

Detection and Segmentation of Defects in Industrial CT Images Based on Mask R-CNN



Jun-Nian Gou^{1,2*}, Xiao-Yuan Wu¹, Li Liu¹

¹ Department of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou, China
junnian@mail.lzjtu.cn, DL_hdljd@163.com, 290354619@qq.com

² Key Laboratory of Optoelectronic Technology and Intelligent Control Ministry of Education, Lanzhou Jiaotong University, Lanzhou, China
junnian@mail.lzjtu.cn

Received 25 September 2019; Revised 15 January 2020; Accepted 1 March 2020

Abstract. For defect detection and segmentation of industrial computed tomography (CT) images, existing methods regard them as two separate tasks, that is, the detection and the segmentation are not integrated. This paper improves mask regions with convolutional neural networks (Mask R-CNN), and the improved network is used to complete the detection and segmentation of industrial CT image defects at the same time. In this study, the three most common defects in industrial CT images are cracks, bubbles, and slags, which are taken as the research objects. The experimental results showed that the network can achieve a detection mAP value of 0.98 and a segmentation accuracy of 0.96 for three typical CT image defects (12×12 pixels $\sim 35 \times 35$ pixels). The method proposed in this paper has high accuracy, in both the defect detection and segmentation of industrial CT images. At the same time, the algorithm also has good robustness and generalization ability. It can be applied to the detection and segmentation of various defects and has great practical significance.

Keywords: defect detection, defect segmentation, industrial CT image, mask R-CNN

1 Introduction

The types of workpiece defects mainly include bubbles, slags, and cracks, which are mostly in the form of computed tomography (CT) [1]. For a long time, traditional detection algorithms and segmentation algorithms based on image features have played a leading role in the industrial field.

For the detection of workpiece defects, prior literature [2] has proposed an adaptive method, which effectively improved the automation degree of defect detection, but the establishment of a defect database is difficult. Another study [3] combined the Firefly algorithm with a neural network to detect industrial CT image defects. The algorithm needs to artificially design the features of the defect, it has strong subjectivity, and it is easy to fall into the local optimum. Since the concept of “Deep Learning” [4] was put forward and convolutional neural network (CNN) [5] is rising, using regions with CNN (R-CNN) [6] and its improved networks, like Fast R-CNN [7] and Faster R-CNN [8], have become hot research directions for the detection of workpiece defects. Different from the above methods of artificial design features, these networks can automatically extract high-quality features of images, and thus they have higher detection accuracy. Faster R-CNN is favored by scholars because it has better performance. For example, previous works [9-11] and [12] studied a workpiece defect detection algorithm based on Faster R-CNN. These algorithms greatly improve the intelligent degree of defect detection, but their common disadvantage is that the rate of missing detection of weak defects is very high, and the location area is larger than the size of the defect.

* Corresponding Author

In the field of workpiece defect segmentation, the traditional image segmentation algorithm is widely used because of its advantages, such as simplicity, speed, and good adaptability. For example, in literature [13], a mathematical morphology and threshold method were combined to segment the defect area located by Faster R-CNN. In literature [14], a defect segmentation method combining the maximum correlation criterion and the threshold method was proposed. Due to the use of low-level features for defect segmentation, the accuracy of these algorithms is not high. At present, many scholars are committed to using depth network to achieve defect pixel to pixel segmentation. In literature [15], a fully convolutional networks (FCN) [16] was used for the segmentation of surface defects. A similar method was applied to the segmentation of aeroengine blade defects in another study [17]. Compared with the traditional image segmentation algorithm, FCN has made great progress in accuracy, but the problem of defect edge information loss is still very serious.

Currently, the main research direction is to use a depth network to detect or segment workpiece defects, but the existing algorithms regard detection and segmentation as mutually independent, and thus the accuracy of these algorithms is still not ideal. To solve this problem, this paper studies the application of mask regions with CNN (Mask R-CNN) [18], which is an instance segmentation network, in industrial CT defect images. Due to the characteristics (small size, fuzzy edge, and low gray contrast) of the research object, to get accurate experimental results, the data set was preprocessed, and the structure and the parameters of Mask R-CNN were improved. The improvements are as follows: a. improved the FPN [19] structure to enhance the role of low-level features in the feature map; b. added dropout layer. When the model was trained, dropout layer was introduced into the feature extraction network to prevent over fitting; c. the anchor size was modified; d. the learning rate was adjusted. The improved model has a high practical value. It not only realizes the integration of defect identification, location, and segmentation but also achieves unprecedented achievements in accuracy.

2 The Mask R-CNN Architecture

Mask R-CNN can be seen as a combination of Faster R-CNN [8] and FCN [16]. They are briefly introduced below.

2.1 Faster R-CNN

The structure of Faster R-CNN is shown in Fig. 1. Faster R-CNN is mainly composed of a region proposal network (RPN) for generating suggestion boxes and Fast R-CNN for detection. In the detection stage, the basic network visual geometry group net (VGG Net) [20] or ZF Net [21] (a kind of CNN model) is used to extract image features, region of interest pooling (RoI pooling) is used to obtain feature maps with a fixed size of each RoI, and finally, the category and regression box of each RoI is output.

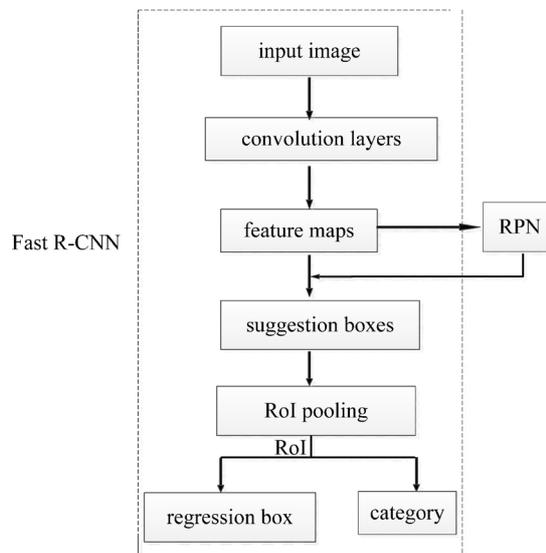


Fig. 1. Faster R-CNN architecture

Faster R-CNN defines the multi-task loss function for each RoI as follows :

$$L = L_{cls} + L_{box}, \quad (1)$$

where L_{cls} represents the classification loss, which is calculated by multi-classification cross-entropy of softmax; L_{box} represents regression box loss, which is mainly affected by Intersection-over-Union (IoU). IoU is the overlap rate of the suggestion box and ground truth (GT), that is, the ratio of the intersection to the union. The suggestion box can be represented as A, the GT as B, and then the IoU can be expressed as

$$\text{IoU} = (A \cap B) / (A \cup B). \quad (2)$$

2.2 FCN

Unlike the classical classification network CNN, which uses fully connected layers to obtain a fixed-length feature vector after the convolution layers, FCN replaces the fully connected layers with convolutional layers to form a fully convolutional network. The input of the FCN can be an image of any size, and the feature maps of the convolutional layer output are up-sampled by deconvolution operation to restore it to the same size as the input image, thus achieving alignment of pixels. At the same time, to realize pixel-to-pixel segmentation, the spatial information of the original image is preserved [16].

Taking VGG Net as an example, the structure of CNN and FCN are compared in Fig. 2 and Fig. 3.

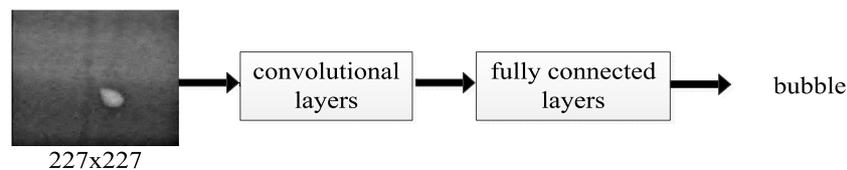


Fig. 2. CNN architecture

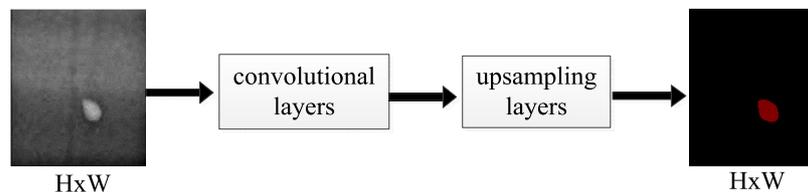


Fig. 3. The FCN architecture

2.3 Mask R-CNN

Mask R-CNN abandons the feature extraction network VGG Net or ZF Net of Faster R-CNN and uses an enhanced network (Residual Network (ResNet) [22] and Feature Pyramid Networks (FPN) [19]) to extract features.

2.3.1 ResNet

The increase in the number of convolutional layers helps to extract more advanced and richer features, thereby improving the performance of the network. However, when the number of layers is enough, two problems may arise: one is that the performance of the network decreases gradually; the other is that the gradient disappears when the network is trained. To solve these two problems, He Kai-Ming and others [22] proposed ResNet.

In addition to the normal convolutional layers output, ResNet increases a branch that connects the input directly to the output, thus achieving an identity mapping of the redundancy layers, which enables the network to achieve the best performance and solve the problem of gradient disappearance. ResNet is composed of several residual blocks, and the structure of each residual block is shown in Fig. 4.

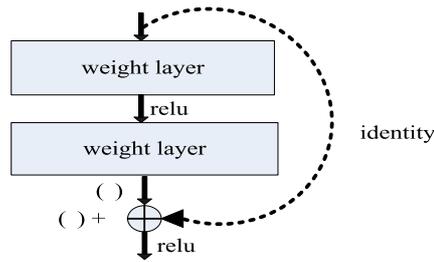


Fig. 4. The Residual block architecture

2.3.2 FPN

In the field of target detection, most networks use the feature maps of the last convolutional layer output (i.e. the top-level feature map) to detect targets. The reason for this is that a high-level feature contains rich semantic information, but one drawback is that it has low resolution, and thus it is difficult to detect small targets. However, a low-level feature has high resolution and low semantic information. Tsung-Yi Lin, Piotr Dollár, and others [19] have proposed FPN and proved that the network significantly improves the accuracy of target detection and segmentation.

FPN combines high-level features with low-level features to obtain multi-scale feature maps with both high-resolution and rich semantic information to improve the overall performance of the network. For ease of understanding, the structure of feature combination with single feature map and FPN is shown in Fig. 5.

The method of feature combination using a single feature map is shown in Fig. 5(a). The method of feature combination using a FPN is shown in Fig. 5(b), where it adopts a top-down method, and the specific operation is shown in the dashed box. A high-level feature is restored to the same size as a low-level feature by double up-sampling ($2 \times$ up), while low-level feature is adjusted to the same channel as a high-level feature by 1×1 convolution operation (1×1 conv). Then, the high-level feature and low-level feature do the addition of pixels to obtain the enhanced feature map on each scale. In Fig. 5, feature maps are indicated by thicker outlines to denote semantically stronger features.

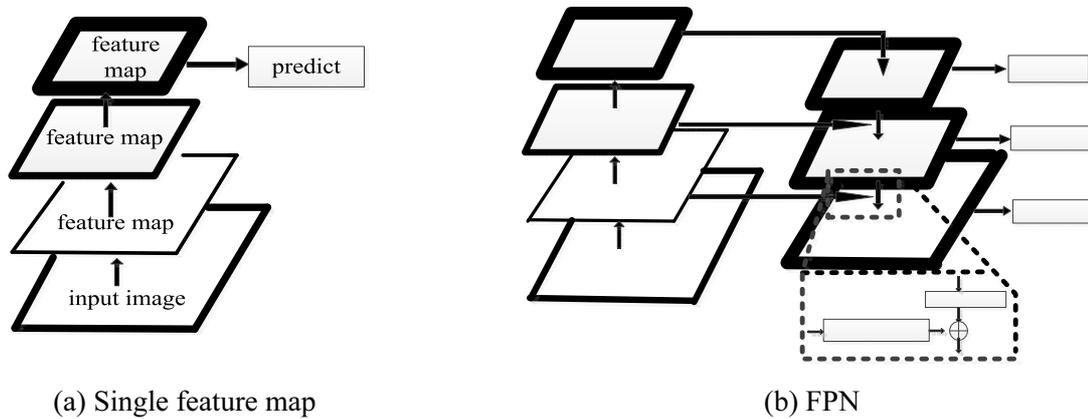


Fig. 5. Feature combination method

2.3.3 Mask R-CNN

The structure of Mask R-CNN is shown in Fig. 6.

(1) The ResNet101-FPN in Fig. 6 is the basic network of Mask R-CNN, and its specific structure is shown in Fig. 7, where C1-C5 respectively represent the feature map of Resnet-101 in five feature extraction stages, 1×1 and 3×3 represent the convolution operation, and $2 \times$ up represents double up-sampling.

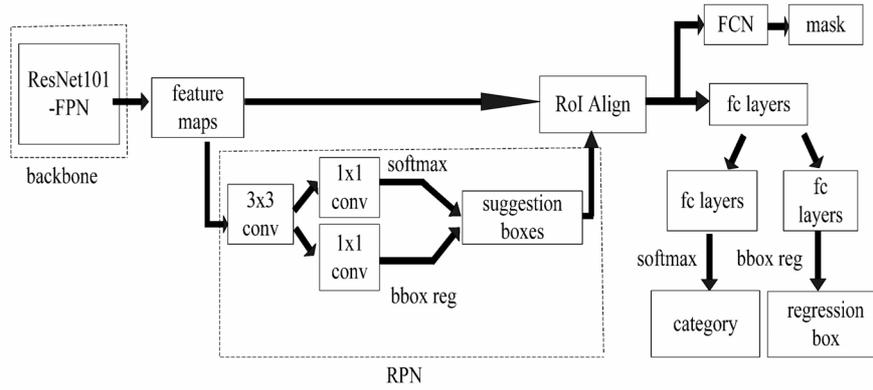


Fig. 6. The Mask R-CNN architecture

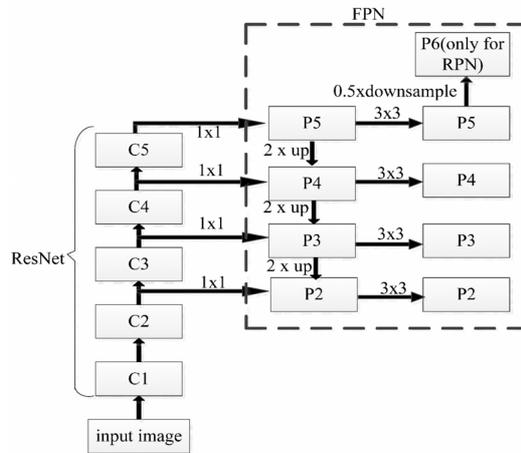


Fig. 7. Backbone network architecture

(2) The position of the pre-selected RoI is usually obtained by model regression, which is generally a floating-point number, and the feature map after pooling requires a fixed scale. Therefore, RoI pooling of Faster R-CNN divides the RoI into several small regions and carries out the max pooling operation on each small region. There are two data quantization processes in this operation: a. the quantification of suggestion boxes to the feature map; and b. the quantization of the feature map to the RoI feature. After these two quantifications, RoI has experienced a serious deviation, which affects the accuracy of detection. To correct the misalignment of RoI pooling, Mask R-CNN proposes RoI Align, which improves the effect of detection and instance segmentation.

RoI Align does not use a quantization operation but uses a “bilinear interpolation algorithm” to deal with floating-point numbers. The algorithm can estimate the corresponding pixel values of the virtual points by using the four real pixel values around the virtual points in the original image. The whole process does not use quantization operations and does not introduce errors. That is, the pixels of the original image and the pixels of the feature maps are completely aligned, which not only improves the accuracy of detection but also is conducive to instance segmentation.

(3) Mask R-CNN adds a mask branch by FCN, which outputs a binary mask with dimension $K * m * m$ for each RoI (where m represents the size of the RoI and K represents the number of categories). Because each layer in the branch can maintain the spatial layout of $m * m$, instead of folding into a vector lacking spatial dimensions, an accurate mask prediction can be achieved.

(4) For each RoI, the multi-task loss function of Mask R-CNN is defined as follows:

$$L=L_{cls} + L_{box} + L_{mask} \tag{3}$$

where L_{cls} and L_{box} are the same as Faster R-CNN and L_{mask} is calculated based on average binary cross-entropy loss of pixel. The specific calculation process is given below.

a. Because the network outputs a mask for each class, the category label predicted by classification branch can be used to select the output mask. For each RoI, if it is detected that it belongs to a certain category, then only the corresponding mask of that category is used for loss calculation (other classes do not contribute to the loss).

b. After the predicted mask is obtained, the sigmoid value is calculated for each pixel of the mask, and then the average cross-entropy of all the pixels in the RoI is taken as L_{mask} .

For L_{mask} , Mask R-CNN and FCN have different calculation methods. In FCN, each pixel in the RoI is classified by a softmax function, and then the softmax-loss of each pixel is taken as a part of the loss function. According to the prediction results of the classification branch, Mask R-CNN divides each pixel into two values (0 and 1), which decouples category prediction and mask prediction well and avoids competition between classes. Prior work [18] has proved that the method of calculating L_{mask} by Mask R-CNN is one of the keys to the network obtaining good segmentation results.

3 Defect Detection and Segmentation of Industrial CT Images

3.1 Image Preprocessing

Due to the influence of noise, artifacts, and production conditions, industrial CT images generally have the characteristics of low contrast, narrow gray ranges, inconspicuous grayscale change, and blurred defect edges [3]. To ensure the accuracy of the subsequent experimental results, enhanced pre-processing of the CT images was first carried out. In this paper, homomorphic filtering and Laplacian sharpening were adopted to enhance the grayscale contrast of images [23]. The images before and after processing are shown in Fig. 8.

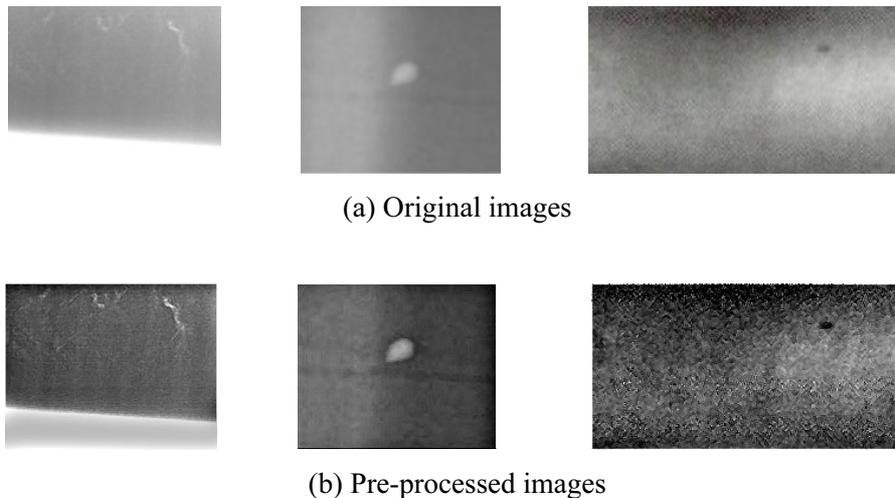


Fig. 8. Image contrast

It can be seen that the target and background of the pre-processed images have significantly enhanced grayscale contrast, which lays a foundation for high-precision detection and segmentation.

3.2 Mask R-CNN Training

According to the data set, the structure of the network was improved and the parameters of the network were modified. Then, the network was trained so that the improved Mask R-CNN could be applied to the detection and segmentation of industrial CT image defects.

3.2.1 Improved FPN

Although FPN uses multi-scale feature information, the fusion path from low-level feature to high-level feature is too long, resulting in a weak role for the low-level features in the whole feature. Based on the

FPN structure, the paper introduced a bottom-up feature fusion path to enhance the whole feature level. The improved feature extraction framework is shown in Fig. 9.

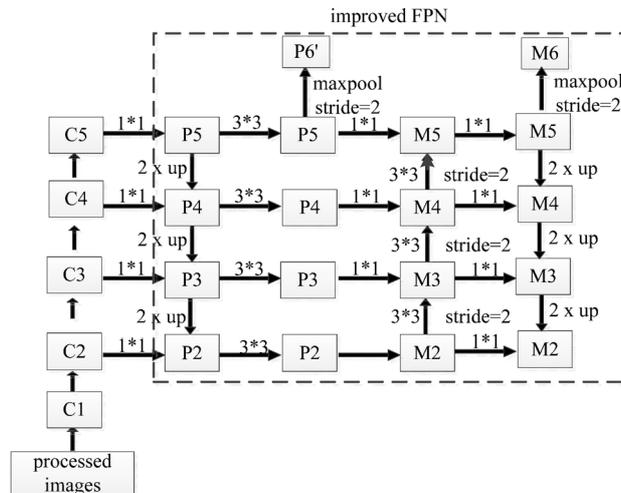


Fig. 9. The improved backbone network architecture

3.2.2 Added Dropout Layer

Because there are too many parameters in Mask R-CNN and too few training samples in this paper, the model is easy to be fitted. To solve this problem, a dropout layer was introduced into the feature extraction network, and the parameter was set to 0.3 (only used in the training stage). This is because the dropout function can suppress the output feature of the last fully connected layer in the feature extraction network.

3.2.3 Anchor Modification

The sizes of the images used in this paper were mostly concentrated between 150×150 pixels $\sim 350 \times 260$ pixels, while the sizes of the defects in the images were mostly concentrated between 12×12 pixels $\sim 35 \times 35$ pixels. Defects only occupied a small part of the background area, that is, they existed in the form of small targets. To improve the accuracy of defect detection, the anchor parameters were modified to increase the overlap between the anchor and the GT as much as possible, to avoid the anchor being discarded as a negative sample or a non-negative and non-positive sample. The anchor ratios were retained, but the scales of the anchor were modified according to the size of GT. The modified anchor parameters are shown in Table 1.

Table 1. Parameters of the anchor

Anchor parameters	value
ratios	[0.5, 1, 2]
scales	(4*6, 8*6, 16*6, 32*6, 64*6)

3.2.4 Learning Rate Modification

A large learning rate helps the fast convergence at the beginning of network training, but it affects the convergence of error at the end of the training. When the learning rate is small, although it is conducive to the full convergence in the later stage, the whole training process converges slowly [24]. Therefore, the learning rate is an important parameter that affects network performance.

The experiment proved that the learning rate is adjusted to 0.001 when the network is pre-trained with coco data set. When the network is fine-tuned, it is adjusted to 0.0001, which has the best detection effect on the data set in this paper. It is worth noting that part of the network improvement is trained separately.

3.3 Mask R-CNN Testing

Using the trained network model to test the test set, that is, when only the test image is input into the model, then the test results can be obtained.

The whole detection process is shown in Fig. 10.

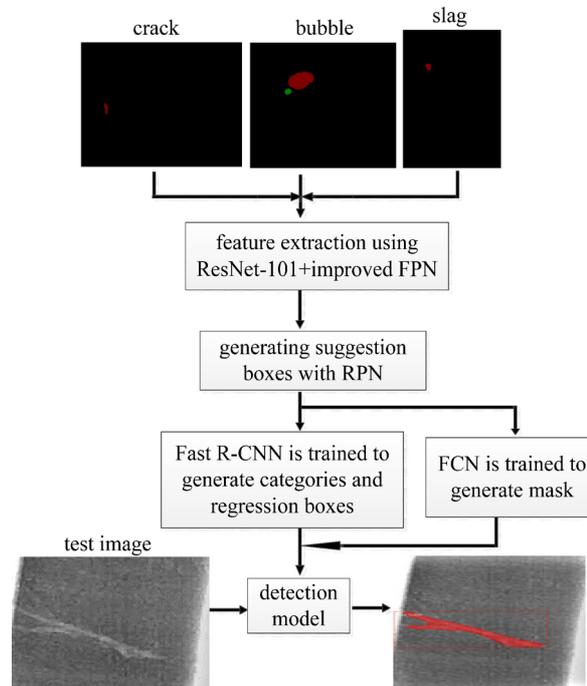


Fig. 10. Defect detection process based on improved Mask R-CNN

4 Experimental Results and Analysis

The experiment was run on an Ubuntu 16.04 system, Tensorflow framework, and Quadro P4000 graphics card. ResNet-101 and FPN were used as the feature extraction networks, and the coco data set was used to pre-train the network. The experimental data (training set and test set) were industrial CT images containing different numbers of cracks, bubbles, and slags.

4.1 Experimental Results

The experimental results of the test set based on improved Mask R-CNN are shown in Fig. 11.

The experimental results show that the trained model is not only suitable for the high-precision detection and segmentation of cracks, bubbles, and slags with large sizes (46×81 pixels) and obvious grayscale changes, but it is also suitable for small defects (4×4 pixels) with low grayscale contrast.

4.2 Analysis of the Experimental Results

To intuitively illustrate that the improved Mask R-CNN has high accuracy in the detection and segmentation of industrial CT defects, this paper compared the experimental results with prior methods [9, 12-13] and FCN.

4.2.1 Comparison and Analysis of the Detection Results

(1) Comparison of the Detection Results

The detection results of the defects using a prior method [9] (Faster R-CNN) and [12] an algorithm (improved Faster R-CNN) are shown in Fig. 12 and Fig. 13.

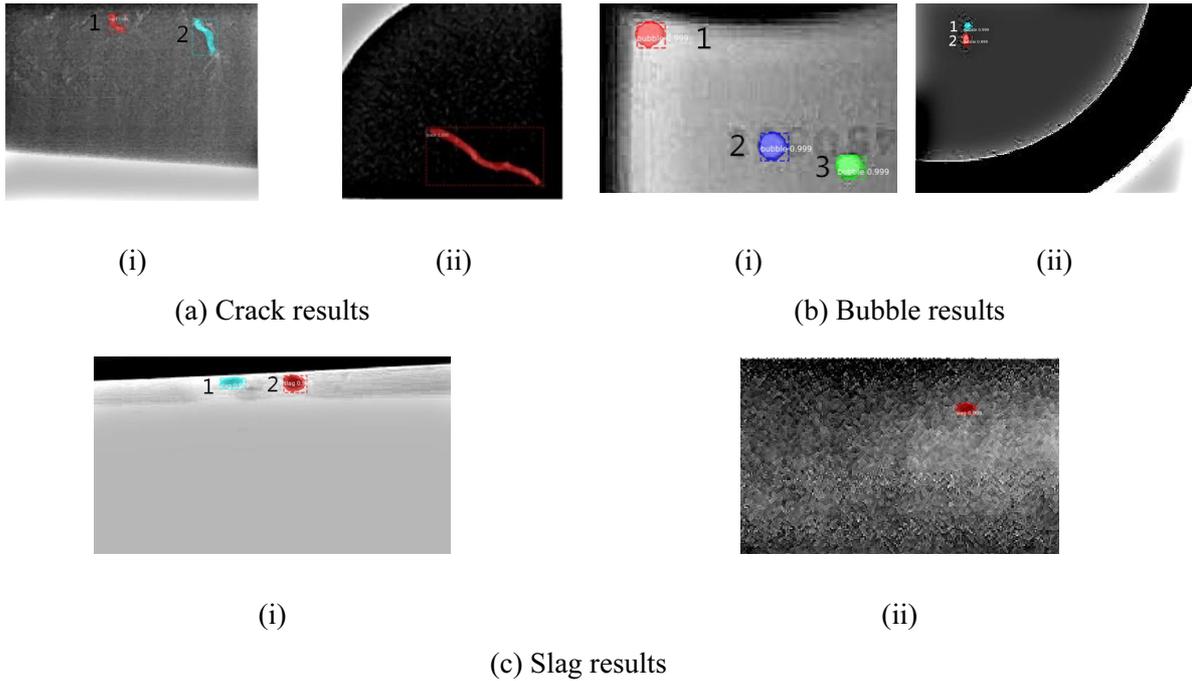


Fig. 11. The experimental results of the proposed algorithm

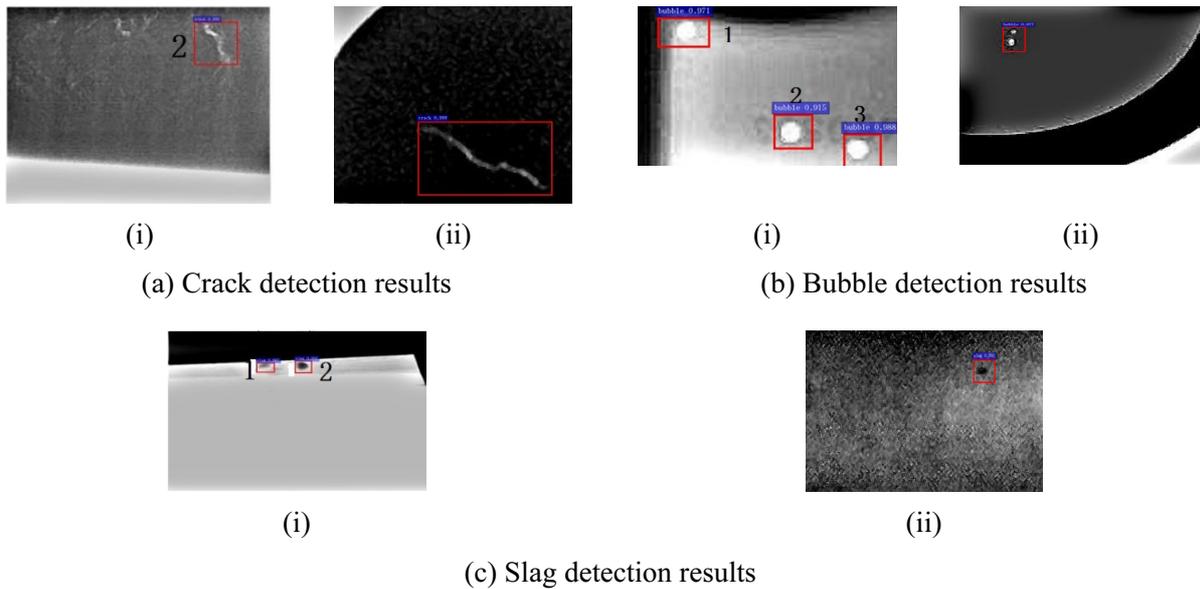


Fig. 12. The defect detection results in the literature [9]

The results of cracks detection using the prior method [9] are shown in Fig. 12(a). Among them, the targets of (i) have the characteristics of small size and inconspicuous grayscale change. The target of (ii) has the characteristics of large size and large grayscale contrast. In comparison to Fig. 11(a), there is the phenomenon of missed detection in (i), and (ii) achieves a more accurate detection.

The results of bubbles detection using the prior method [9] are shown in Fig. 12(b). Among them, the sizes of the bubbles in (i) are larger but the grayscale contrast is lower while the grayscale contrast of the bubbles in (ii) are larger but the sizes are smaller. In comparison to Fig. 11(b), although the bubbles were accurately detected in (i), the locating area is much larger than the size of the defect. Moreover, there is the phenomenon of false detection in (ii), that is, two small bubbles were mistaken as a whole.

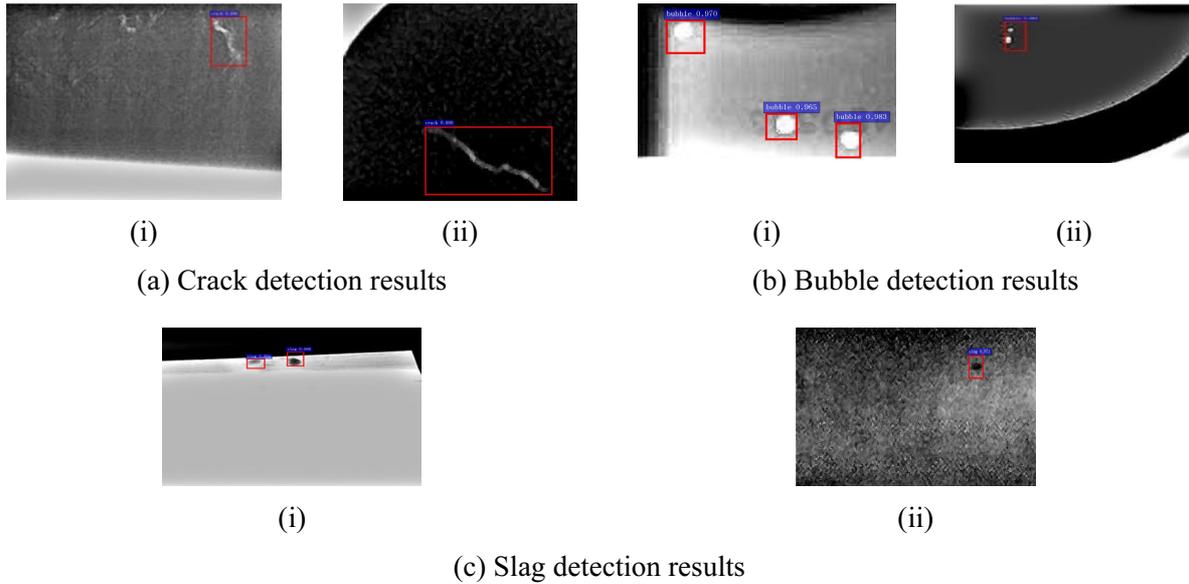


Fig. 13. The defect detection results in the literature [12]

The detection results of the slags using the prior method [9] are shown in Fig. 12(c). Both (i) and (ii) achieved accurate detection, but for (ii) there is no obvious gray change, and the location area is larger than the defect itself. For smaller slags, Faster R-CNN and the improved Faster R-CNN may not be able to achieve a correct detection, as shown in Fig. 14.

The results of defects detection using the prior method [12] are shown in Fig. 13. Compared with Fig. 12 and Fig. 13, the detection results of literature [12] and literature [9] were similar, but literature [12] had better location results than literature [9].

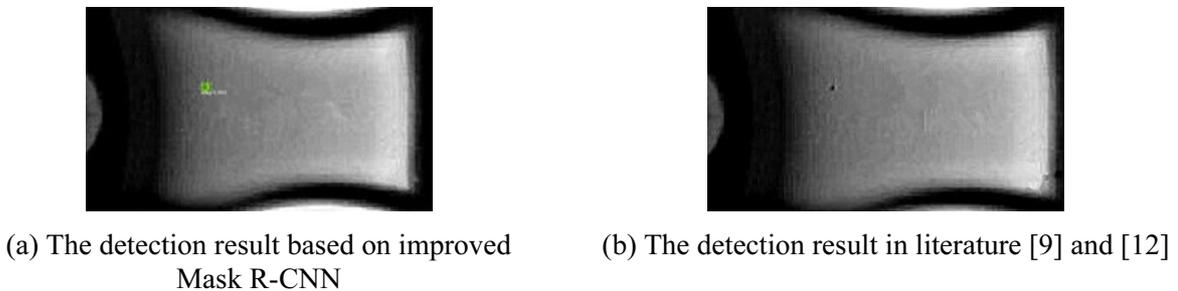


Fig. 14. The detection results of small slag

As shown in Fig. 14(a), the proposed achieved the correct detection of small slag; while Fig. 14(b) shows that prior methods [9, 12] could not detect this target.

The sizes and grayscale contrasts of the defect had little impact on the performance of the proposed algorithm, and it could accurately recognize, locate, and segment all kinds of defects (Fig. 11, Fig. 12, Fig. 13, and Fig. 14). However, although prior methods [9] and [12] could both accurately recognize and locate defects with obvious gray change and large size, there was a very high rate of missed and false detections of the defects with small size or small gray change, and the location area was larger than the size of the defects.

(2) Analysis of the Detection Results

To quantitatively explain the accuracy of detection results, the mean average precision (mAP) was used for analysis and evaluation. The larger the mAP value, the better the detection results.

Under the same IoU threshold, the mAP values of detection results in this paper, literature [9] and literature [12] are shown in Table 2.

As shown, under the same IoU threshold, the mAP value of the proposed algorithm is higher than that of literature [9] and literature [12], which shows that the proposed method has more accurate detection results.

Table 2. Comparison of mAP

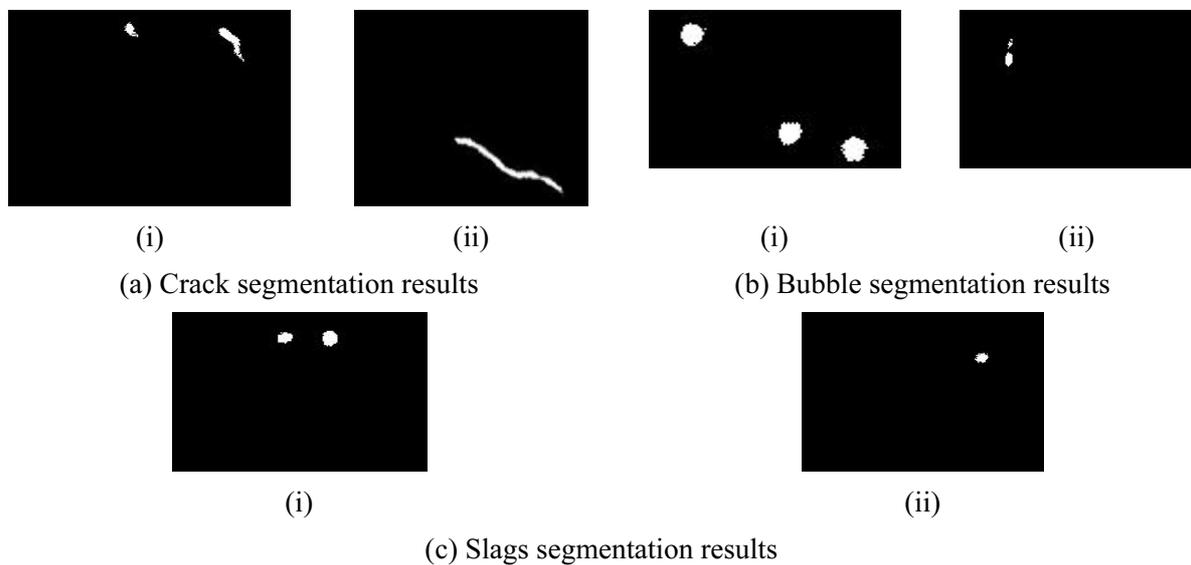
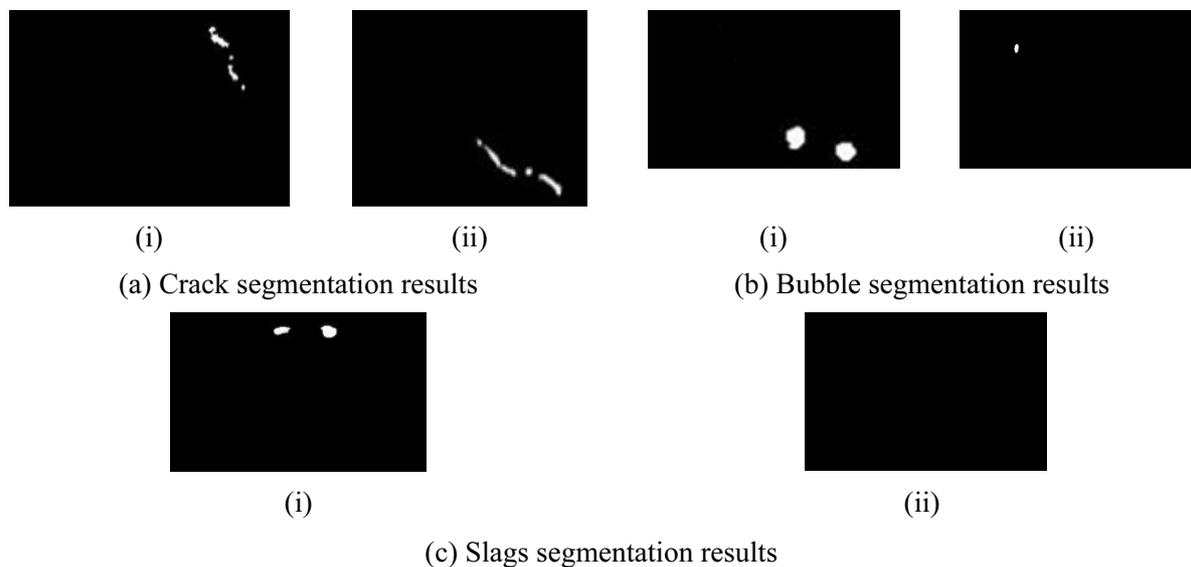
Algorithm	mAP ₅₀	mAP ₇₀	mAP _{c70}	mAP _{b70}	mAP _{s70}
our algorithm	0.98	0.96	0.97	0.96	0.96
literature [9]	0.88	0.85	0.90	0.80	0.88
literature [12]	0.90	0.88	0.93	0.85	0.88

Note. In the Table, mAP₅₀ and mAP₇₀ represent the mAP values when the IoU threshold equals 0.5 and 0.7, respectively. mAP_{c70}, mAP_{b70} and mAP_{s70} represent the mAP values of cracks, bubbles, and slags when the IoU is equal to 0.7, respectively.

4.2.2 Comparison and Analysis of the Segmentation Results

(1) Comparison of the Segmentation Results

As a kind of convolutional neural network, the semantic segmentation network FCN is widely used in image segmentation. In this paper, FCN and a prior [13] algorithm were used as segmentation comparison algorithms, and the segmentation results are shown in Fig. 15 and Fig. 16, respectively (corresponding to Fig. 11).

**Fig. 15.** FCN segmentation results**Fig. 16.** Segmentation results of literature [13]

(2) Analysis of the Segmentation Results

In this paper, the segmentation results were evaluated by three indexes: segmentation accuracy, area ratio (the ratio of segmentation result area to the GT area) and correlation coefficient (γ). The specific values are shown in Table 3. The higher the precision, the better the segmentation result, the larger the value of γ , the higher the similarity between the segmentation result and the GT; the closer the area ratio is to 1, the better the segmentation result is (for the convenience of observation, the above images were magnified at different scales).

Table 3. Segmentation evaluation

Defects	indexes	our algorithm	FCN	literature [13]
(i) crack1	precision	0.961	0.460	0
	area ratio	1.026	1.339	0
	γ	0.996	0.9283	0
(i) crack2	precision	0.973	0.899	0.034
	area ratio	0.998	1.207	1.441
	γ	0.902	0.833	0.502
(ii) crack	precision	0.984	0.972	0.853
	area ratio	0.979	0.941	0.903
	γ	0.995	0.984	0.926
(i) bubble1	precision	0.936	0.923	0
	area ratio	1.012	1.117	0
	γ	0.997	0.990	0
(i) bubble2	precision	0.962	0.927	0.908
	area ratio	1.003	1.007	0.970
	γ	0.996	0.994	0.988
(i) bubble3	precision	0.945	0.915	0.924
	area ratio	1.020	1.054	0.914
	γ	0.996	0.995	0.994
(ii) bubble1	precision	0.930	0.675	0
	area ratio	1.079	1.106	0
	γ	0.985	0.948	0
(ii) bubble2	precision	0.949	0.938	0.736
	area ratio	1.100	1.208	1.198
	γ	0.995	0.989	0.973
(i) slag1	precision	0.938	0.899	0.901
	area ratio	1.119	1.130	1.125
	γ	0.993	0.911	0.960
(i) slag2	precision	0.957	0.765	0.847
	area ratio	1.058	1.116	1.082
	γ	0.996	0.931	0.974
(ii) slag	precision	0.922	0.909	0
	area ratio	1.092	1.133	0
	γ	0.984	0.973	0

Table 3 shows that the algorithm in literature [13] had a high miss rate and low segmentation accuracy for the three kinds of defects. Compared with the results of literature [13], the FCN algorithm greatly improves the segmentation effect of the three defects. However, the proposed achieved better segmentation results than the FCN algorithm for these three defects.

5 Conclusion

In this paper, Mask R-CNN, an instance segmentation network, was applied to the detection of defects in industrial CT images. Combined with the characteristics of defects, the network was improved. The experimental results show that compared with the existing methods of workpiece defect detection and segmentation, the proposed algorithm not only achieves the integration of detection and segmentation

(that is, both tasks of detection and segmentation are completed simultaneously using a network), but also greatly improves the accuracy of detection and segmentation. When the IoU threshold was 0.5 or 0.7, the mAP value of our algorithm was 10% higher than that in literature [9] and 8% higher than that in literature [12]. Compared with FCN and the algorithm in literature [13], the segmentation accuracy of cracks, bubbles, and slags in this paper improved respectively by a maximum of 50.1%, 96.1% and a minimum of 1.2%, 13.1%; a maximum of 25.5%, 93.6% and a minimum of 1.1%, 2.1%; and a maximum of 19.2%, 92.2% and a minimum of 1.3% and 3.7%.

Although our algorithm has made remarkable achievements in the field of workpiece defect identification, location, and segmentation, it still has many problems that need to be improved in follow-up research. For example, (1) the resource consumption is large, and the training time is increased; therefore, it is necessary to strengthen the innovation of training mechanism, monitor the changes of training parameters and convolution features, so that the training process is more controllable and observable; (2) there is no comparison with more basic networks, and more backbone networks will be used for experiments in the future.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Opening Foundation of Key Laboratory of Opto-technology and Intelligent Control, Ministry of Education (KFKT2018-14).

References

- [1] H.-J. Tian, Origin and Application of Industrial CT Technology, in: Proc. 2019 Chongqing Annual Foundry Conference, 2019.
- [2] A. Kehoe, G.A. Parker, An intelligent knowledge based approach for the automated radiographic inspection of castings, *NDT & E International* 25(1)(1992) 23-36.
- [3] J.-Y. Shi, Research on intelligent detection technology of workpiece defect based on industrial CT image, [dissertation] Lanzhou: Lanzhou Jiao-tong University, 2017.
- [4] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313(5786)(2006) 504-507.
- [5] F.-Y. Zhou, L.-P. Jin, J. Dong, Review of convolutional neural network, *Chinese Journal of Computers* 40(6)(2017) 1229-1251.
- [6] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
- [7] R. Girshick, Fast R-CNN, in: Proc. IEEE International Conference on Computer Vision, 2015.
- [8] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39(6)(2015) 1137-1149.
- [9] H.-T. Chang, J.-N. Gou, X.-M. Li, Application of faster R-CNN in image defect detection of industrial CT, *Journal of Image and Graphics* 23(7)(2018) 1061-1071.
- [10] H. Han, Research on industrial defect detection based on deep learning, [dissertation] Chongqing: Chongqing University of Posts and Telecommunications, 2019.
- [11] W.-X. Yan, Research on deep learning and its application on the casting defects automatic detection, [dissertation] Guangzhou: South China University of Technology, 2016.

- [12] Z.-Z. Zhou, Q.-H. Lu, Z.-F. Wang, H.-J. Huang, Detection of micro-defects on irregular reflective surfaces based on improved faster R-CNN, *Sensors (Basel)* 19(22)(2019). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6891612/>.
- [13] X.-Yuan Wu, H.-T. Chang, J.-N. Gou, Research on defect segmentation algorithm of industrial CT image after Faster R-CNN positioning, *Application of Electronic Technique* 45(1)(2019) 76-80.
- [14] X.-T. Li, G. Zhang, Q. Yang, L.-Q. Yin, A method for enhancing and segmenting image of surface defect of non-uniform illumination strip, *Control Engineering of China* (2019) 1-7.
- [15] Z.-Y. Yu, Fully Convolutional Networks for Surface Defect Inspection, [dissertation] Harbin: Harbin Institute of Technology, 2018.
- [16] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39(4)(2017) 640-651.
- [17] X. Bian, S.-N. Lim, N. Zhou, Multiscale Fully Convolutional Network with Application to Industrial Inspection, in: *Proc. 2016 IEEE Winter Conference*, 2016.
- [18] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: *Proc. IEEE International Conference on Computer Vision*, 2017.
- [19] T.Y. Lin, P. Dollár, R. Girshick, K. He, S. Belongie, Feature pyramid networks for object detection, in: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [20] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: *Proc. International Conference on Learning Representations-ICLR*, 2015.
- [21] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: *Proc. 13th European Conference on Computer Vision-ECCV*, 2014.
- [22] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [23] H.-Y. Chen, S. Xu, K. Liu, H.-X. Sun, Surface defect detection of steel strip based on spectral residual visual saliency, *Optics and Engineering* 24(10)(2016) 2572-2580.
- [24] L. Qiao, E.-D. Zhao, J.-J. Liu, B. Cheng, Research of workpiece defect detection method based on CNN, *Computer Science* 44(11A)(2017) 238-243.