface defects of industrial products.

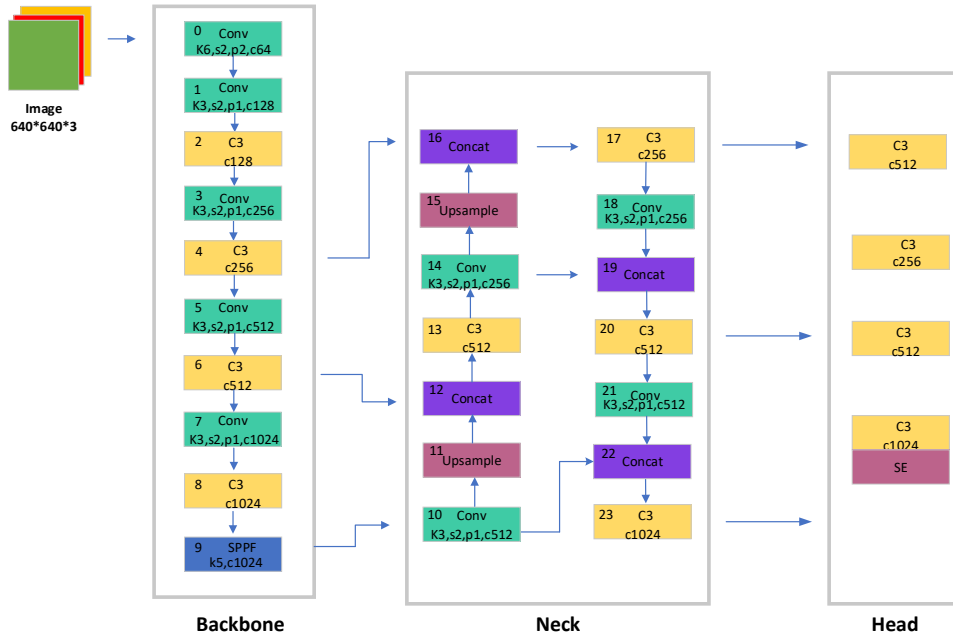


Fig 1. Improved YOLOV5s network structure

## 2.1. Introduction of the SE Attention Mechanism

In traditional CNN architectures, convolutional layers and pooling layers are usually used to extract image features. However, this approach does not explicitly model the relationships between feature channels, resulting in some channels contributing relatively little to a particular task, while others are more important. The SE module is designed to solve this problem. The pooling process of the SE attention mechanism is shown in Fig. 2. The SE module models the relationship between channels by introducing a Squeeze action and an Excitation operation. In the Squeeze phase, it compresses the output feature map of the convolutional layer into a feature vector through the global average pooling operation. Then, in the Excitation phase, by using fully connected layers and nonlinear activation functions, learn to generate a weight vector for a channel. This weight vector is applied to each channel on the original feature map to weight the features of the different channels. The SE attention mechanism helps the network to better focus on important feature channels, thereby improving model performance.

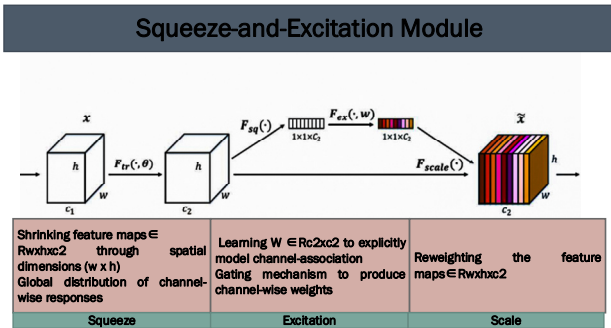


Fig 2. SE attention mechanism

## 2.2. EIOU

Because the aspect ratio of the CIOU Loss prediction box and the GT box is linearly scaled, it will hinder the effective optimization similarity of the model. It is proposed to introduce EIOU in the original YOLOV5s model. The influence factor of the aspect ratio of the prediction box and

the real box is separated, and the length and width information of the target box and Focal focus prediction box are added. The length and width of the prediction box and the real box are calculated respectively to solve the problem of penalty failure in the proportional change of aspect ratio in CIOU. The regression process focuses on high-quality anchor frames, which solve the problems of sample imbalance, slow model convergence, and imprecision. At the same time, the prediction box regression accuracy is improved.

## 3. Result

### 3.1. Dataset and Experiment Environment Configuration

The dataset discloses the steel surface defect dataset NEU-DET dataset practice. The original data has a total of 5404 images, and there are only 30 verification images in the original data verification set file. To improve the accuracy of test recall, validation set preprocessing in the original data is performed first. Enrich the validation set file image and tags. 80% of the original data is used for training, 10% for validation, and 10% for testing. Tab. 1 describes the configuration of the experimental environment.

Table 1. Experimental Environment Configuration

Environment	Name
Operation system	Windows 10
CPU	Intel (R) Core (TM) i7-10700F CPU @2.90 GHz
GPU	NVIDIA GeForce RTX 2070 SUPER
Code version	YOLOV5-6.0
Learning framework	Pytorch1.8.0

### 3.2. Ablation Studies

Ablation experiments were carried out to investigate the feasibility of improving the model based on average pooled YOLOV5s object detection. The results of different

improvement strategies for ablation experiments are shown in Tab. 2. Observe the experimental data to find an improved object detection network model through ablation experiments.

**Table 2.** Comparison of Defect Detection Accuracy of Different Improvement Strategies

Model	+SE	+EIOU	mAP@0.5 (%)	FPS
Primeval YOLOV5s			74.4	90
YOLOV5s-A	√		81.1 (+6.7)	103 (+13)
YOLOV5s-B		√	80.8 (+6.4)	101 (+11)
YOLOV5s-OUR	√	√	81.8 (+7.4)	99 (+9)

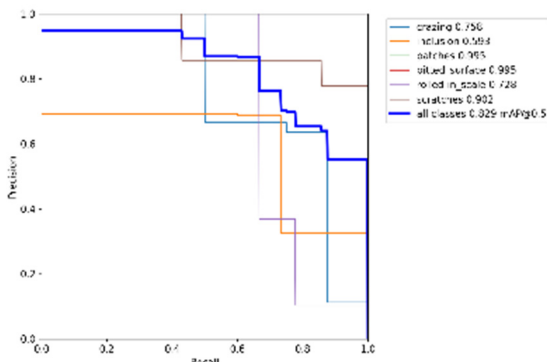
### 3.3. Experimental Comparison

Firstly, under the condition of ensuring the experimental variables, a comparative experiment of the YOLOv7 network model was added. The average accuracy (mAP) of the YOLOv7 object detection network model can reach 80.3%, which is 1.5% lower than that of the improved model. To verify the performance of the improved YOLOV5s object detection method based on average pooling. Compare experiments under the same data set division. In this experiment, four target detection networks were used: YOLOV5s, YOLOV5n, YOLOV5x, and YOLOV7. Tab. 3 shows the training results of different YOLO series algorithm models. The data in row 1 is the training results of the YOLOV5s model, and the mAP@0.5 is 74.4%. The data in row 2 is the training result of the YOLOV5n model, and the mAP@0.5 is 77.6%. The data in row 4 is the YOLO V7 model training results, and the mAP@0.5 is 80.3%. The YOLOV5 series performs well in both speed and precision. The average accuracy (mAP) of the YOLOV5s target detection model based on average pooling can reach 81.8%, and the overall performance is better than that of other YOLO models, which indicates that the improved method in this paper can identify and locate small targets more accurately.

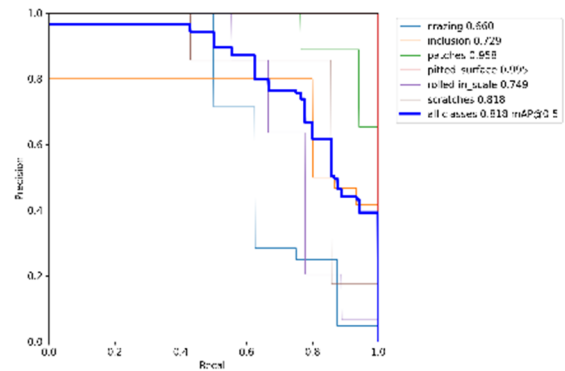
**Table 3.** Data Comparison

Model	mAP@0.5(%)	P (%)	R (%)	FPS
1 YOLOV5s	74.4	76.7	67.8	90
2 YOLOV5n	77.6(+3.2)	65.4	81.4	108(+18)
3 YOLOV5x	82.2(+7.8)	76.9	77.9	77(+13)
4 YOLOV7	80.3(+5.9)	81.2	76.2	61(+29)
SOUR	81.8 (+7.4)	86.8	87.3	99 (+9)

## 4. Conclusion



**Fig 3.** Original -YOLOV5s



**Fig 4.** YOLOV5s-OUR

In the PR image, the closer the polyline to the upper right, the better the performance of the object detection network model. The performance of the original YOLOV5s model is shown in Fig. 3 Original -YOLOV5s, and the improved performance of the YOLOV5s model based on average pooling is shown in Fig. 4 YOLOV5s-OUR. It can be found that the improved performance is better than the original YOLOV5s object detection network model. EIOU bounding box regression loss is introduced to further improve the detection accuracy [16] [17]. The SE attention mechanism object detection network model is selected to effectively reduce the loss of feature map information and improve the recognition accuracy. The YOLOV5s network structure with fast detection speed and excellent accuracy value is obtained.

A detection classification method based on improved YOLOv5s is proposed. The SE module is integrated in the backbone feature extraction network stage, so that the backbone network highlights useful features and pays more attention to small target features. In addition, the EIOU module is introduced to better integrate the characteristics of each scale. The test results of the proposed method are compared with the YOLOv5n, YOLOv5x, and YOLOv7 calculations. The experimental results show that compared with the original YOLOv5s algorithm, the improved network has a great improvement in the false and missed detection situation, and can accurately and quickly detect the target.

## References

- [1] Ying Z, Lin Z , Wu Z.A modified-YOLOv5s model for detection of wire braided hose defects[J].Measurement, 2022(190-):190.DOI:10.1016/j.measurement.2021.110683.
- [2] Abdulghani A M, Abdulghani M M , Walters W L. Multiple Data Augmentation Strategy for Enhancing the Performance of YOLOv7 Object Detection Algorithm[J].Tech Science Press, 2023.DOI:10.32604/jai.2023.041341.
- [3] Mao Q, Wang M , Hu X . Intelligent Identification Method of Shearer Drums Based on Improved YOLOv5s with Dark Channel-Guided Filtering Defogging[J]. 2023.
- [4] Xi D, Qin Y , Wang S .YDRSNet: an integrated Yolov5-Deeplabv3-real-time segmentation network for gear pitting measurement[J].Journal of Intelligent Manufacturing, 2023.
- [5] Shaikh S A, Chopade J J , Sardey M P .Real-Time Multi-Object Detection Using Enhanced Yolov5-7S on Multi-GPU for High-Resolution Video[J].International Journal of Image and Graphics, 2023.DOI:10.1142/S0219467824500190.
- [6] Panigrahi S , Raju U S N .MS-ML-SNYOLOv(3): A robust lightweight modification of SqueezeNet based YOLOv(3) for pedestrian detection[J]. Optik: Zeitschrift fur Licht- und

- Elektronenoptik: = Journal for Light-and Electronoptic, 2022 (260-): 260.
- [7] Iyer R , Bhensadiya K P, Priyansh Ringe.Comparison of YOLOv3, YOLOv5s and MobileNet-SSD V2 for Real-Time Mask Detection[J]. 2021.
- [8] Ying Z, Lin Z , Wu Z. A modified-YOLOv5s model for detection of wire braided hose defects[J]. Measurement, 2022 (190-): 190.DOI:10.1016/j.measurement.2021.110683.
- [9] Mao Q, Wang M , Hu X . Intelligent Identification Method of Shearer Drums Based on Improved YOLOv5s with Dark Channel-Guided Filtering Defogging[J]. 2023.
- [10] Xu Q, Liang Y , Wang D . Hyperspectral Image Classification Based on SE-Res2Net and Multi-Scale Spatial Spectral Fusion Attention Mechanism[J].Journal of Computer-Aided Design & Computer Graphics, 2021, 33(11):1726-1734.DOI: 10.3724/SP.J.1089. 2021.18778.
- [11] Hu X, Jing L , Sehar U .Joint pyramid attention network for real-time semantic segmentation of urban scenes[J].Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2022(1):52.
- [12] Zhang Y F, Ren W , Zhang Z. Focal and Efficient IOU Loss for Accurate Bounding Box Regression[J]. 2021.DOI: 10.48550/ arXiv.2101.08158.
- [13] Shen Y, Zhang F , Liu D. Manhattan-distance IOU loss for fast and accurate bounding box regression and object detection[J]. Neurocomputing, 2022.
- [14] Tang J, Liu S , Zhao D . PCB-YOLO: An Improved Detection Algorithm of PCB Surface Defects Based on YOLOv5[J]. 2023.
- [15] Huang S , He Y , Chen X A .M-YOLO: A Nighttime Vehicle Detection Method Combining Mobilenet v2 and YOLO v3[J]. Journal of Physics Conference Series, 2021, 1883(1): 012094. DOI:10.1088/1742-6596/1883/1/012094.
- [16] Yang B, Wang J .An Improved Helmet Detection Algorithm Based on YOLO V4[J].International Journal of Foundations of Computer Science, 2022.DOI:10.1142/S0129054122420205.
- [17] Li D, Zhang Z , Wang B. Detection method of timber defects based on target detection algorithm[J].Measurement, 2022.