Research on Iron Surface Crack Detection Algorithm Based on Improved YOLOv4 Network

Huafu Deng¹, Jianghua Cheng¹, Tong Liu¹, Bang Cheng¹ and Zhenhua Sun²

¹College of Electronic Science and Technology, National University of Defense Technology, 410073, Changsha, China
²Academy of Military Sciences PLA China, 100091, Beijing, China
Email: hankson@126.com

Abstract. Metal surface defect detection has always been an important branch of target detection. In certain places, if the cracks on the metal surface can be found in time, the existing safety hazards can be eliminated. In this paper, a polarized imaging camera with strong environmental applicability is used for sampling, and the degree of polarization image is applied to the detection of iron material cracks. The iron material polarization image crack data set (PICD-iron) is established, and the Cascade-YOLOv4 (C-YOLOv4) network model is used for crack detection. Then it solves the problems of target detection in the dark environment, complex and variable crack detection, small target detection. Experimental tests verify that the detection accuracy of the C-YOLOv4 net-work has improved while comparing with the YOLOv4 network, the detection speed has also increased by 28%.

1. Introduction
The requirements of modern industry on the quality of metal materials are becoming higher and higher, which has prompted the detection of metal surface defects to become a research hotspot in related industries. In recent years, many application competitions have been held. In the 2018 Alibaba Cloud Tianchi Big Data Competition [1], the defect detection of industrial parts has become a hot topic, and the research results of the competition are directly related to the industrial applications of enterprises. Steel defect detection in the Kaggle competition [2] in 2019 attracted 2431 teams from all over the world to participate. In academia, it is also a hot research topic. Landstrom [3] used morphology to deal with steel plate defect detection problems and used logistic regression for statistical classification to achieve high crack detection accuracy. However, this algorithm is to detect specific defects, and their overall applicability is not good. Xu [4] solved the difficult problem of defect detection for non-planar industrial parts, advanced algorithm Faster R-CNN was introduced to detect defects in metal arc industrial parts. After the deep learning algorithm is adopted, the missing detection rate of arc surface defects is reduced, and the detection rate is greatly improved. Li [5] introduced the YOLO algorithm with excellent detection speed into the detection of surface defects of steel strips to solve the problem of real-time detection of surface defects of steel strips, and the detection rate was further improved. These two convolutional neural network algorithms are used for the quality screening of industrial parts before leaving the factory. When dealing with the detection of cracks and defects on metal surfaces in actual use, the environment is more complicated and the algorithm is not applicable. Kazama and Oshige [6] used polarization images to detect surface defects of steel strips for the first time and achieved good results using tradition-al algorithm detection. The X-3Dvision defect detection system [7] of the German IMS Messsysteme GmbH company can display
the depth information while displaying the defect, making the defect detection more one-dimensional reference information. These two schemes do not use ordinary visible light images to study the detection of defects on metal surfaces, which has opened up new ideas for us to study such problems and has reference significance.

This article mainly studies the detection of surface defects of iron materials in the actual use environment. Due to the complex environment of this industrial equipment, the light and shade change greatly, and cracks are small targets. The crack may be arc-shaped and complex in shape, and some water and oil stains will also appear in the defects. In this regard, this paper proposes an improved C-YOLOv4 network, and preprocesses the image at the input end of the network, and inputs the degree of polarization (DOP) image for training and detection. The scheme of the DOP image + C-YOLOv4 network is used for crack detection, and the precision and speed of crack detection are greatly improved.

2. Discussion on Related Issues
Crack detection in the use scene of iron material industrial parts, because of the complicated environment, there are several problems to be solved in crack detection (figure 1). (1) Ambient light is very complex, most of them are dim; (2) Small crack target detection problem; (3) The causes of crack defects are different, the shape is complex and variable. The corresponding crack detection algorithm also revolves around how to solve these three problems. With the continuous improvement of computer performance, deep learning has shown strong capabilities in image processing. The convolutional neural network is inspired by the biological vision system, and it uses convolution kernels to simulate neuronal cells in the visual cortex to find specific target features and has good target detection performance [8]. We choose the convolutional neural network methods and use the YOLOv4 network with excellent target detection speed and detection accuracy for crack detection. However, the convolutional neural network needs to learn more features from a large number of samples, and it is also necessary to establish a corresponding image database. The predecessors used convolutional neural networks in crack detection algorithms. Most of them used ordinary visible light images as the input images of the network, ordinary visible light images, especially grayscale images, were greatly affected by lighting conditions. But the degree of polarization image can still reflect the characteristics of different materials and target structures in the dark environment [9] (figure 2).

![Figure 1](image1.png)
(a) Dark environment, (b) Cracks are small targets, (c) Crack shape is complex and changeable.

![Figure 2](image2.png)
(a) Visible light image, (b) Degree of polarization image.

We use a CCD polarizing camera equipped with a Sony IMX250-MZR sensor to take pictures to build an image database. The image it takes is a visible light polarization image with 1224*1024
The visible light polarization image is the image obtained after the target light passes through the polarizer. It can acquire visible polarization images in four directions (0°, 45°, 90°, 135°) at the same time. To obtain a crack image closer to the actual application environment, we shoot under different lighting conditions and choose three different angles to shoot the defect. The image data set is expanded by flipping the image, rotating the image, and adding Gaussian noise to the image to prevent training from overfitting. Finally, the created iron material polarization image crack data set (PICD-iron) (figure 3) contains visible light polarization images (7165 sheets) and degree of polarization images (3416 sheets). There are two main types of cracks, one is a crack that explodes under pressure, and the other is a severely corroded crack (figure 4).

![Figure 3](image1.png)  
**Figure 3.** (a) Visible light polarization image data set, (b) Degree of polarization image data set.  

![Figure 4](image2.png)  
**Figure 4.** (a) Burst cracks, (b) Rust cracks.

3. Algorithm Research and Experimental Results

3.1. Defect Detection in Complex Ambient Light

In grayscale images, the cracks on the surface of the iron material under different lighting conditions, the darker the light, the easier the cracks are hidden. The degree of polarization image is related to the target material and structure, and the broken grain at the cracks has structural characteristics. Under different lighting conditions, the cracks in the degree of polarization image can be well distinguished from the background (figure 5). Therefore, this paper uses the degree of polarization image for the detection of cracks on the surface of the iron.

![Figure 5](image3.png)  
**Figure 5.** Defects in from the light to the dark environment. First row: visible polarized image. Second row: the degree of polarization image.

We use the Stokes vector $S = (S_0, S_1, S_2, S_3)^T$ to analyze the various polarization states of the image. The $S$ component of the electrical vector and the light intensity value $I$ of the polarization component have the following relationship [10]:

$$S = (S_0, S_1, S_2, S_3)^T$$

$$I = S_0$$
\[ \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} I(0,0) + I(\frac{\pi}{2},0) \\ I(0,0) - I(\frac{\pi}{2},0) \\ I(\frac{\pi}{4},0) + I(3\frac{\pi}{4},0) \\ I(\frac{\pi}{4},0) - I(3\frac{\pi}{4},0) \end{pmatrix} = \begin{pmatrix} I_+ + I_- \\ I_+ - I_- \\ I_{45} + I_{-45} \\ I_{45} - I_{-45} \end{pmatrix} \]

\[ S_0 \] is the total light intensity of the light wave. \( S_1, S_2 \) can be obtained according to the addition and subtraction of the polarization component in the corresponding direction of the light wave. \( S_3 \) is the subtraction of two circular polarization components \( (I_R, I_L) \). \( S_4 \) is usually assumed to be zero [11].

The degree of polarization (DOP):

\[ DOP = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \]  

The current deep learning target detector is mainly composed of 4 parts (figure 6): Input, Backbone, Neck, Prediction. YOLOv4 network is developed based on the YOLOv3 network, and it is a target detector with better speed and accuracy. In the YOLOv4 network model, CSPDarknet53 is used as the backbone, and YOLO is still used as the Prediction [12]. In this paper, the DOP image is used as the input image of the convolutional neural network. During crack detection, the preprocessing added at the Input end is used to extract the target DOP image.

![Figure 6. YOLOv4 network structure framework.](image)

Comparing the visible polarization image with the ordinary visible light image, the visual change is not great. To verify the effectiveness of the polarization image for crack detection, we added a comparative experiment. The YOLOv4 net-work is trained on the visible light polarization image data set and the DOP image data set (learning_rate = 0.0001, batch_size = 64, epochs = 1000). After that, the test was performed in a test set containing 155 images (figure 7). It can be seen that in a dim environment, the DOP image is used for defect detection, and the detection effect is better than the visible light polarization image. The error detection situation of these two tests is shown in figure 8.
Figure 7. Detection results of different data set after training. (a) (c) visible light polarization image, (b) (d) The DOP image.

Figure 8. Error detection situation after training on the different data set. (a) Visible light polarization image, (b) The DOP image.

3.2. Cracks Are Variable and Small Target Detection Problems

3.2.1. Cracks Are Complex and Variable. Due to the complex and variable shapes of cracks, convolutional neural networks require a lot of image data to learn their features. On the other hand, if the entire crack is marked with a box, it also requires a lot of image data. However, it is difficult to list all shapes of cracks. If the image data set is insufficient, it is easy to overfit. In this paper, the method of section defect marking is used to solve the problem of difficult to identify complex cracks (figure 9). A large number of small cracks are marked in an image so that if the number of images in the training set is not very large, there can be enough training samples.

Figure 9. (a) Crack mark frame, (b) Crack identification frame.

3.2.2. Small Target Detection. The YOLOv4 network uses 3×3 anchor boxes to return to the bounding box. The average cross union ratio (Avg IOU) of the rectangular frame is used to constrain, and the size of the anchor frame is obtained by k-means clustering on the training set. YOLOv4 author anchor frame selection is ([12 16], [19 36], [36 75], [76 55], [72 146], [142 110], [192 243], [459 401]).

\[
I_{IOU} = \frac{\text{area}(\text{box}_{true} \cap \text{box}_{pred})}{\text{area}(\text{box}_{true} \cup \text{box}_{pred})} \tag{3}
\]

\[
d(\text{box}, \text{cen}) = 1 - I_{IOU}(\text{box}, \text{cen}) \tag{4}
\]

\[
f = \arg \max \frac{\sum I_{IOU}(\text{box}, \text{cen})}{n} \tag{5}
\]

In the formula, the box represents the size of the rectangular box, and the cen represents the size of the rectangular box selected by the k-means clustering algorithm. \(d(\text{box}, \text{cen})\) is the distance measure of the k-means clustering algorithm, and \(I_{IOU}\) is the ratio of the prediction box and the real box. \(N\) represents the total number of samples. The average intersection ratio of the cluster determines the size
of the anchor frame by the objective function $f$. The choice of anchor frame affects the convergence and detection accuracy of the model. The size of the selected anchor frame will also affect the performance of the target detector, so the choice of anchor frame is very important [13]. Therefore, according to the size of the cracks we marked, reselect the anchor frame. As can be seen from figure 10, when the anchor frame $K$ is selected as 8, clustering Accuracy reaches 80%, then $K$ increases and the Accuracy value does not change much. Here we still select 9 anchor frames, and re-select the anchor frame as ([44 59], [52 79], [60 35], [63 146], [65 108], [75 48], [83 78], [100 56], [150 71]).

The YOLOv4 network outputs three different scale detection maps for detecting objects of different sizes, thereby achieving end-to-end target detection (figure 11). Although it has an excellent detection speed, it is a regression-based detection method. Because it does not need to generate a suggestion box, there is a certain positioning error, resulting in a small target detection accuracy is not high [14]. In this regard, we propose a Cascaded YOLOv4 (C-YOLOv4) network model (figure 12), which improves the robustness and crack detection accuracy of the network model.

Figure 10. Correspondence curve of anchor frames $K$ and accuracy of $K$-means clustering.

Figure 11. Schematic diagram of the correspondence between the three feature maps and anchors. (a) 19x19 feature map, (b) 38x38 feature map, (b) 76x76 feature map.

Figure 12. C-YOLOv4 network model.

In the first training stage, we discard the predictive regression of the (76x76) feature map. We mark crack defects in stages, the detection targets are very small, and the corresponding detection in the (76x76) feature map is a small size target. When the number of training images is not sufficient, the network does bounding-box regression on the (76x76) feature map. It is easy to cause overfitting and consume more time, which increases the cost of training time. It should be noted that the (76x76) feature map needs to be retained because the FPN layer of YOLOv4 needs to predict the (19x19) feature map and (38x38) feature map through the (76x76) feature map through two PAN structures. In the second training stage, the (19x19) features are discarded before training. The (19x19) feature map is used to detect large targets, and the crack information is very little. Besides, the image size of our
input network is 1224×1024, and when doing bounding-box regression, a grid pixel corresponding to the 19×19 feature map is 64×54. K-means clustering statistics show that nearly 70% of the crack pixels are below 60 (figure 13), and the corresponding anchors of the 19×19 feature map are larger ([83 78], [100 56], [150 71]). Many operations are useless when doing bounding-box regression. The (19×19) feature map is discarded, and the large target recognition frame is also discarded. In this way, the bounding-box regression can be more accurate. The detection results of the C-YOLOv4 network model are shown in figure 14. It can be seen that compared with the YOLOv4 network model, the C-YOLOv4 network model can identify more crack targets, and the positioning is more accurate.

Figure 13. Crack width and height distribution, a red dot is the cluster center.

Figure 14. (a) (c)YOLOv4 network detection crack results, (b) (d)C-YOLOv4 network detection crack results.

For the detection of surface defects of iron materials in the actual use environment, we used the DOP image + C-YOLOv4 network for crack detection and tested on 155 images. The comprehensive comparison is shown in table 1. The false detection rate of cracks is 0.1%, and the rate of missed detection of the crack image is 7%, and the detection effect is best. In terms of detection speed, the detection time of Faster R-CNN for defect detection of industrial parts in Ref. [4] is 300ms. In this paper, the YOLOv4 network is used to detect cracks, and the detection time is 245ms. The C-YOLOv4 network has the fastest detection speed and the detection time is 191ms. Compared with the YOLOv4 network our scheme detection speed increased by 28%.

Table 1. Crack detection on test set images.

| Scheme                  | Number of cracks detected | Number of error detection | Number of missed images | Crack misdetection rate | Crack image missing detection rate | Average detection time per picture (ms) |
|-------------------------|---------------------------|---------------------------|-------------------------|-------------------------|------------------------------------|---------------------------------------|
| Polarization image + YOLOv4 | 469                       | 23                       | 26                      | 4.9                     | 16.8%                              | 279                                   |
| DOP image + YOLOv4      | 902                       | 4                        | 13                      | 0.4%                    | 8.3%                               | 245                                   |
| DOP image + C-YOLOv4    | 1296                      | 1                        | 11                      | 0.1%                    | 7%                                 | 191                                   |

4. C-YOLOv4 Network Video Detection Results
To verify the reliability of our algorithm, video crack detection was verified in the simulated scene. The experimental computer equipment is configured with Windows 10 system, Python version 3.6.10, Tensorflow-GPU version 1.8.0, and Keras version 2.1.6. The hardware equipment is NVIDIA GeForce GTX1080Ti graphics card and a CCD polarized camera with a Sony IMX250MZR sensor. Use a polarizing camera to slowly rotate the angle to shoot a simulated scene in real-time to identify
the cracks on the iron surface in the scene. The iron objects in the scene include iron racks, iron plates, round iron pipes, square iron pipes, and a broken iron basin. In dimly lit scenes, the DOP image + C-YOLOv4 network scheme is used for crack detection, and the display results are shown in figure 15. By intercepting the number of frames, it can be seen that cracks in the iron basin can be accurately detected.

![Figure 15](image-url)

**Figure 15.** (a) Simulated scene. (b) First row: visible polarization image captured by video, Second row: the DOP image + C-YOLOv4 real-time detection results.

## 5. Conclusion

For the problems that need to be solved in the detection of cracks on the surface of iron materials in the actual application environment. We established an iron material polarization image crack data set, segmented annotation cracks, and built a C-YOLOv4 network. The network can detect small targets better. The scheme of the DOP image + C-YOLOv4 network can effectively detect the crack defects in the dark environment. The results show that our network has good detection accuracy and detection speed, and realizes an end-to-end detection solution for cracks on the surface of iron materials in complex environments, which has certain practical application value.

## Acknowledgment

This work was supported by the Hunan Natural Science Foundation of China under Grant number 2020JJ4670.

## References

[1] https://tianchi.aliyun.com/competition/entrance/231681/introduction.
[2] https://www.kaggle.com/c/severstal-steel-defect-detection/data.
[3] Landstrom A and Thurley M J 2012 Morphology-based crack detection for steel slabs *IEEE Journal of Selected Topics in Signal Processing* 6 (7) 866-875.
[4] Xu H J 2017 Research on metal cambered workpiece’s surface defect detection and false detection result exclusion technology[M] *South China University of Technology* pp 14-20.
[5] Li J Y, Su Z F, and Geng J H 2018 Real-time detection of steel strip surface defects based on improved YOLO detection network *IFAC PapersOnLine* 51 (21) 76-81.
[6] Kazama A and Oshige T 2012 Surface inspection technique for steel strip using relative intensity characteristics of multi azimuth polarized images *Transactions of the Society of Instrument & Control Engineers* 2012 (47).
[7] http://www.surcon-gmbh.de/.
[8] Xu K 2012 *Study of Convolutional Neural Network Applied on Image Recognition* (China: Zhejiang University) pp 11-14.
[9] Zhao C X 2016 *Research of Image Defogging Algorithm Based on the Polarization Information* (China: Changchun University of Science and Technology) pp 15-18.
[10] Qiao J 2017 *Research on Image Fusion Technology Based on Polarization* (China: Changchun University of Science and Technology) pp 7-9.
[11] Liu X, Wang F and Xue M G 2008 The study of camouflage target detection based on polarization characteristics *Optical Technology* 2008 (05) 787-790.
[12] Bochkovskiy A, Wang C Y and Liao H Y M 2020 YOLOv4: Optimal speed and accuracy of object detection *Computer Vision and Pattern Recognition* (DOI arXiv:2004.10934).

[13] Ju M R, Luo H B and Wang Z B 2019 Improved YOLOv3 algorithm and its application in small target detection *Acta Optica Sinica* 039 (007) 245-252.

[14] Zheng Q M, Wang L L and Wang F H 2020 Small object detection in traffic scene based on improved convolutional neural network *Computer Engineering* 46 (06) 26-33.