A New Method in Wheel Hub Surface Defect Detection: Object Detection Algorithm Based on Deep Learning

Kai Han,Muyi Sun, Xiaoguang Zhou*, Guanhong Zhang,Hao Dang,Zhicai Liu
School of Automation
Beijing University of Posts and Telecommunications
Beijing, China
bupt2012211769@bupt.edu.cn,

***Abstract*—In wheel hub surface defect detection, a unified image background is required. However, it is a challenging task because of the various categories of wheel hubs, and the complicated image background of the defect areas caused by the collection of the images with the defect areas in a narrow field of vision. Compared to the traditional method, the deep learning algorithm is more robust, which doesn't need the unified image background. We use Faster-RCNN with ResNet-101 as the object detection algorithm. And our related experiments show that our deep learning method is able to detect the scratches and points on the wheel hub in an image with a complicated background, as shown in Figure5. Furthermore, the model can detect defects on any part of the wheel hub of various types, and obtain the position and the class of the defective area. Particularly, the method achieves 86.3% mAP on our own data set.**

Keywords—wheel hub; object detection; deep learning; surface defect area

I. INTRODUCTION

Defect detection is a very important part of the wheel production line. The traditional wheel defect recognition mainly depends on human vision, and in the case of long hours of work, the efficiency and accuracy of defect recognition will decline. Therefore, a stable, highly accurate automated detection algorithm is needed.

For defective region detection method based on machine vision, firstly, the feature extraction is carried out by histogram statistic, wavelet transform, Fourier transform and so on, and then the threshold method, decision tree, support vector machine is applied for classified. The limitation of the above method is that the unified image background is needed, that is, the grayscale value of the image does not change greatly. However, due to the large size of the hub, and limited by the receptive field, the background of the defect area is very complex. So we need a more robust algorithm.

In the recent days, deep neural networks achieved excellent performance in classification task, and many algorithms in computer vision area focus on deep convolutional networks. Meanwhile, traditional image processing algorithms and shallow machine learning algorithms can hardly satisfy our needs in many popular computer vision task, such as semantic segmentation, object detection, and transfer learning. Generally speaking, we treat deep neural networks as an end-to-end combination of a feature extractor and a classifier in traditional feature learning theory, so it is spontaneous for us to transfer deep convolutional network from classification task to more complex tasks, such as defect detection task. Based on this, we have applied Faster-RCNN [1], which achieves state-of-the-art performance in object detection task, in the wheel surface defect detection, and performs very well. In addition, the training of the deep learning model requires a lot of images, so we have established a database consisting of defect area images, the relevant content will be described in detail in the third session.

II. RELATED WORK

According to the domestic and international investigation, we have not found the research literature on surface defect detection of the wheel hub. However, relevant experts and scholars have carried out the defect detection and other related researches on the material surface based on machine vision [2][3]. For example, Mean filter and dynamic threshold segmentation is applied to detect the large defect features, Gaussian filter and global threshold segmentation is used to detect the small defect features. Besides, the morphological processing of the defect features such as open operation and close operation is employed to improve the detection accuracy of the location and size of the defects.

With the development of deep learning, the theory based on CNN is widely applied in the field of object detection. Object detection, a task that carries out object classification and object location on an image at the same time, are based on sliding window detector or traditional image processing methods. R-CNN [4] proposed in 2014, which was a combination of CNNs, object detection task, and regression methods, achieved an important breakthrough in object detection area. In following years, SPP-Net [5] introduced Spatial Pyramid Pooling Layer in object detection task,

978-1-5386-2602-3/17/$31.00/ ©2017 IEEE

which improved detection accuracy by concatenating multi-scale semantic information of feature maps. Then Fast-RCNN [6] and Faster-RCNN developed from SPP-Net achieved bold-alteration performance not only in accuracy but also in speed, and complete end-to-end network becomes the mainstream of object detection task. The first Faster R-CNN model is based on VGG16 [7], and now the ResNet101-based [8] model has improved performance to application-level. For example, in tasks such as pedestrian detection or automatic drive, models can be applied to open scenes. Referring to object detection models, our wheel hub detection model treats scratches and flaw points as a kind of object, using deep convolutional neural networks to make bounding box regression to these object and locate their position. We achieved a remarkable performance in this task.

III. DATA SET

A. Data Cleaning

Aiming at setting up a standard object detection dataset of surface defect area, we cleaned the image of the local area of the hub, and abandoned the image that did not contain the defect areas, and got 110 images with points and 60 with scratches.

B. Data Augmentation

In the original image, the defect area only occupies a small part. Under these circumstances, the model is hard to learn the features of the defect areas because of the extreme foreground-background class imbalance. To solve the problems, we use a 4x4 grid to cut the original image to 16 parts, and leave the part which contains the defect areas. In this way, the receptive field of the defect area is blown up on the fixed image.

Then the data augmentation is applied for the fixed images, including the flip horizontal, Flip vertical, Gaussian noise and rotating. And we totally gained 2000 images as our data set. At last, according to the proportion of 8:2, the dataset is divided into training set and testing set.

C. Label Encoding

Training object detection model needs a label file, including the four coordinates of the rectangular boxes called bounding box containing defect areas, as well as the class of each defect area, which is saved to a XML file. The detail of our dataset is similar to PASCAL VOC 2012.

IV. NETWORK ARCHITECTURE

A. Feature Extraction

In 2012, Krizhevsky et al. rekindled interest in CNNs by showing higher image classification accuracy on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Their outstanding performance resulted from training an 8-layer CNN on 1.2 million images with their label which called AlexNet [9]. In next years, thanks to the evolution of training strategy and network architecture, and more challenging classification dataset such as ImageNet, a series of better and deeper CNNs was developed, including VGG [7], GoogLeNet [10], ResNet [8], etc. After ResNet was developed in 2016, models with residual learning and identity mapping theory achieved state-of-the-art performance on almost every influential image classification dataset. More and more people build ResNet, which also can be called a deep feature extractor, as their baseline in semantic segmentation and object detection tasks. To pursue better performance in defect detection task, our work developed from shallow machine learning method to model based on deep convolutional networks that proposed in this paper, similar to the development in the field.

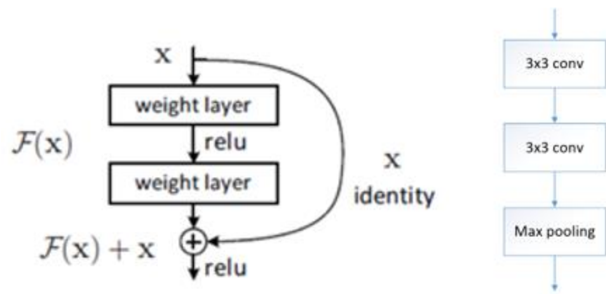

Figure1 Residual(left) and VGG(right) building block

We finally choose ResNet101 and VGG16 as the feature extractions in the network. And the performance in the defect detection of ResNet101 and VGG16 is analyzed by our experiments, as is shown in TABLEI.

The building blocks of ResNet101 and VGG16 are shown in Figure1. The difference between the residual building block and the VGG architecture is the shortcut connection in the forward propagation. In this way, the information of previous layers is retained to a certain extent. Because of this, the vanishing gradient problem is avoided and the very deep CNN architecture is possible.

B. Meta-architecture:Faster-RCNN

Faster R-CNN, is composed of two modules. The first module is a deep fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector which uses the proposed regions. The architecture of Faster-RCNN is shown in Figure2 [11]. Besides, this network uses the recently popular terminology of neural networks with the attention mechanisms, the RPN module tells the Fast RCNN module where to look.

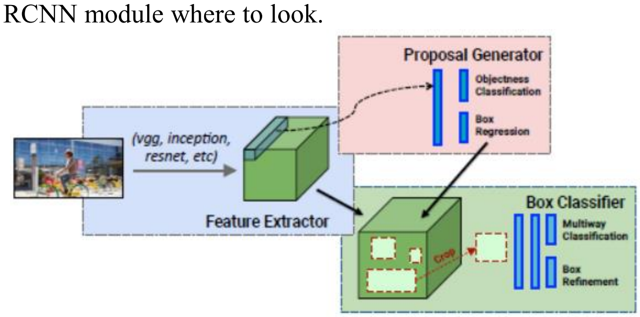

Figure2 Faster-RCNN. Top: RPN. Down: Fast-RCNN. Left: Feature extraction

1) Region Proposal Networks

RPN---Region Proposal Networks, is the most important part of Faster-RCNN, which takes an image as input and outputs a set of rectangular object proposals, each with a predictive score. The process is modeled by a fully convolutional network which shares the computation with the Fast R-CNN object detection network by using a set of convolutional layers. In our experiments, we investigate the ResNet model to generate region of interests by sliding a small network over the last convolutional feature map output. Each sliding window is mapped to a lower-dimension feature that the number of the dimension is equal to the number of the last layer's feature map. Then this lower-dimension feature is fed into a regression layer and classification layer. One for bounding box regression and another for object classification. Finally, these coarse results are used in Fast R-CNN as region proposals.

In the RPN process, there is an important trick, we call it "anchor". At each sliding-window location, there are k possible proposals. An anchor is centered at the sliding window in question, and is associated with a scale and aspect ratio. Generally, the model uses 3 scales and 3 aspect ratios, so k usually equals to 9.

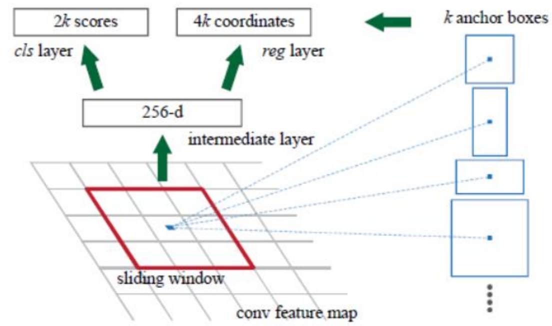

Figure3 Region Proposal Network

2) Fast R-CNN

Fast R-CNN is the basic object detection architecture. RPN gives region of interests to the final chain of the convolution and full-connected layers. The main layer of this part is the ROI pooling layer, which get the idea from the SPP-net. The end-to-end network architecture prompts the training procedure, which can update all network layer with the iteration. The output of the Fast R-CNN regression layer and the classification layer is the final results of the object detection.

TABLEI The Results of Our Algorithm on the Dataset

Method	Backbone	Steps	mAP	AP[point]	AP[scratch]	Recall
Faster-RCNN	VGG-16	30k	64.25%	54.12%	74.39%	75%
Faster-RCNN	ResNet-101	30k	81.82%	87.19%	76.45%	85%
Faster-RCNN	ResNet-101	60k	86.31%	89.38%	83.24%	90%

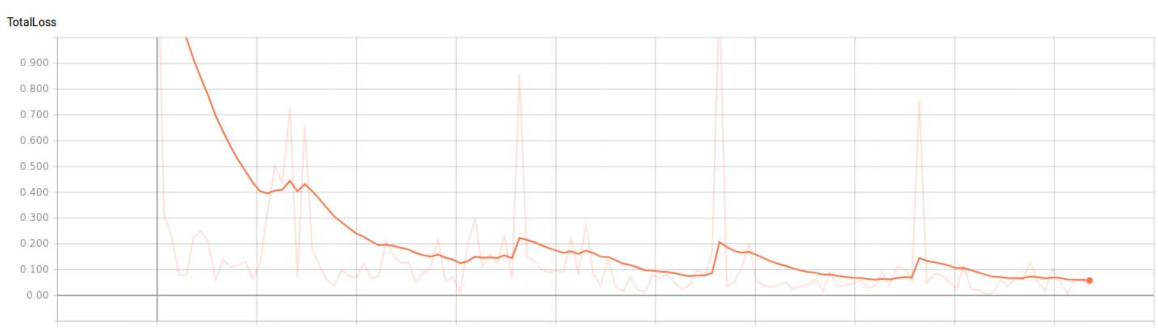

Figure4 Training Loss

V. EXPERIMENTS

A. *Pre-training*

The previous work shows that using a pre-trained model can not only accelerate the convergence of the model, but also improve the detection accuracy. As a result, ResNet-101 is applied as a meta-architecture of our model, which is pre-trained on ImageNet, the biggest image classification database in the world.

B. *Joint Training*

In fact, there are two tasks in defect detection. One is the location of the defect areas, which is actually a regression task called bounding box regression; another is the softmax classification of the above-mentioned defect areas.

To achieve joint training of the two tasks in a unified model, we combined the loss functions of regression and classification tasks.

C. *Hyper Parameters*

We initialized the weights of the fully connected layers used for bounding box regression and softmax classification from zero-mean Gaussian distributions with standard deviations 0.01 and 0.001. Besides, biases are all initialized to 0.001. In addition, stochastic Gradient Descent optimization with a momentum of 0.9 and a mini-batch size of 16 are employed in our experiments. And the learning rate is set to 0.003 at the beginning, divided by 10% every 20 thousand steps. Moreover, we applied a weight decay of 5e-4 to all layers and a batch normalization layer [10] to every convolution layers.

D. Results and Analysis

As is shown in Figure4, the loss decreases with the training steps. And the model is converged at about 50 thousand step. Note that the light-colored line presents the true value of the loss, while the dark-colored line is the smoothed value.

In addition, the results on the testing set are shown in TABLEI. As we can see, the capability of ResNet101 is much better than VGG16, which is applied as a feature extraction on all indexes. Particularly, the mAP of ResNet101 architecture surpasses the one of VGG16 by nearly 20%, which is a huge gap. Besides, the value of training steps also has crucial influence to the experiment results. We can see that the accuracy of the model with 30k training steps is lower compared with the model with 60k training steps, especially the average precision of scratches. The reason is that 30k training steps are not enough to learn the information of the defect areas, that is under-fitting.

The detection examples are shown in Figure5. As we can see, although the background of scratches or the points is so complex, our method is capable of carrying out accurate location and detecting what the defect area is. Note that the green bounding box means the point, while the blue one is the scratch. And the percentage on the bounding box indicates the probability of the defect area.

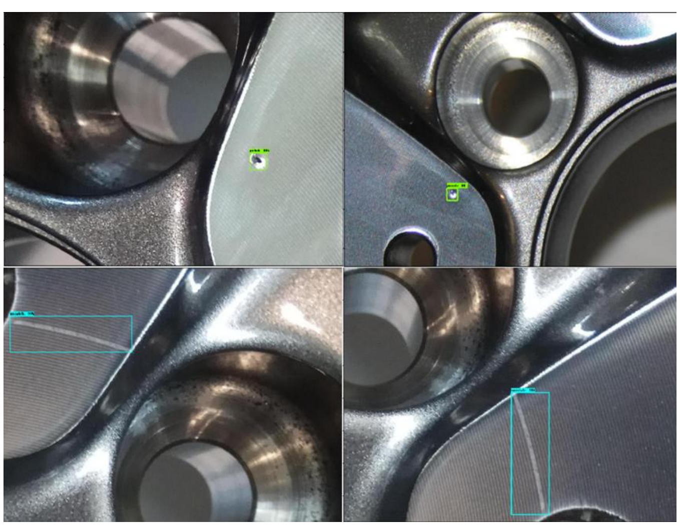

Figure5 Detection Examples

VI. CONCLUSION

In this paper, a new method is proposed for wheel hub surface defect detection, which is based on deep learning algorithm. And the challenge of detecting defect area in complicated background is conquered. We demonstrate the network architecture and the training details, then show the results of the model on our own data set. In our experiments, our method performs very well in defect area detection and achieves 90% Recall on the testing set. Our algorithm provides a practical solution for automated inspection of wheel hub production line.

ACKNOWLEDGMENT

This research is sponsored by the Fundamental Research Funds for the Central Universities (No. 2017RC27) and the China National Common Weal Industrial Special Scientific Research Funds for Grain Industry (No. 201413006).

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015.

[2] CHANG Hong-jie, SUN Ju-yun, YUE Yan-fang, YANG Guang, "Application of Machine Vision in Detection of Wood Surface Defects," School of Mechanical Engineering, Hebei University of Science and Technology, Hebei Shijiazhuang 050018, China

[3] LUO Jing, DONG Tingting, SONG Dan, XIU Chunbo, "A Review on Surface Defect Detection," (1. Key Laboratory of Advanced Electrical Engineering and Energy Technology, Tianjin Polytechnic University, Tianjin 300387, China 2. School of Electrical Engineering and Automation, Tianjin Polytechnic University, Tianjin 300387, China)

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In ECCV, 2014.

[6] R. Girshick. Fast R-CNN. arXiv:1504.08083, 2015.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.

[9] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *NIPS*, pp. 1106–1114, 2012.

[10] Christian Szegedy1, Wei Liu2, Yangqing Jia1. Going Deeper with Convolutions. In CVPR 2014.

[11] Jonathan Huang, Vivek Rathod, Chen Sun. Speed/accuracy trade-offs for modern convolutional object detectors. In CVPR 2017.