ISBN: 978-981-18-3959-7
2022 the 12th International Workshop on Computer Science and Engineering (WCSE 2022)
doi: 10.18178/wcse.2022.06.027

Surface Defect Detection of Components Based on SE-YOLOv5

Zheng Lehui [1+], Sun Junjie [1] Niu Rui [1], and Huang Ying [2]

[1] Graduate School, Engineering University of PAP, Xi'an 710086, China;

[2] Graduate School, Information and Communication, Engineering University of PAP, Xi'an 710086, China;

Abstract. With the continuous development of modern industry, all kinds of equipment have higher and higher requirements for components, and minor faults of components may cause great impact on equipment. Therefore, it is very important to detect component defects under modern industrialization conditions. The contrast and defect characteristics of industrial components are different, and the traditional visual detection method has poor effect and low accuracy, which cannot meet the requirements of modern industrial detection. Aiming at the deficiency of traditional visual detection for small defect target detection, a fault diagnosis method for surface defects of industrial components based on SE-YOLOv5 is proposed in this paper. YOLOv5 is very suitable for visual defect detection because of its fast calculation speed and good expression of bearing defect characteristics. However, its accuracy of defect detection is not high enough. Therefore, SE attention module is embedded on the basis of YOLOv5 and the new model SE-YOLOV5 is obtained, which not only greatly improves the calculation speed, but also improves the accuracy of visual defect detection. Moreover, higher accuracy of defect diagnosis is obtained.

Keywords: SE-YOLOV5 defect detection SE attention module

1. Introduction

Defect detection is an essential part of modern industry, and it is also an important research direction in mechanical vision field. As an important component of mechanical equipment, various components play an important role in mechanical engineering. Mechanical failure is mainly caused by component failure, and bearing and tile magnetic components as an important part of the component, its probability of failure is the largest. Therefore, the existence of surface defects of bearings and tile magnets will inevitably affect the life of machinery. In the production process, surface defects of bearings and tile magnets and other components must be detected. How to solve the problems of real-time detection of defects of components and components and the difficulty of detection due to the low contrast and small defects of bearing and tile magnetic components is the focus and difficulty of defect detection. The traditional machine vision algorithm is difficult to complete the modeling and migration of defect features, which has low reusability and requires to distinguish working conditions, which will waste a lot of labor costs. As deep learning has achieved very good results in feature extraction and location, more and more scholars and engineers begin to introduce deep learning algorithms into the field of defect detection.

Dong et al. [1] conducted real-time detection of key components of power lines based on YOLOv3 model, which improved the detection effect of target detection network on small target objects. Guo et al. [2] detected the magnetic tile surface defects through YOLOv3 model and the deeper network of Resnet residual structure. The detection effect was no worse than resNet101-based Faster R-CNN method, and its average detection speed was more than 5 times Faster. Electric Power Research Institute of Shanxi Electric Power Company, State Grid [3] takes YOLOv3 as the basic framework to build a standard YOLOV3-TINY network and an insulator detection model to achieve end-to-end efficient and accurate detection. The YOLOv3 model used by them can improve the detection speed to some extent, but the accuracy of defect detection is far from the standard of modern industry. Chang Haitao et al. [4] applied Faster R-CNN to industrial defect detection to correct the position of prediction box through pre-processing, achieving good

+ Corresponding author. Tel.:+18717327102;
E-mail address:981543794@qq.com.

detection results, but this also greatly reduced the speed of model detection and failed to achieve the expected results. Li Ming et al. [5] expanded the data set through GAN, and then conducted defect detection through Faster R-CNN, which solved the difficulty in collecting industrial data sets to a certain extent, but did not fundamentally improve the detection speed. Sun Hui et al. [6] realized online detection of surface defects by improving SSD. Zhang et al. [7] improved the accuracy of bearing fault diagnosis by constructing an integrated learning model (ASPNN) based on AdaBoost. Shi et al. [8] improved HHT through the set empirical mode decomposition improved by multi-population differential evolution and sensitive inherent mode function screening method, and extracted time-frequency characteristics of fault signals, which greatly improved the accuracy of fault diagnosis and solved the problem of over-fitting. She [9] decomposed the signal through empirical mode decomposition (EEMD) and SVM, and combined with CNN model, solved the endpoint divergence problem and improved the accuracy. Through TCNN and STFT based on transfer learning, online CNN and offline CNN are constructed in literature [10] to improve real-time performance and achieve desired diagnostic accuracy within limited training time. Although they overcome the shortcoming of small sample number to a certain extent and improve real-time performance, they still cannot meet the requirements of precision and rapidity of modern industrial defect detection. In the continuous development of YOLO model, YOLOv5 model has also been continuously applied in the field of detection. Yu et al. [11] proposed a mask wearing detection algorithm based on YOLOv5 to achieve real-time detection in complex scenes, greatly improving the detection speed. Zhang et al. [12] proposed a firework detection algorithm integrating yOLOV5-RESNET cascade network, which achieved better results compared with other fire detection algorithms on the open data set and had better comprehensive performance than other existing fire detection algorithms. Wang et al. [13] proposed an improved lightweight YOLOv5 algorithm to realize the defect detection of insulators, which greatly improved the detection speed and accuracy and made it easy to deploy in the embedded middle end. It can be seen that the detection speed and accuracy of YOLOv5 are greatly improved compared with the previous YOLO model.

Therefore, this paper takes bearings and tile magnets as examples, uses YOLOv5 model with fast detection speed and high accuracy, and makes improvements on the basis of YOLOv5 model. SE attention module is embedded on the basis of YOLOv5 model, and the new model SE-YOLOV5 is obtained.

2. Algorithm Description of YOLOv5

YOLOv5 model is a model developed from YOLO model in recent years, and there are YOLOv3 and YOLOv4 models before it. Its advantage lies in that it can directly process single image, multiple images, video and even the port input of webcam effectively, and its running speed is fast, so it is very suitable for defect detection in industry. In this paper, SE-YOLOv5 is improved based on YOLOv5 model.

The network structure diagram of Yolov5 is divided into four parts: input terminal, Backbone, Neck and Prediction.

(1) Input: Mosaic data enhancement and adaptive anchor frame calculation
(2). Backbone: Focus structure, CSP structure
(3) FPN+PAN
(4) Output end Prediction: GIOU_Loss, Bounding box

2.1. The Input

(1) Mosaic data enhancement

Yolov5's input uses the same Mosaic data-enhanced approach as Yolov4's. Mosaic data enhancement uses random scaling, random clipping and random arrangement for stitching, and has a very good detection effect for small targets.

(2) Adaptive anchor frame calculation

In Yolo algorithm, there are anchor frames with initial length and width for different data sets. In the network training, the network outputs the prediction box on the basis of the initial anchor frame, and then compares it with the groundtruth of the real frame, calculates the gap between them, and then reversely updates and iterates the network parameters.

(3) Adaptive picture scaling

In common target detection algorithms, different images are different in length and width, so the common way is to uniformly scale the original image to a standard size, and then send it to the detection network. For example, Yolo algorithm often uses sizes of 416×416 ×608 ×608.

2.2. Backbone

(1) Focus structure

Taking the structure of Yolov5 as an example, the original 608×608×3 image was input into the Focus structure, and the slice operation was adopted to transform it into a feature map of 304×304×12 first, and then through a convolution operation of 32 convolution kernels, it finally became a feature map of 304×304×32. See Figure 1

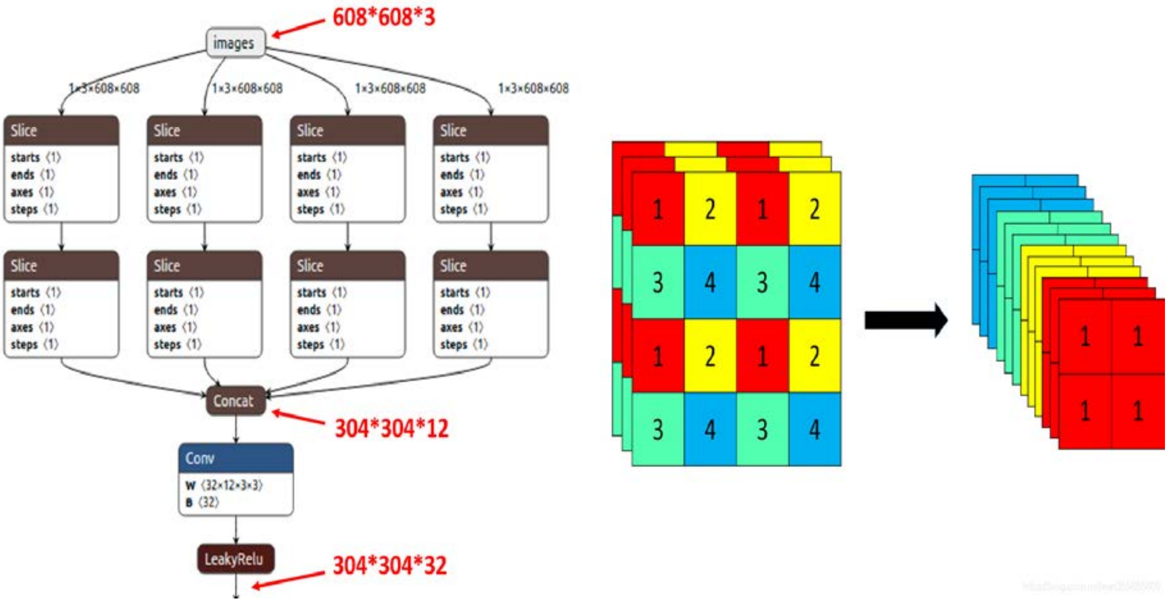

Fig. 1: Function diagram structure of Focus junction

2.3. Neck

The Neck of Yolov5 now adopts the FPN+PAN structure just like that of Yolov4. However, when Yolov5 first came out, only FPN structure was used, and then PAN structure was added. In addition, other parts of the network were also adjusted. In Yolov5's Neck structure, the CSP2 structure designed by reference to CSPNet is adopted to strengthen the ability of network feature fusion.See Figure 2

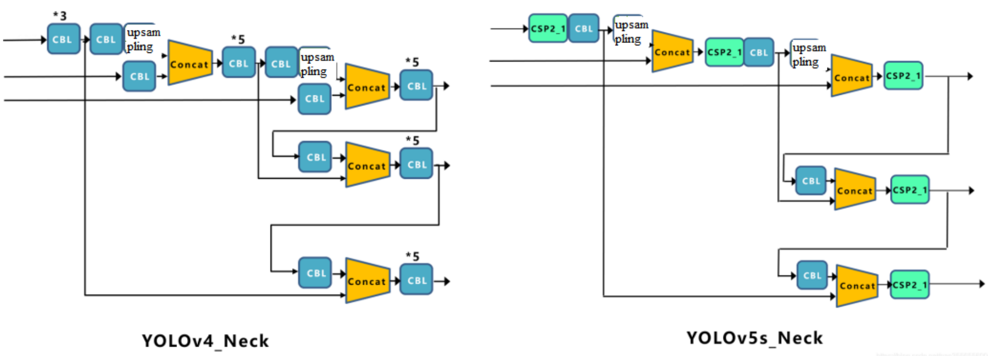

Fig. 2: Comparison of neck structure of Yolov4 and Yolov5

2.4. The Output End

(1) Bounding box loss function

In Yolov5, GIOU_Loss is used as the loss function of Bounding box.

(2) NMS is non-maximum suppression

In the post-processing process of target detection, NMS operation is usually required for the screening of many target boxes, and Yolov5 adopts the weighted NMS method.

Two parameters depth_multiple and width_multiple are designed in YOLOv5. Depth_multiple is used to control the number of residual structures in CSP1_x. For example, the number of residual structures set in backbone is 9. Depth_multiple in YOLOv5s is 0.33, and the number of residual structures in the final network structure is 3. Width_multiple is used to control the number of channels when feature graphs are transmitted in the network, that is, to control the number of convolution kernels in the convolution layer. For example, the number of convolutional kernels originally set in the convolution layer of Foucs structure is 64, while the width_multiple of YOLOv5 is 0.5. The number of convolutional kernels in the Foucs structure of YOLOv5 network is 32, and the number of channels output by the feature map through the Foucs structure is also 32. The reasoning speed of the target detection model trained by YOLOv5 based on Bdd100k data set can reach 7ms, while the model size is only 14.8m, which is suitable for GPU mounted on embedded platform with low computing power. It is suitable for defect detection with fast operation speed and short detection time.

3. Experimental Process

3.1. Data Set Acquisition

The data set in this paper is the images collected from the actual production line of components, including common faults of components: surface scratches, cracks, scarring, and lateral wear of the outer ring of components. There are 1200 single-channel images of 494×648 for each defect category, 2400 images in total, which are evenly divided into training set, verification set and test set according to 2:1:1. Some defect images are shown in Figure 3.

Fig. 3: Partial defect graph of data set

3.2. Image Preprocessing

Due to the influence of light and other environment, the resolution and contrast of the photos obtained can not reach the expected target when obtaining the photos of the components with faults. Defect detection needs to obtain a good image of the inspection target, so it is necessary to select the appropriate light source according to the characteristics of the workpiece and the field environment when obtaining the image of the component surface defects. After obtaining the image of the target workpiece, the image should be preprocessed to improve the resolution of the image, so as to achieve a good detection effect.

There is only one notch defect in the obtained image, occupying a small size of the original image, and the gray value is distributed in a certain area, and the gray value of the defect and background is mainly in this area, which is very unfavorable to the defect feature extraction. After pretreatment, the contrast of the defect area is improved to a certain extent, and a certain amount of gray value in the histogram is evenly distributed to other areas, so as to improve the contrast of the defect area and limit the increase of gray value in the background area.

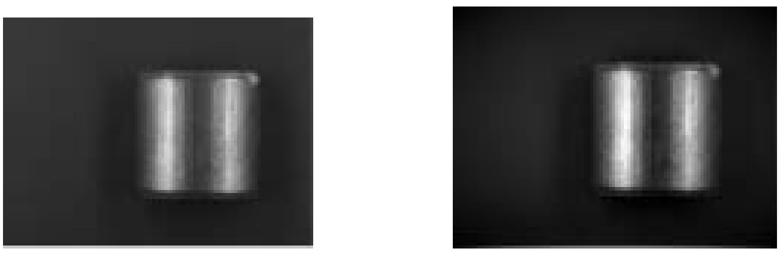

Fig. 4: Unpretreated versus pretreated

The YOLOv5 model used in this paper can enhance the data set by rotating, mirroring, changing the brightness, Gaussian filtering, translation and scaling, and expanding the data set by 11 times. After pretreatment, the image has obvious improvement in sharpness and gray contrast. 80% of the data set was used as the training set, 15% as the verification set and 5% as the test set, and the image data was expanded. The expanded photos showed obvious improvement in brightness and contrast, which laid a good foundation for the subsequent model establishment.

3.3. Experimental Environment and Evaluation Index

The neural network training platform used in this experiment was Ubuntu18.04 with Intel®Core™i7-9700K@3.60HZ×8 CPU and GeForce RTX™2080 GPU. NVIDIA CUDA10.2 computing platform and CUDNN7.6.5GPU acceleration library are adopted. Neural network framework is Pytorch1.5.

The operating system of the test platform used in this experiment was Ubuntu18.04. The CPU was composed of two ARM V8 clusters with high-performance consistent interconnection structure. The GPU was an NVIDIA Pascal GPU equipped with 256 CUDA-supporting cores and adopted NVIDIA CUDA10.0. Computing platform and cudn7.6.3 GPU acceleration library, neural network framework using Pytorch1.5. In order to evaluate the performance of the network and explain the effectiveness of the target detection network, the following evaluation indicators were selected:

Accuracy and recall rate: Accuracy refers to the ratio of positive samples detected by the network to all detected samples. Recall rate refers to the ratio between the number of positive samples detected by the network and the number of marked real samples. The calculation formula of accuracy rate and recall rate is as follows:

$$\mathrm{Precision}=\frac{TP}{TP+FP} \tag{1}$$

$$\mathrm{Recall}=\frac{TP}{TP+FN} \tag{2}$$

In the formula, True Positive (TP) represents the number of samples in which the detected object class is consistent with the real object class. False Positive (FP) is the number of samples in which the detected object class is inconsistent with the real object class. False Negtive (FN) is the number of samples that exist in the real object but are not detected by the network.

(2) Average accuracy (AP) and Mean Average Precision (mAP) : The ideal target detection network should maintain high accuracy while increasing recall rate. The reality, however, is that an increase in recall often requires a loss of Precision. Therefore, the precision-recall (P-R) curve is usually used to show the balance between Precision and Recall of an object detector. For each class, the average accuracy of the class is defined as the area under the p-R curve. Mean average accuracy is the average of the average accuracy of all categories. AP and mAP are calculated as follows:

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{3}$$

$$mAP=\frac{\sum_{i=1}^{N} AP_i}{N} \tag{4}$$

(3) Frames per second (FPS) : the number of images that the target detection network can detect per second. This index is used to evaluate the detection speed of the target detection network.

4. Bearing Fault Detection Results and Analysis

We first trained 200 epochs of the original YOLOv5 model on the training set, and then saved and tested the obtained models. Then, we improved YOLOv5 based on deeply separable convolution, replaced the traditional convolution in backbone with deeply separable convolution, obtained depth-Yolov5, and trained 200 epochs on the training set of the improved model. The detection results are shown in Table 1.

Table 1: AP detection results of depth-YOLOv5 model

	Fault 1	Fault 2	Fault 3	Fault 4	mAP
Precision	0.614	0.508	0.587	0.667	
Recall	0.807	0.785	0.851	0.891	
AP	0.797	0.766	0.841	0.900	0.823

The SE attention module was embedded at the end of CSP1_x to evaluate the importance of each CSP1_x output channel and enhance it according to the importance. The improved model was trained with 40 Epochs and named SE-YOLOV5. The detection results are shown in Table 2.

Table 2: AP detection results of SE-YOLOv5 model

	Fault 1	Fault 2	Fault 3	Fault 4	mAP
Precision	0.637	0.527	0.604	0.691	
Recall	0.836	0.812	0.884	0.925	
AP	0.826	0.795	0.872	0.911	0.851

In order to facilitate intuitive comparison, we compared the detection AP of the three models, and their data pairs are shown in Table 3. According to the comparison in Table 3, it is not difficult to find that THE AP value of SE-YOLOV5 is significantly improved compared with the unimproved YOLOv5 model and the depth-YOLOV5 model, indicating that the new SE-YOLOV5 can improve the accuracy of defect detection compared with the previous model.

Table 3: AP comparison of YOLOv5, depth-YOLOv5 and SE-YOLOv5

	YOLOv5	depth-YOLOv5	SE-YOLOv5
Fault 1（AP）	0.735	0.797	0.826
Fault 2（AP）	0.742	0.766	0.795
Fault 3（AP）	0.812	0.841	0.872
Fault 4（AP）	0.876	0.900	0.911

According to the detection results of the new model SE-YoloV5, the detection results of the constructed model after training are shown in the figure 5.

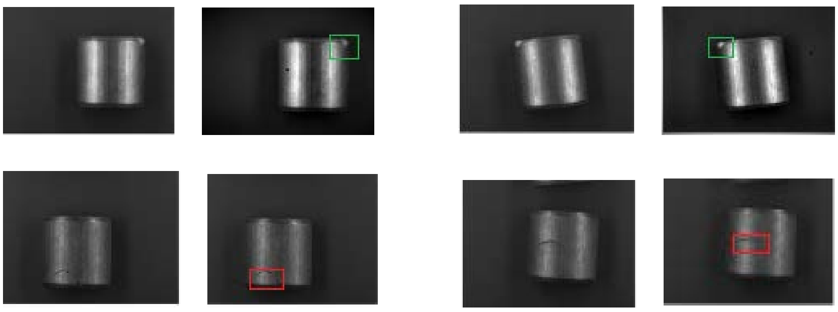

Fig. 5: SE-YOLOv5 test result diagram

To test the change of SE-YOLOV5s computing speed, mAP tests of YOLOv5, depth-YOLOV5-ELU and SE-YOLOV5 are shown in Table 4

Table 4: Speed comparison of YOLOv5, depth-YOLOv5 and SE-YOLOv5

Algorithm	FPS
YOLOv5	11.53
depth-YOLOv5	13.61
SE-YOLOv5	19.26

Compared with YOLOv5 and depth-YOLOV5, THE mAP and FPS of SE-YOLOV5 are both greatly improved, which proves that the detection accuracy and speed of SE-YOLOV5 are greatly improved, and the SE-YOLOV5 has good generalization ability, which is more suitable for defect detection in industrial environment.

5. Conclusion

This paper discusses the evaluation indexes such as the accuracy rate, detection speed and positioning accuracy mAP in the defect detection of industrial components such as bearings and tiles. In this paper, on the detection model of YOLOv5, a preprocessing component defect detection algorithm is added to enhance the data set through rotation, mirror, brightness change, Gaussian filtering, translation and scaling, and data set expansion, so as to better ensure the feature extraction of small defects and low contrast defects. At the same time, SE attention module is embedded after YOLOv5, which improves the effect of small target defect detection to a certain extent. In general, the algorithm can quickly locate and classify defects, improve the speed and accuracy of defect detection to a certain extent, and has good generalization ability. In the next step, the speed and accuracy of defect detection will be further improved with the help of better models and networks.

6. References

[1] DONG Zhaojie. Real-time detection of key components of power line based on YOLOv3 [J]. Electronic measurement technology, and 2019 (23) : 173-178. The DOI: 10.19651 / j.carol carroll nki emt. 1903141.

[2] GUO Longyuan, TONG Guanghong, DUAN Houyu, et al. Journal of chengdu institute of technology,2019,22(3):25-30. DOI:10.13542/j.cnki.51-1747/tn.2019.03.006.

[3] State Grid Shanxi Electric Power Company Electric Power Research Institute. A Fault Detection Method for Aerial Images of Insulators in Transmission and transformation Lines Based on Improved YOLOv3 :CN201910310921.0[P].2019-07-18.

[4] Chang Haitao, GOU Junnian, Li Xiaomei. Application of Faster R-CNN in defect detection of industrial CT images [J]. Journal of Image and Graphics, 2018, 23(7):129-139.

[5] LI Ming, JING Junfeng, LI Pengfei. Defect detection of colored fabric using GAN and Faster R-CNN [J]. Journal of Xi 'an Polytechnic University, 2018, 32(6):44-50.

[6] SUN H, SUN X K, SUN Y, et al. Online detection system for surface defects of hot galvanized sheet based on improved SSD model [J]. Electronic Technology, 2018, 47(12):112-115.

[7] Zhang Xixi, GU Xingsheng. Fault Diagnosis of Motor Bearing Based on Ensemble Learning Probabilistic Neural Network [J]. Journal of East China University of Science and Technology (Natural Science), 2020, 46(1): 68-76.

[8] Shi Jie, WU Xing, LIU Tao. Hybrid fault diagnosis of bearings based on HHT algorithm and convolutional neural network [J]. Transactions of the csae,2020,36(4):34-43.

[9] He Jiangjiang, Li Xiaoquan, Zhao Yuwei, Zhang Baoshan, Ding Haibin. Rolling bearing fault diagnosis based on improved EEMD convolutional neural network [J]. Journal of chongqing university (natural science edition),2020,43(1):82-89.

[10] MAO Guantong, Liu Hong, WANG Jinglin. Online Fault Diagnosis of Rolling Bearings based on Transfer Learning [J]. Aeronautical Science and Technology,2020(1):61-67.

[11] Yu Shuo, Li Hui, GUI Fangjun, Yang Yanqi, Lv Chenyang. Research on mask wearing real-time detection algorithm based on YOLOv5 in complex Scene [J]. Computer Measurement and Control, 201,29(12):188-194.

[12] Quan Zhang, Wei Zhang, Xianfeng Yang, Bo Bo, Shuyan Liu. Pyrotechnic detection method based on YOLOV5-RESNET Cascade Network [J]. Journal of Safety and Environment,,,:1-10.

[13] Wang Shukun, Gao Lin, Fu Desu, Liu Wei. Research on Improved Lightweight YOLOv5 Insulator Defect Detection Algorithm [J]. Journal of Hubei University for Nationalities (Natural Science Edition), 201,39(04):456-461.