Research on Vehicle Parts Defect Detection Based on Deep Learning

Wang Liqun\(^a\), Wu Jiansheng\(^b\), Wu Dingjin\(^c\)

University of Science and Technology Liaoning, School of software, Anshan, China;
\(^a\)18341269842(+86), hbcdlq415@163.com, \(^b\)16642207618(+86), ssewu@163.com, 
\(^c\)15541225805(+86), 347288623@qq.com

ABSTRACT. At present, automobiles have become a common means of transportation, but with the increase of vehicles, safety issues have gradually emerged. Therefore, the assembly, manufacture and production of vehicles require systematic testing and rigorous inspection. Therefore, defect detection of vehicle parts is particularly important. Vehicle parts defect detection has evolved from manual detection of traditional classification methods to machine vision methods. In this paper, the deep learning method is used to firstly detect the defects of vehicle parts through the training of VGG16 network structure model. The accuracy rate is 94.36. Secondly, the VGG16 network structure model is improved. By introducing the inceptionv3 module, the width of the model is increased on the basis of depth, the image is better recognized with an accuracy of 95.29. However, the accuracy of the traditional HOG+SVM classification method is only 93.88, and the efficiency of both methods is higher than that of the traditional method.

1. INTRODUCTION

Image classification is a way for computer to obtain image data and imitate human biological system according to its interpretation. Image classification is of great significance to the development of today's society. Because every second around the world will produce a large number of image data, human beings can obtain valuable knowledge from these large amounts of picture data to solve a variety of problems. Defect detection of vehicle parts is an application of image classification. With the rapid development of computer technology and artificial intelligence technology, machine vision [1-2], machine learning and deep learning provide a more efficient and reliable classification method. The two most extensive methods of image classification are SVM [3] based on machine learning and CNN based on deep learning. A large number of experiments and studies have proved that SVM is suitable for learning small sample data, and can achieve good results. CNN is suitable for the training and learning of a large number of data. It can effectively extract the features of the data and classify them accurately and efficiently.

Cognex Germany has added the PatMax software vision tool to the vision sensor. This tool not only recognizes the shape of 25 different filters, but also determines the location of the filter. This tool is highly accurate, up to 99%. Domestically, visual inspection has been gradually used instead of manual inspection, which not only improves production efficiency, but also reduces the rate of missed detection during the inspection process. It has been widely used in the surface characters of automobile parts and the detection of missing parts of automobile parts.

At present, in the whole vehicle factories of some automobile companies, machine vision [4] is used to detect defects in the process of vehicle assembly, and establish a surface defect detection
simulation system based on the Siemens PLC platform to automatically detect the car. The process of the simulation system first collects the information to be detected by the camera, reads the image preprocessed result through the visual sensor, trains the processed image, establishes a relatively complete information model, and then performs the target image. The feature is extracted and matched with the information template, and the quality of the assembly process is judged according to the matching result.

However, there are still difficulties in the detection of vehicle component defects. Due to the variety of vehicle parts, the shape of the vehicle components is different. Due to the variety of vehicle component defects, the position, length and area of the vehicle component defects are diversified. Secondly, the image structure of vehicle parts is more complex, the structure of the parts themselves is not prominent, there are a lot of irrelevant factors around the image, and there are a lot of noise, which makes feature extraction more difficult.

Based on the above problems, this paper designs an improved model based on VGG16, and introduces inceptionv3 module. Defect detection of the seven most common categories of front axle left front and middle bolts, transmission shafts, parking locks and shift mechanisms, special steering tools, right headlight positioning bolts, shock absorbers and shock absorbers. Compared with the unmodified VGG16 model and the traditional machine learning HOG + SVM method, the accuracy is improved.

2. RELATED WORK

2.1 Deep Learning
Deep learning [5] mainly realizes an abstract expression of data by simulating the process of multi-layer abstract learning of human brain. Compared with the shallow structure, the model structure of deep learning has stronger ability to express data. Deep learning is deep learning network through multi-layer nonlinear units. The output data of the lower layer network is used as the input data of the higher layer network, and the effective high-order feature data is screened out layer by layer from a large number of data. Image recognition and classification are carried out by using high-order feature data containing a lot of information.

2.2 Convolutional Neural Network
The convolutional neural network [6] is a feed forward neural network whose artificial neurons can respond to a surrounding area of a part of the coverage and have excellent performance for large image processing. The convolutional neural network model can transform the image into a feature map shared by weights through a series of convolutional and pooling layers, and classify the features expressed by each image in a fully connected form, wherein the convolution layer The convolution kernel completes the function of perceptual field of view, stimulating the underlying local area to the upper level. As shown in Figure 1.

![Figure 1. Convolution neural network.](image)

2.3 VGGNET
VGGNet [7-8] is a deep convolutional neural network model developed by researchers at the Visual Geometry Group and Google DeepMind of the University of Oxford. The VGGNet network model mainly explores the relationship between the depth of the convolutional neural network and its
performance. The network model constructs a convolution neural network with a depth of 16 to 19 layers by repeatedly stacking $3 \times 3$ small convolution kernels and $2 \times 2$ maximum pooling layers.

The VGGNet model [9-10] has 5 segments of convolution, with 2 to 3 convolution layers in each segment, and a maximum pooling layer at the end of each segment to reduce the size of the feature map. The number of convolution kernels in each segment is the same. The closer to the back end of the network model, the greater the number of convolution kernels: 64-128-256-512-512. Among them, there are often multiple exactly the same $3 \times 3$ convolution layers stacked together in the network model, and the design of this network structure is very important for the overall structure of VGGNet. As shown in Figure 2, Two $3 \times 3$ convolution layers in series are equivalent to a $5 \times 5$ convolution layer, that is, one pixel will be associated with the $5 \times 5$ pixels around it, which indicates that the receptive field of the structure is $5 \times 5$. On the other hand, the effect of three $3 \times 3$ convolution layers in series is equivalent to that of a $7 \times 7$ convolution layer. In addition, three $3 \times 3$ convolution layers in series have less parameters than one $7 \times 7$ convolution layer, and only $(3 \times 3 \times 3) / (7 \times 7) = 55\%$ of the latter. Most importantly, three $3 \times 3$ convolution layers have more nonlinear transformation than one $7 \times 7$ convolution layer (the former can use cubic ReLU activation function, while the latter only needs one ReLU activation). Such a network structure design, It ensures that CNN has a stronger ability to learn features. $3 \times 3$ convolution kernel and $2 \times 2$ pooling kernel are all used in VGGNet model, which can improve the performance by deepening the network structure. Figure 3 shows the network structure diagram at all levels of the VGGNet.

![Figure 3. VGGNet network structure diagram.](image-url)
3. NETWORK STRUCTURE AND METHOD

In this paper, experiments show that the accuracy of using VGG network structure alone has been improved compared with the traditional classification methods. After studying the influence of network depth on classification accuracy, the classification accuracy is improved by properly increasing the width of the network. Affected by the design of 3 × 3 convolution kernels in VGG networks, some large convolution kernels can be replaced by a series of 3x3 convolution kernels, or even smaller.

Therefore, this paper combines the idea of inceptionV3 network model structure. InceptionV3 network has two main characteristics. The first is to introduce the method of decomposing convolution kernel to integrate a larger two-dimensional convolution into two smaller one-dimensional convolution. For example, the network of 7 × 7 convolution is divided into three 3 × 3 convolution, which produces more parameters than dividing the network of 7 × 7 convolution into 1 × 7 and 7 × 1 convolution, that is, the latter is more economical. In this paper, we divide the 3 × 3 convolution kernel in some layers of VGG network into 1 × 3 and 3 × 1 convolution (Figure 4). Its advantage is that it reduces a variety of parameters, improves the operation speed, and reduces the occurrence of over-fitting. In addition, the expressive ability of a layer of nonlinear extended model is added. The result of the resolution of this asymmetric convolution structure is more obvious than that of symmetrically dividing into several identical small convolution kernels, which can deal with more and richer spatial features and increase the diversity of features. The second advantage is the introduction of inceptionV3 (Figure 5) modules. InceptionV3 optimizes the structure of inception module. At present, inception module has three different structures: 35 × 35, 17 × 17 and 8 × 8. These inception module appear only at the back of the network, and the front of the network is still a common convolutional layer.

Figure 4. The 3 x 3 convolution kernel is divided into 1 x 3 and 3 x 1 convolutions.

Therefore, inspired by the structure idea of inceptionV3 network model, this paper combines the structure of inceptionV3 model with the structure of VGG16 network model, and constructs a new 10-layer classification network to realize the defect detection of vehicle parts. Compared with the original VGG16 network structure, the effect is better. Table 1 shows the structure and parameter settings of the model in this paper.
3.1 Model structure advantages

VGG16 network structure can effectively solve the defect detection and classification of vehicle parts. In this paper, InceptionV3 modules are introduced to broaden the width of the network on the premise of ensuring the depth of the network. In the case of determining the input dimension, it is necessary to consider the complexity of the model and ensure that the classification information is not lost, so the model designed in this paper belongs to the structure from wide to narrow, and after using pooling and other methods, the feature map of the final output is the size of $7 \times 7$.

The use of convolution kernel. Small convolution kernels $(3 \times 3)$ are used instead of large convolution kernels such as $7 \times 7$ and $5 \times 5$. That is, three $3 \times 3$ small convolution kernels are equivalent to a $7 \times 7$ large convolution kernel, and two $3 \times 3$ small convolution kernels are equivalent to a $5 \times 5$ large convolution kernel. The advantage of this is that while keeping the size of the receptive field unchanged, the depth of the model is deepened and the number of parameters is reduced.

Solve the problem of covariate shift. The method of Batch Normalization (BN) is adopted, and this method is applied to Conv1, Conv2 and Fully Connected of the model, so that the training speed of the model is accelerated and the generalization ability is enhanced.

In order to avoid the over-fitting phenomenon of data, the method of spatial dropout is used in this paper. Dropout is the simplest neural network regularization method. The principle is very simple: arbitrarily discard the input in the neural network layer, which can be an input variable in the data sample or an activation from the previous layer. It can simulate neural networks with a large number of different network structures, and in turn makes the nodes in the network more robust.

| Name     | Input          | Output         | Remarks   |
|----------|----------------|----------------|-----------|
| Conv1    | 224×224×3      | 112×112×64     | ReLu, BN  |
| Conv2    | 112×112×64     | 56×56×128      | ReLu, BN  |
| Pool1    | 56×56×128      | 28×28×128      | max pooling |
| Inception2 | 28×28×128     | 28×28×320      | BN        |
| Pool2    | 28×28×320      | 14×14×320      | max pooling |
| Dropout2 | 14×14×320      | 14×14×320      | spatial dropout |
| Inception3 | 14×14×320     | 14×14×192      | BN        |
| Pool3    | 14×14×192      | 7×7×192        | max pooling |
| Dropout3 | 7×7×192        | 7×7×192        | spatial   |
The structural parameters of the inception in the experiment are as follows: The specific structure of inception module is as follows: Because the padding method taken is same, the module does not change the input W×H, but may change its channel number. The inception layer parameters used in Table 1 are as follows: The output channels of the four branches of the first inceptionV2 from left to right are 64, 96, 96, 64, respectively, and the number of channels output by the two branches (that is, 1×1 conv) in the middle position is 64. The output channels of the four branches of the second inceptionV3 from left to right are 32, 64, 64, and 32, respectively, with a total of 192, and the number of channels output by the two branches (that is, 1×1 conv) in the middle position is 96.

3.2 Relu activation function
This paper uses relu as the activation function. When the input value is less than zero, the output value is zero. If the input is greater than or equal to zero, the output is equal to the input. When the input value is positive, the derivative is 1. Its mathematical expression is as follows:

\[ f(x) = \max(0, x) \]

The advantage of using the relu activation function is that when the network is trained, the convergence rate of the network is much faster than using the sigmoid function or the tanh function. Compared with the sigmoid function or the tanh function, the relu function only needs a threshold to get the activation value, and does not need to do some complex operations. At the same time, the use of relu function will not have a significant impact on the generalization accuracy of the model.

3.3 Softmax loss function
Softmax is the normalization of multiple values obtained by the neural network, so that the resulting values are between [0, 1]. The results show in the form of probability that the greater the probability of a category, the more likely it is to classify the sample into that category. The softmax function is recorded as:

\[ f_j(z) = \frac{e^{z_j}}{\sum_i e^{z_i}} \]

The cross entropy loss is obtained by taking the negative logarithm of the softmax function.

\[ L_i = -\log \left( \frac{e^{z_{y_i}}}{\sum_j e^{z_j}} \right) = -z_{y_i} + \log \sum_j e^{z_j} \]

\( z_j \) represents the score on category \( j \), and \( y_i \) represents the real category. The range of loss values is \([+\infty, 1]\), and the higher the score in the real category, the lower the loss. The loss of Softmax total samples is as follows:

\[ L(X,Y) = -\frac{1}{N} \sum_i \sum_j I\{j = y^{(i)}\} \log (p_{i,j}) \]

Where \( p_{i,j} = \frac{\exp(z_{i,j})}{\sum_j \exp(z_{i,j})} \) represents the probability that the sample \( i \) prediction category is \( j \). \( z_j \) is a linear combination of the output \( a^{L-1}_j \) of the previous layer: \( z_j = w_j a^{L-1}_j + b_j \).
Partial derivative: \[
\frac{\partial L}{\partial w_j} = -y_j (1 - p_j) a_{j-1}^{(j)}, \quad \frac{\partial L}{\partial b_j} = p_j - y_j.
\]

4. EXPERIMENT AND EVALUATION

4.1 Experimental environment
The experimental environment of this paper is carried out in Python3.5 environment, using Python language, deep learning tensorflow framework to complete feature extraction, parameter setting and network training. Experiment and Evaluation.

4.2 Experimental data.
This paper mainly uses VGG16 network and combines inceptionv3 model to realize the classification of vehicle parts. The experimental samples are divided into training set and test set. Each of the seven categories of vehicle defect detection is divided into positive and negative samples. In the process of training samples, to prevent the data form being single and produce over-fitting, at the same time to enhance the performance ability of sample defect features to adapt to different shooting angles, using the way of data enhancement to expand the training sample, the way of data enhancement is as follows.

Image rotation: The usual method of image rotation is to rotate the center of the image clockwise. All the pixels in the image are rotated at the same angle, that is, the pixels in the image are rotated 270 degrees, and the size of the rotated image does not change with the original image. When the plane rotates 270 degrees around the center of the circle, the coordinates of the pixels on the plane change. As shown in Figure 6.

The coordinates before rotation are:
\[
x = r \cos \alpha \\
y = r \sin \alpha
\]

The rotation transformation formula of the coordinates is as follows, where \( \theta = 180^\circ \).
\[
x' = r \cos(\alpha + \theta) = r \cos \alpha \cos \theta - r \sin \alpha \sin \theta = x \cos \theta - y \sin \theta \\
y' = r \sin(\alpha + \theta) = r \sin \alpha \cos \theta + r \cos \alpha \sin \theta = x \sin \theta + y \sin \theta
\]

The matrix form becomes:
\[
\begin{bmatrix}
x' \\
y'
\end{bmatrix} =
\begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix}
\]

Figure 6. Image before (left) and after rotation (right).

Image mirror transformation: image mirror transformation is divided into two methods: horizontal mirror and vertical mirror. No matter what kind of specular transformation, the height and width of the image will not change. As shown in Figure 7.

Horizontal mirror transformation: that is, taking the vertical central axis of the original image as the center, the left and right parts of the image are transformed symmetrically.
Figure 7. Horizontal mirror transformation.

Vertical mirror transformation: taking the horizontal central axis of the original image as the center, the upper and lower parts of the image are transformed symmetrically. As shown in Figure 8.

Figure 8. Vertical mirror transformation.

Training the original positive sample has 24556 images, negative sample has 17057 images, the data is enhanced, through rotation, vertical transformation and other series of operations, increase the positive sample 14600 images, increase the negative sample 18203 images. In the testing phase, vehicle defect detection can detect both positive and negative samples. As shown in Table 2.

| Sample processing                      | The amount of data |
|----------------------------------------|--------------------|
| positive sample                        | 24556              |
| negative sample                        | 17057              |
| Positive sample after data enhancement | 47520              |
| negative sample after data enhancement | 35260              |

4.3 Analysis of experimental results

The experiment in this paper is divided into three parts, which use the traditional HOG+SVM method, the unimproved VGG16 network model and the improved VGG16 network model to detect the defects of vehicle parts. The experimental results are analyzed by the accuracy of recognition and the loss value of detection and recognition.

Firstly, the image is extracted into multi-dimensional features in the way of HOG, and the result of feature extraction is shown in Figure 9. Then the SVM algorithm is used to classify the image. The penalty coefficient in SVM is set to 1.0, the squared_hinge loss function is used, the dual optimization algorithm is adopted, and the maximum number of iterations is set to 1000.

Figure 9. HOG Feature Extraction Result Map.

The parameters of the improved VGG16 network model are as follows: the learning rate is 0.001. Although the more iterations, the better the effect, but the greater the amount of computation, so the
number of iterations adopted in this paper is 500. The epoch is set to 10. The improved VGG16 network model is trained iteratively to obtain high-level features that represent attribute categories and information, and the final recognition effect of the image is obtained by linear training of the classifiers. The accuracy of recognition is used as an important index to judge the mail box of the algorithm. Figure 10 and Figure 11 show the accuracy and loss values of the improved VGG16 network model training set obtained by tensorboard visualization.

Figure 10. Accuracy of improved VGG16 training set.

Figure 11. Loss of improved VGG16 training set.

Figure 12 and Figure 13 show the accuracy and loss values of the improved VGG16 network model test set visualized by tensorboard.

Figure 12. Accuracy of unimproved VGG16 test set.

Figure 13. Loss of unimproved VGG16 test set.
As can be seen from Table 3, the accuracy of the improved VGG network model on the training set is 96.68 and the loss value is 0.089. The accuracy of the test set is 95.29 and the loss value is 0.147. The experimental results show that the improved VGG network model has a remarkable effect on the classification of vehicle parts defect detection.

Table 3. Amount of data before and after data enhancement

|       | Accuracy | Loss  |
|-------|----------|-------|
| Train | 96.68    | 0.089 |
| Test  | 95.29    | 0.147 |

Table 4 shows the evaluation and comparison of test data sets in the three methods. As you can see from Table 3, the accuracy of defect detection of vehicle parts by traditional HOG+SVM method is 93.88%, and the accuracy of defect detection and recognition of vehicle parts by using unimproved VGG16 network model is 94.36%. The accuracy of defect detection and recognition of vehicle parts based on the improved VGG16 network model is 95.29%. This shows that the depth learning method has higher recognition rate and robustness to vehicle parts defect detection than the traditional classification method, and improves the detection efficiency and accuracy of the whole network model. At the same time, the improved VGG16 network extraction method is easy to implement, the training method is simple, and the computational complexity is low. It is proved that the introduction of inceptionv3 model broadens the width of the network on the basis of ensuring the depth of the network. So that the model can be more specific to extract the information features of the image, so as to achieve a higher and faster recognition effect.

Table 4. Amount of data before and after data enhancement

| Method     | Accuracy | Loss  |
|------------|----------|-------|
| HOG+SVM    | 93.88    | -     |
| VGG16      | 94.36    | 0.198 |
| Improved VGG16 | 95.29 | 0.147 |

5. SUMMARY

This paper mainly puts forward a method of using depth learning to solve the problem of defect detection of vehicle parts. Using VGG16 network model structure, the defect detection method is tested and a high accuracy of defect detection is obtained. After that, the VGG16 network structure model is improved, and the inceptionv3 module is introduced to increase the width of the model on the basis of depth, and the accuracy of the model is improved from 94.36% to 95.29%, so that the image can be better.

Although the current vehicle parts defect detection methods have made preliminary development and achieved certain results, but there are shortcomings in many places, the scientific basis and theoretical methods also need to be innovated and improved. It is believed that in the future, the defect detection of vehicle parts based on deep learning will play a positive role and make a great contribution to the target detection. At present, most of the defect detection of vehicle parts is to convert color images into grayscale images and then detect them. In the future, it is worth studying whether we can deal with the detection of vehicle parts because of lighting, occlusion and unclear images. It is believed that these problems can be solved by deep learning.

REFERENCES

[1] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. international conference on learning representations, 2015.
[2] Bell S, Upchurch P, Snavely N, et al. Material recognition in the wild with the Materials in Context Database[J]. computer vision and pattern recognition, 2015: 3479-3487.
[3] Zhang F Y, Liu R J. Study on the Parts Surface Defect Detection Method Based on Modified SVM Algorithm[J]. Applied Mechanics & Materials, 2014, 541-542:1447-1451.

[4] Li Y, Zhao W, Pan J. Deformable Patterned Fabric Defect Detection With Fisher Criterion-Based Deep Learning[J]. IEEE Transactions on Automation Science & Engineering, 2017, 14(2):1256-1264.

[5] Wang X B, Jie L I, Yao M H, et al. Solar Cells Surface Defects Detection Based on Deep Learning[J]. Pattern Recognition & Artificial Intelligence, 2014, 27(6):517-523.

[6] Zhu Q, Ai X. The Defect Detection Algorithm for Tire X-Ray Images Based on Deep Learning[C]/ 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). 2018.

[7] Qu Z, Shen J, Li R, et al. PartsNet: A Unified Deep Network for Automotive Engine Precision Parts Defect Detection[C]/ the 2018 2nd International Conference. 2018.

[8] Chen Y. Research on Defect Detection Based on Machine Vision for Bubbles in Toughened Glass Insulator[J]. Process Automation Instrumentation, 2010, 31(7):51-53,56. I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[9] Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning[J]. 2017.

[10] Bayar B, Stamm M C. A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer[J]. 2016.