Defect detection of automotive injector valve seat based on feature fusion

Zonghong ZHU1, Chungui Li2*

1School of Electrical and Information Engineering, GuangXi University of Science and Technology, Liuzhou 545001, China
2School of Computer Science and Communication Engineering, GuangXi University of Science and Technology, Liuzhou 545001, China, e-mail: liza4323@163.com

Abstract. Automobile fuel injector seat is a device that injects gasoline into the automobile cylinder. It plays a very important role in the control of automobile fuel quantity. However, due to the small parts and susceptibility to processing technology, in the production process of automobile fuel injector seats, it is inevitable to leave scratches, defects, rust spots, white spots and other defects. This paper will use the depth detection technology to complete the flaw detection of the injector seat, the depth detection algorithm model Faster R-CNN is improved. Defect detection can be seen that when extracting features, the Faster R-CNN algorithm model only uses the features of the last layer of convolution output, and has a large loss and affects the effect of small target detection. In order to improve the detection ability and solve the problem of missing detection caused by multi-scale and small targets, we introduce the idea of multi feature fusion in the stage of feature network extraction, and compares the improved algorithm model with the original Faster R-CNN model on the data set of injector seat. It is found that the improved model can be better applied to the flaw detection of automobile injector seats.

1. Introduction

Due to the rapid development of society, automobiles have become an increasingly important travel tool in daily life. As a device for injecting gasoline into automobile cylinders, the valve seat of automobile injectors plays a very important role in fuel quantity control. How to improve the quality of the parts has become an important issue of concern, but because of the small size of parts, it is easy to be limited by the processing technology. During the production process, it will inevitably leave scratches, defect, rust spots, white spots and other types of defects inside, which affects the performance of the automotive injector seat. Therefore, picking out defective parts from many parts has become an inevitable project. With the rapid increase of image data and the rapid progress of hardware computing ability, the deep learning detection technology, represented by convolutional neural network, has been applied to the related tasks of flaw detection. Compared with the traditional algorithm, the performance has been greatly improved. In 2014, Ross Girshick [1] and others proposed the R-CNN algorithm to extract candidate regions through a selective search algorithm, but the algorithm is computationally intensive and slow. Subsequently, the target detection algorithm SPP-Net is proposed, which solves the problem of object deformation, and then Fast R-CNN is proposed by introducing multi-task loss and RoI Pooling, which uses multi-task learning to complete classification and regression. However, the regional method adopted by the algorithm will still consume a lot of time. Therefore, Ren [2] proposed the Faster R-CNN algorithm. The algorithm introduces the RPN network...
on the basis of the Fast R-CNN algorithm, which has been greatly improved in speed and performance. The Faster R-CNN algorithm can achieve better results in object detection than other algorithms. Although the Faster R-CNN algorithm has good detection performance in the detection of objects, the seat defect size of automobile fuel injector is relatively small and there are many kinds of defects. Therefore, the Faster R-CNN detection is used in the process, it is impossible to accurately complete the identification and positioning of defects, which is likely to cause a missed inspection. In this paper, we introduce the idea of feature fusion on the Faster R-CNN algorithm, fuse the features of different convolution layers, improve the expression ability of the detection algorithm, and make it more accurate to detect the defects of the valve seat of the automobile injector.

2. Dataset Construction

2.1 Image Data Processing

In the process of collecting defects in the valve seat of the automobile injector with the help of hardware such as CCD industrial cameras, tooling, PC, etc., due to the interference of the environment, current, operation and other factors, the collected pictures will increase the difficulty of subsequent operations, in order to simplify Subsequent work requires effective methods to preprocess the images in actual production. First, during the image acquisition process, there will be problems such as image redundancy and naming irregularities during saving. Redundant images will not only affect the work the efficiency has a great impact, and it will increase the difficulty of subsequent work. Therefore, it is necessary to remove duplicate pictures. Secondly, in the collection In the process of the picture, due to the influence of current and noise, some irrelevant information will be generated. Therefore, it is necessary to use the Gaussian filtering method to denoise the image and retain the useful information for detection and recognition. Therefore, in the detection of the injector valve seat, the picture needs to be compressed, and the picture size is processed to 800 ×600, after obtaining the unified standard image data, the data enhancement method is used to avoid data shortage, and the model generalization ability is enhanced. Data enhancement is an important part of training deep learning models [3]. There are generally two ways to increase data. One is to add a data perturbation layer to the network model to allow the image to be trained every time, there is another way that is more straightforward and simple, the image samples are enhanced by image processing before training, we expand the data set using image enhancement methods such as geometry and color space, and use HSV in the color space, as shown in Figure 1.

![Figure 1. Picture preprocessing.](image)

(a) Primal  
(b) Hue diminished  
(c) Saturation reduction  
(d) Decrease in value  
(e) Flip  
(f) Hue enhancement  
(g) Saturation enhancement  
(h) Zoomed

Figure 1. Picture preprocessing. (b)–(f) are the transformed images of 20 pixels of chroma increase and decrease; (c)–(g) indicate that the saturation becomes 0.1 and 1.3 times of the original; (d) the brightness is increased by 0.5 times of the original; (e) and (h) are the flip and zoom graphs in the geometric space, respectively.
3. Improvement of Faster R-CNN defect detection model

In the Faster R-CNN algorithm model, first of all, you need to extract the features of the input picture, and the extracted output features can directly affect the final detection effect. The core of object detection is feature extraction. The common feature extraction network in the Faster R-CNN algorithm model is the VGG-16 network. This network model was first used in image classification [4], and then it has been excellent in semantic segmentation [5] and saliency detection [6]. The feature extraction network in the Faster R-CNN algorithm model is set to VGG-16, although the algorithm model has a good performance in detection, it only uses the feature map output from the last layer in image feature extraction, so there will be some losses and the feature map cannot be fully completed, which will lead to inaccuracy in detection of small target objects and affect the final recognition effect. In convolutional networks, shallower convolutional layers can better extract local features, which is beneficial to the positioning of objects in detection; while deeper networks can better extract global features, which is beneficial to classification in detection. In this paper, based on the VGG16 network structure, the features of different convolutional layers are merged to form a multi-feature map to achieve complementary features, thereby improving the efficiency of object detection. Specifically, the output features of Conv2_2, Conv4_3, and Conv5_3 are fused, and the above network feature layer is selected. The fusion of deep networks will enhance semantic attributes and improve category judgment ability.

The input image size is 800×600, the second layer convolution conv2_2 output characteristics is 400×300, the fourth layer convolution conv4_3 output characteristics is 100×75, and the fifth convolution layer conv5_3 output feature map is 50×37, so conv2_2 is down-sampled, and conv5_3 is deconvolved to obtain a fixed convolution map, and the size of the output feature map is consistent with the resolution of the conv4_3 output feature map to 100×75.

However, because the activation values of different convolution layers are different, directly connecting the feature maps after sampling and deconvolution operation may result in the suppression or enhancement of the characteristic information, and make the feature distribution too large, therefore, before the fusion of the characteristic images, it is necessary to carry out the local response normalization operation (LRN) for the output features of different layers, smooth the activation values between the features, standardize the feature mapping, and unify them the formula is shown in (1):

\[ b_{i,j}^j = a_{i,j}^j / (k + a \sum_{j=\max(0,j-n/2)}^{\min(N-1,j+n/2)} (a_{i,j}^j)^2)\beta \]  

(1)

Where i represents the subscript, j represents the squared cumulative index, a represents the specific pixel value of i in the feature map, N represents the number of innermost vector columns, and k, a, n/2, and \( \beta \) represent bias, alpha, and deep_radius, respectively. belta and other hyperparameters, and then concatenate the feature maps by concat. The low convolution layer conv2_2 extracts 128 channels, the middle convolution layer conv3_3 extracts 128 channels, and the high convolution layer conv5_3 extracts 256 channels to form a multi-feature fusion map. The concat formula is shown in (2):

\[ Z_{concat} = \sum_{i=0}^{c} X_i \ast K_i + \sum_{i=0}^{c} Y_i \ast K_{i+r} \]  

(2)

In the formula, Xi and Yi respectively represent channels, and \( \ast \) represents convolution. Finally, use 1×1 convolution to change the dimension of the feature map. The overall detection process of the
feature fusion map is shown in Figure 2.

![Figure 2. Feature fusion detection model](image)

4. Experimental results and analysis

The experiments in this article will be implemented on the Tensorflow deep learning framework. Other configurations of the experimental environment are Ubuntu16.04 operating system, Intel Xeon(R) CPU E5-2603 V4@1.70 GHz processor, 9 TB Seagate hard drive, 48GB memory, GPU model is NVIDIA GeForce GTX TITAN X. Algorithm programming languages are Python and C++. After the relevant hardware is ready, install the corresponding NVIDIA driver and OpenCV and CUDA libraries to facilitate data processing and accelerated training of network models. Different from the commonly used four-step alternating training method, this article uses an end-to-end training method, momentum stochastic gradient descent (SGD) for weight optimization, the momentum value (momentum) is set to 0.9, and the weight decay factor (weight decay) is set to 0.002, the batch size of each step is set to 96, the initial learning rate is set to 0.0001, when the number of batches reaches 60000, the model training is stopped.

4.1 Model Recognition Results

In the defect detection of the valve seat of the automobile injector, there are usually small size defects such as defect, rust spot and white spot in the valve seat, and there are many kinds of defects in one picture. In order to have a clearer understanding of the valve seat defects of the injector, the method proposed in this paper firstly identifies four kinds of common defects, 5(a) is the defect identification map, 5(b) is the scratch identification map, 5(c) is the white spot, identification map, and 5(d) is the rust spot identification map. It can be observed that the defects are relatively small, but all of them can achieve good recognition results. From the picture, it can be observed that the defects are relatively small, but all can achieve a better recognition effect. In practical applications, there are many kinds of small size defects inside the valve seat. If the original algorithm model is directly used for detection, there will be missed detection. Some of the missed detection results are shown in Figure 5(e), 5(g). In order to solve the problem of missed detection caused by the multi-scale and small feature in the detection of valve seat defects of automobile fuel injectors, VGG16 feature network module will be extracted from, the original Faster R-CNN network model the features are merged, and the detection results of the algorithm model after the improved feature extraction network are shown in Figures 5(f)
and 5(h). It can be seen from the figure that the improved algorithm model can effectively solve the problem of missed detection to a certain extent. The display result is shown in Figure 3.

![Defect test chart](image)
(a) Defect test chart

![Scratch test chart](image)
(b) Scratch test chart

![White spot test chart](image)
(c) White spot test chart

![Rust spot test chart](image)
(d) Rust spot test chart

![Faster R-CNN method](image)
(e) Faster R-CNN method

![Improve method](image)
(f) Improve method

![Faster R-CNN method](image)
(g) Faster R-CNN method

![Improve method](image)
(h) Improve method

**Figure 3.** test result

**Acknowledgments**

Fund Projects: Guangxi Postgraduate Education Innovation Program (YCSW2019206); Guangxi Education Hall Project of Improving the Basic Ability of Young Teachers (KY2016YB252); Guangxi colleges and universities young and middle-aged teachers scientific research basic ability improvement project (2019KY0358);

**References**

[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2014: 580-588.

[2] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]// Advances in Neural Information Processing Systems, 2015: 1137-1146.

[3] Barret Z, Ekin D, Jonathon S, et al. Learning Data Augmentation Strategies for Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019.

[4] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv: 1409. 1556, 2014.
[5] Chen, G. Papandreou, Murphy, K. Semantic image segmentation with deep convolutional nets and fully connected. In ICLR, 2015.

[6] Kruthiventi, K. Ayush, and R.V. Babu. Deepfix: A fully convolutional neural network for predicting human eyes. arXiv preprint arXiv:1510.02927, 2015.