WILEY | Hindawi

Research Article

Positioning of the Moving and Static Contacts of the Switch Machine Based on Double-Layer Mask R-CNN

Jiacheng Yin [ID],[1] **Zhaomin Lv** [ID],[1] **Xingjie Chen** [ID],[1] **and Kun Yang**[2]

[1]*School of Urban Railway Transportation, Shanghai University of Engineering Science, Shanghai 201620, China*
[2]*Shanghai Trusted Industrial Control Platform Company Limited, Shanghai 200333, China*

Correspondence should be addressed to Zhaomin Lv; zhaomin_lv@sues.edu.cn

With the continuous development of rail transit, the maintenance of the switch machine becomes more and more important, and the contact depth of the moving contact and static contact in the switch machine is a key part of it. At present, the manual measurement method is the main measure of contact depth, which has the problems of low efficiency and strong subjectivity. The measurement of contact depth based on machine vision includes two steps: moving and static contact positioning and distance conversion. The positioning result will have an important impact on distance measurement. Therefore, a positioning method for moving and static contact based on double-layer Mask R-CNN (DLM) is proposed in this paper: first, the moving contact is roughly positioned by Mask R-CNN to obtain the predicted target area; second, the subgraph of the target area is preprocessed; finally, the precise positioning is used to determine the precise position of the moving and static contact. The accuracy and robustness of the proposed DLM are verified by the internal image of the switch machine.

1. Introduction

In recent years, the technology development in the field of rail transit becomes more and more mature, and safety is an essential attribute of rail transit. Trackside facilities such as switch machines play an important role in the safe operation of rail transit. Once the switch machine fails, serious train derailment accidents will occur, so the switch machine must be in good working condition at all times. The switch machine will inevitably wear in daily use, which needs to be observed and maintained on time. The contact depth of the moving and static contacts of the switch machine determines whether the switch machine can work normally, so they become the key to maintenance.

Moving contact and static contact are indispensable parts to complete this task. As shown in Figure 1, the red hollow box is the static contact area, one at the top and one at the bottom. The small red solid circle is the static contact in this paper, while the moving contact in this paper is the blue solid circle. It swings up and down with the state of the switch machine and contacts with the tape spring to conduct current to make the switch machine work. The distance between the yellow lines is the contact depth, which can be calculated from the relative position of the moving and static contact. The contact effect is determined by the depth of the moving contact driving into the static contact.

Among them, the contact depth of the moving contact column and the static contact base of the switch machine become the key to maintenance. If the contact depth is not up to the standard, the electric circuit and some serious consequences would be caused by the poor contact depth. At present, manual participation is still needed for the inspection of the switch machine. Due to the complexity of the switch machine structure, the inspectors need a lot of professional knowledge and testing experience, the test method cannot be promoted, and so on. These reasons have caused the low efficiency of the maintenance and protection of the switch machine.

The maintenance efficiency can be improved by using the automatic method to measure the contact depth of the moving contact and static contact, and artificial intelligence has made great progress in data-driven modeling [1, 2]. The
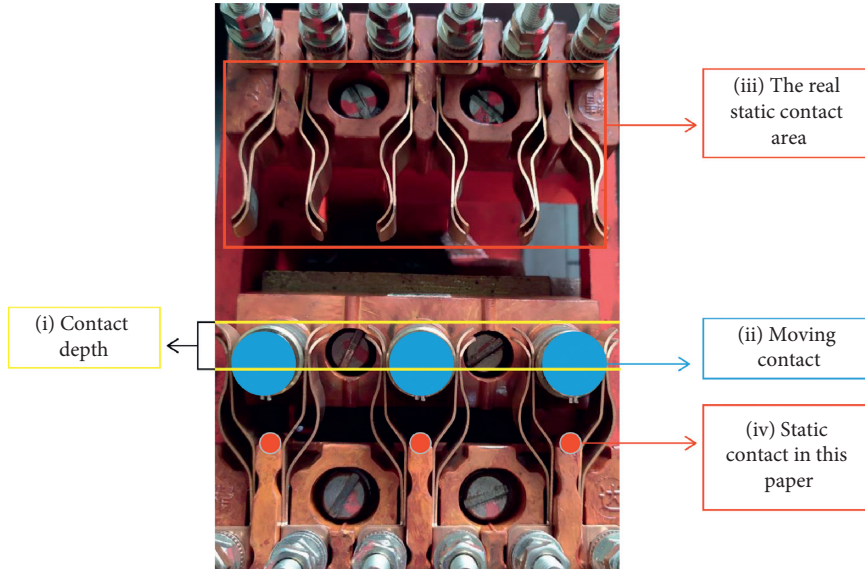
FIGURE 1: Internal structure diagram of switch machine. (i) The contact depth: the depth of the moving contact driving into the static contact. (ii) The moving contact. (iii) The real static contact area. (iv) The static contact in this paper.

automatic detection method is mainly noncontact detection, which will not affect the surface deformation and wear of the device. At present, the common noncontact distance detection methods are mainly divided into ultrasonic detection method [3, 4], laser detection methods [5], measurement based on stereo vision, machine vision measurement, and so on. Ultrasonic and laser detection methods require high reflectivity of the object surface to be measured. If the reflectivity of the object surface does not meet the standard, the measurement effect will be poor. The measurement technology based on stereo vision is more stringent on the number and placement conditions of cameras. However, the on-site conditions in the switch machine case do not meet the requirements of the above methods, and the above distance measurement methods are not suitable for the detection of the driving depth of the mobile static contact of the switch machine. Therefore, the machine vision measurement method is selected.

When using machine vision for automatic measurement, the first step is accurate positioning of moving and static contact's target area and the second is distance measurement based on the image. The result will be greatly affected by the accuracy of each step, and the target positioning has the greatest impact on the accuracy. Therefore, it is necessary to select an accurate positioning method.

At present, the methods of target location are divided into traditional methods and modern methods. Traditional target location methods include feature extraction and feature classification. Template matching [6, 7] is a common traditional location method. In the aspect of image feature extraction, there are local binary pattern feature [8], Histogram of Oriented Gradient (HOG) [9, 10], Haar feature [11], and other features. After obtaining image features, the similarity measure [12, 13] is used to classify and locate the target. A dynamic positioning algorithm based on template matching [14] is proposed by Yin et al. to detect the area,

width, and distance of groove shape. The limitation of this method is that it requires high image quality and cannot recognize the rotation or size change of the matching target. These traditional image feature extraction methods have high normalization to the image and cannot adapt to the complex and changeable environment. In the application of engineering practice, the detection accuracy and anti-noise ability are poor, the robustness of image data processing is weak, and the effect needs to be improved.

With the deepening of image processing and image classification based on deep learning [15–17], modern methods of target location have achieved important research results, such as You Only Look Once (Yolo) [18], Convolutional Neural Networks (CNNs), Regional Convolutional Neural Networks (R-CNNs), Fast Region-Convolutional Neural Networks (Fast R-CNNs) [19], Faster Region-Convolutional Neural Networks (Faster R-CNNs) [20], and Mask R-CNN [21–24]. At present, Mask R-CNN is an excellent deep learning network compared with the previous generation. It adds segmentation branches and carries out target detection and segmentation tasks synchronously. It introduces Region of Interest Alignment (RoI Align) to replace Region of Interest Pooling (RoI Pooling) in Faster R-CNN, which greatly improved the accuracy of region segmentation. Compared with the traditional methods, Mask R-CNN has a stronger ability of high-level semantic abstraction, translation invariance of convolutional neural network, and scale invariance within training, which are also necessary for image classification. Mask R-CNN is the most efficient and widely used method in the target location.

When we use Mask R-CNN for one-step positioning of the static contact of the switch machine, we find that the positioning effect of the moving contact is good but the positioning effect of the static contact is poor. This is because the internal structure of the switch machine chassis is

complex, there are many parts with similar shapes and colors, and the shape of the static contact to be positioned is too small, which will lead to the low positioning accuracy of the static contact area. Therefore, the static contact area cannot be accurately located by the one-step Mask R-CNN method.

To solve the above problems, this paper proposes to use two Mask R-CNNs to locate the moving and static contact, which are divided into rough positioning and precise positioning. The definition of rough positioning in this paper is as follows: Mask R-CNN is used to get the subgraph containing the moving and static contact area so as to reduce the influence of irrelevant environment and improve the subsequent positioning accuracy. The definition of precise positioning is as follows: after the subgraph is preprocessed, the Mask R-CNN is used to locate the moving and static contacts and finally get their accurate positions. The problem of inaccurate positioning of static contact can be effectively solved by DLM.

Compared with R-CNN, the improvement of DLM lies in the following: because of using the multistep positioning method, the positioning accuracy of DLM will be higher than that of R-CNN, especially for objects with a small pixel value. When R-CNN locates the target, it uses a rectangular box to locate, while DLM is accurate to the pixel level, which is helpful for the subsequent contact depth measurement. The training process of R-CNN is very complicated, including pretreatment training sample, parameter fine-tuning, SVM training, and bounding box regression training. Multiple GPUs are used to accelerate the DLM training time, which improves the efficiency compared with R-CNN.

The main innovations of this paper are as follows. (1) A machine vision-based automatic positioning method for moving and static contact of switch machine is proposed to carry out accurate positioning in a complex environment. (2) A multimodel method is proposed for object location and subgraph segmentation. The DLM method is used to solve the problem that it is difficult to locate small pixel objects in large pixel images, and the positioning accuracy is improved.

The remainder of this paper is organized as follows. Section 2 introduces the theory of Mask R-CNN. Section 3 describes the theory of the proposed DLM switch machine moving and static contact positioning method. Section 4 introduces the experimental results of DLM theory, and the improvement of the accuracy of the DLM is verified compared with the single Mask R-CNN and Yolo. Section 5 summarizes and prospects the full text.

2. Basic Introduction of Mask R-CNN

Image classification is the basic task of image recognition. Mask R-CNN is mainly used for object detection and entity segmentation. It inputs images and outputs them as image categories and object masks. Its network architecture is shown in Figure 2.

Compared with Fast R-CNN, some changes have been made as follows:

(1) Replace the ROI pooling layer with ROI align

(2) Add parallel FCN layers

(3) Add feature extraction network to resnet101 + FPN to enhance feature extraction ability

Mask R-CNN adopts multitask loss function as follows:

$$L = \overline{L}_{cls} + L_{box} + L_{mask}. \tag{1}$$

The loss function of each ROI region consists of three parts: classification loss value of bounding box, location regression loss value of bounding box, and loss value of mask. The mask branch has $K * m^2$ dimensions output for each RoI (K is the number of masks and m^2 is the resolution). There are 2 (2 is the number of categories: target and background) binary masks with $m * m$ resolution during training. For a class of RoI, L_{mask} only considers the mask of this target class. The input of other kinds of masks will not be calculated into the loss function. In the feature layer of $K * m^2$, each value is a binary mask: 0 or 1. In the first target region prediction, the first feature layer with the first resolution of $m * m$ is selected, and then the average binary cross-entropy loss is calculated, which is the loss function of the mask branch.

3. Moving and Static Contact Positioning Based on DLM

3.1. Algorithm Process. This section will introduce the basic process of the algorithm, the specific process is shown in Figure 3, and the overall process is divided into two steps. The first step is rough positioning, which inputs the original image of the switch machine into the step 1 Mask R-CNN for rough positioning, so as to obtain the moving contact area and eliminate the adverse effect of other irrelevant area pixels on subsequent measurement. The second step is precise positioning, which processes the image of the moving contact area obtained in step 1 to obtain the subgraph with fixed size and including the moving and static contact. The subgraph is input into the step 2 Mask R-CNN, and the static contact can be accurately obtained through the pretrained subgraph model for accurate contact depth measurement.

On the basis of Mask R-CNN, the DLM can achieve accurate positioning in complex environment and solve the problem that the internal structure of switch machine is complex and the effect of the ordinary positioning method is not good. The DLM can be used to locate small objects in large images, and the experimental results show that it can improve the positioning accuracy.

3.2. Rough Positioning. In most cases, the direction of the manual photos of the switch machine is not the same. Because the size of the input image in the neural network will be reset in the input layer, some images will be compressed here, resulting in deviation in distance measurement. Therefore, before resetting the size, the length and width of the image will be calculated to decide whether to rotate the image 90°.
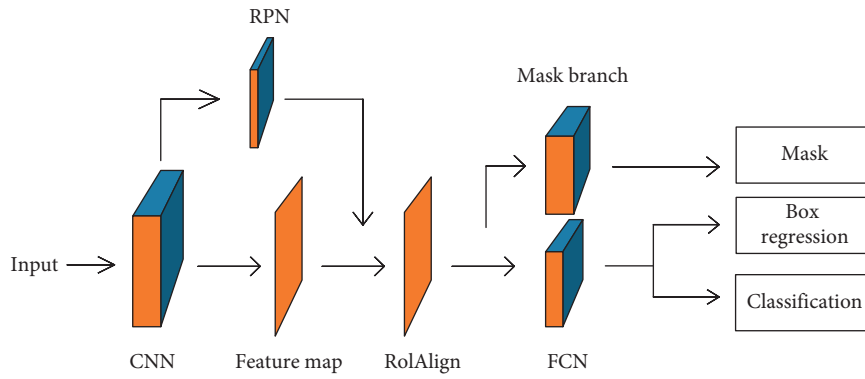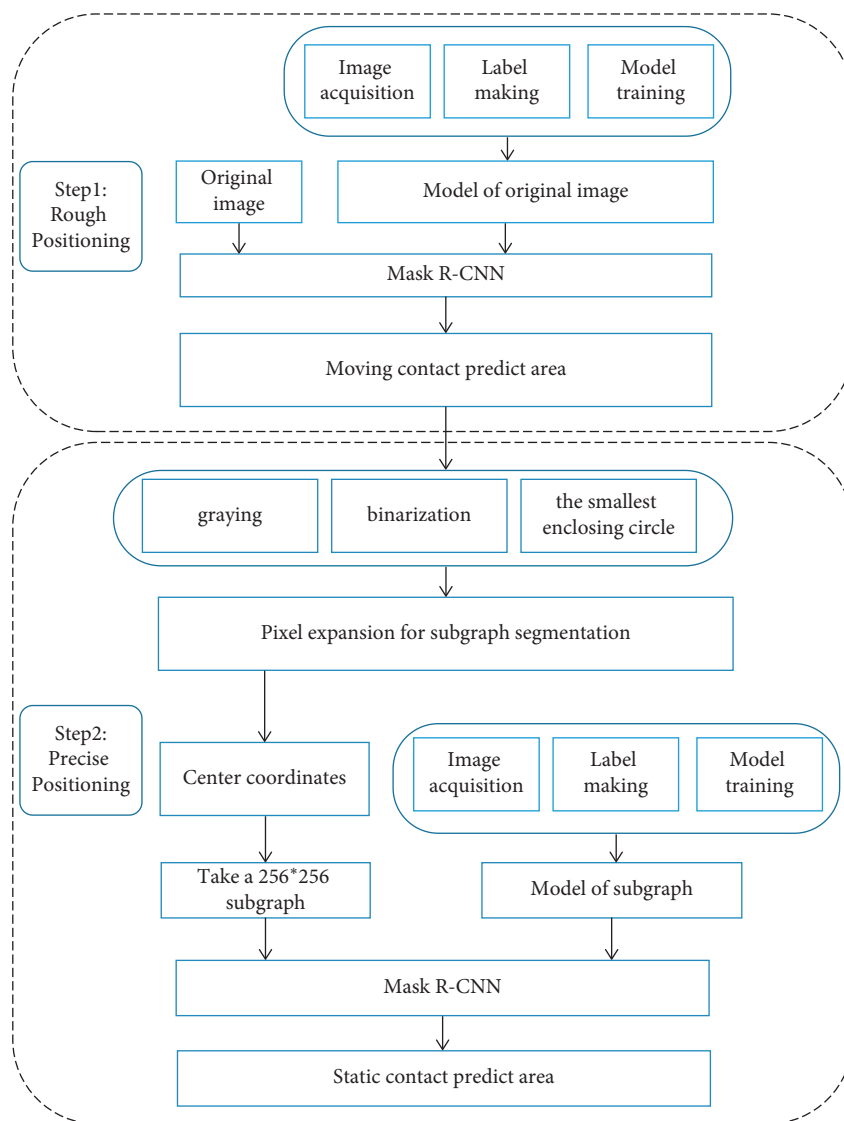
Figure 2: Overview of Mask R-CNN.

Figure 3: Framework of the positioning of switch machine moving and static contact based on DLM.

The pictures with fixed size and direction are put into the network, model 1 is loaded for testing, and the prediction of moving contact area is obtained, as shown in Figure 4. The black background is the schematic diagram of the original diagram, and the white part is the area of interest located by the Mask R-CNN in step 1, which is the moving contact area.
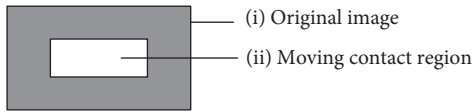
FIGURE 4: Moving contact area. The white area is the moving contact area. (i) The original image. (ii) The moving contact region.

3.3. Precise Positioning. Due to the large size of the image of the data set, the static contact accounts for fewer pixels in the image, so it is difficult to locate directly from the original image, which is easy to cause positioning offset and incomplete mask. In order to get the location of the static contact better, this experiment adopts the step-by-step method. After obtaining the location of the moving contact from the rough positioning, it carries out the preprocessing and subgraph segmentation and then predicts the location of the static contact from the subgraph. There are four steps in this method as follows:

(1) Gray-scale processing is performed on the prediction area of the moving contact mask. Because the color of the predicted area is different from the background, it is easy to segment the edge after graying.

(2) The gray image is binarized to form a black-and-white image, which is easy to find the smallest enclosing circle, as shown in Figure 5 binarization.

(3) After getting the binary image, the top, bottom, left, and right pixels of the white area are marked as A, B, C, and D, and the radius and center of the smallest circle C1 surrounding the four points are calculated, as shown in Figure 6(a). Then, repeat that operation once for each pixel. If there is a point e beyond the boundary of C1, the point e is added to A, B, C, and D. Continue to search for the smallest enclosing circle C2, as shown in Figure 6(b), until all the pixels do not have the point out of bounds, so that the center coordinates of the circle are the center coordinates of the moving contact.

(4) Taking the obtained center coordinates as the center to intercept the small graphs with a fixed size, the graph is the subgraph of subsequent processing. The processing process is shown in Figure 5, including graying, binarization, the smallest enclosing circle, and obtaining the final moving contact area.

The fixed area subgraph is intercepted by the moving contact column center coordinates obtained in the previous step, as shown in Figure 7(a). If the moving contact column coordinates obtained by the first model are too close to the image edge, the image area to be intercepted will exceed the limit, as shown in Figure 7(b), resulting in the program crash.

In order to solve this problem, some black pixels are filled on each edge of the image, and the moving contact center has moved corresponding pixels in two dimensions, which solves the problem that the center of the screenshot is close to the edge of the image and the screenshot fails. The effect after processing is shown in Figure 7(c). The subgraph

required for fine positioning can be successfully obtained through pixel expansion.

By sending the subgraph into the Mask R-CNN in the precise positioning step, we can get the accurate moving contact and static contact area. The accuracy of positioning determines the accuracy of contact depth measurement, so it is necessary to use the DLM method.

4. Experimental Result

4.1. Experimental Environment and Data Set Construction

4.1.1. Experimental Environment Configuration. The experimental environment used in this experiment is as follows: the operating system is Windows 10 professional 64 bit; the CPU is Intel i7-8700 @ 3.2 ghz; the GPU is NVIDIA GeForce RTX 2080; the deep learning framework is Tensorflow 1.5.0; the memory is 32 G; and the programming language is Python 3.5.

4.1.2. Data Set Construction. The data pictures used in this paper are from the switch machine pictures taken by manual mobile phones. Due to the different sizes of the pictures, it is necessary to unify the picture size to $960 * 1280$. In order to simulate the recognition rate under special working conditions under poor shooting conditions, some operations are taken, such as angle transformation, brightness transformation, and adding noise. The image states of the data sets are very different, so the positioning is challenging.

The training set of rough positioning consists of 500 internal images of the switch machine, and the training set of precise positioning consists of 1500 subgraphs. Each test set has 50 pictures, a total of 6 test sets. The method proposed in this paper is used to collect six times of positioning results, and the positioning accuracy of the moving and static contact of the switch machine is obtained.

This experiment adopts the transfer learning method, and the coco data are used as the pretraining model to speed up the convergence time of the model.

In this experiment, the key to judging whether it is a positive or negative sample is determined by the Intersection over Union (IoU). The judgment of whether the candidate frame anchor is a positive or negative sample is taken as an example. If the calibration threshold is 0.5, then calculate the IoU between each reference box and ground truth. If it exceeds the set threshold, it is a positive sample, otherwise a negative sample, where the ground truth is obtained from the coordinates of the upper left corner and the lower right corner of the mask in the training image label.

$$IoU_{anchor} = \frac{box_{anchor} \cap box_{GT}}{box_{anchor} \cup box_{GT}}, \qquad (2)$$

where box_{anchor} is the anchor reference box and box_{GT} is the ground truth.

Since this experiment involves the location of interest, mask segmentation, and target classification, three judgment thresholds are set. The specific values are shown in Table 1.
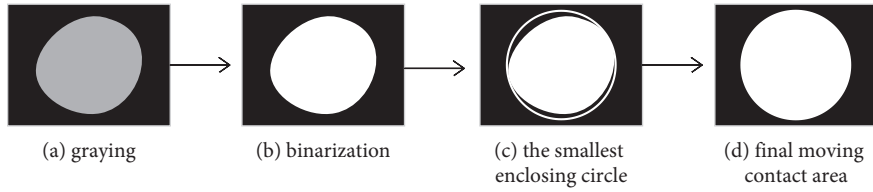
(a) graying   (b) binarization   (c) the smallest   (d) final moving
                                                    enclosing circle   contact area

FIGURE 5: Image preprocessing.

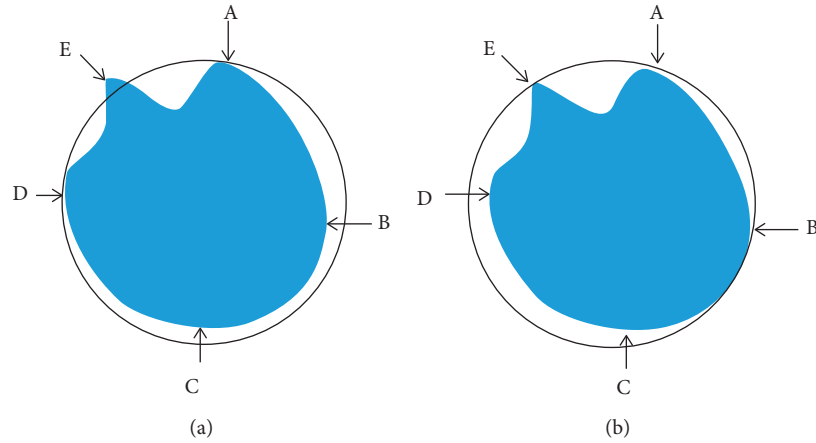(a)   (b)

FIGURE 6: The smallest enclosing circle: (a) the schematic diagram of the first encirclement; (b) the schematic diagram of the second encirclement. A, B, C, D, and E are the end points of the graph, respectively.

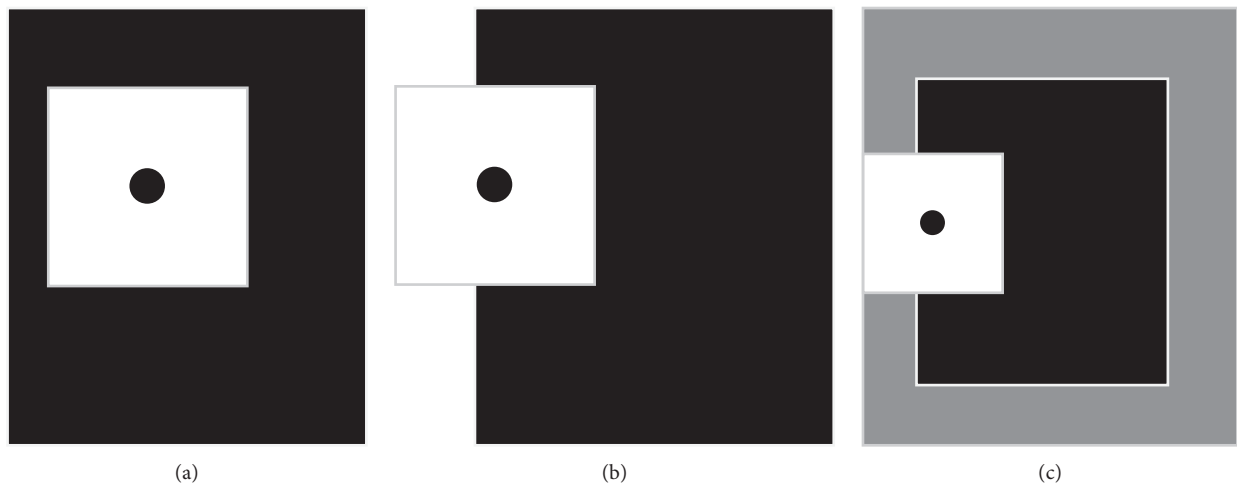(a)   (b)   (c)

FIGURE 7: Pixel expansion. (a) A fixed-size subimage with the center of the circle as the center is taken. (b) The center of the circle is too close to the boundary. (c) After filling, there will be no out of bounds.

TABLE 1: The specific threshold data of the experiment.

Experiment part	Evaluation index	Specific threshold
Anchor	The IoU of anchor and ground truth	0.7
RoI	The IoU of RoI and ground truth	0.5
Mask segmentation	The IoU of reference mask and true mask	0.5

4.2. Analysis of Single-Layer Mask R-CNN Positioning Results. In order to verify the high accuracy and robustness of the DLM method, the single location method is used to design the comparative experiment in this paper. Only single-layer Mask R-CNN is used to locate the contact of switch machine picture.

The moving contact area obtained by single positioning is very accurate, which is similar to the rough positioning method in the DLM method. However, for the positioning of the static contact, because the target area is small and the original image is a 960 ∗ 1280 large pixel image, the static contact area is hard to determine. The positioning effect is poorer than the DLM method in this article. As shown in Figure 8, the mask of the moving contact is accurate, but the mask of the static contact is incomplete or even disappeared.
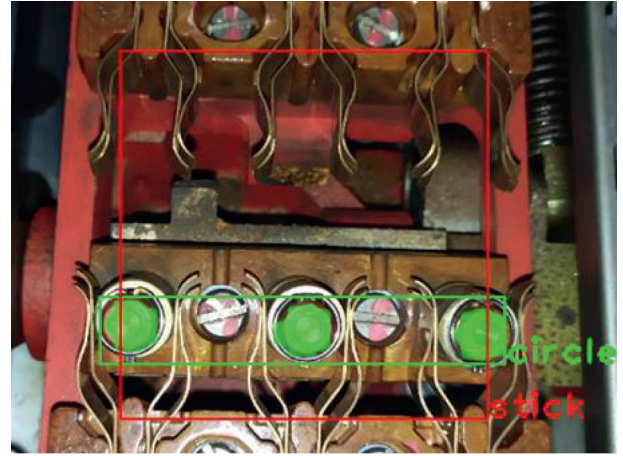

FIGURE 8: The single-layer Mask R-CNN positioning result.

4.3. Analysis of Rough Positioning Results

4.3.1. Rough Positioning of Switch Machine Moving Contact. The image is input to be predicted into the Mask R-CNN in step 1 to get the mask segmentation image. The comparison of the moving contact positioning effect is shown in Figure 9. Figure 9(a) is the original image captured by the mobile phone, and Figure 9(b) is the result of locating and labeling the original image by the Mask R-CNN.

The region of interest of moving contact obtained in Figure 9(b) is processed before segmentation. After graying, binarization, and the smallest enclosing circle, the circle obtained by experiment has high overlap with the original image. The specific experimental picture is shown in Figure 10.

It can be seen from Figure 10 that the method mentioned in Section 3.3 can be used to locate the moving contact under extreme conditions, and the mask covers a complete area without causing the center of the circle to shift.

4.3.2. Subgraph Acquisition and Region Recognition. Before the region segmentation of the moving contact in the original image, in order to prevent the segmentation failure, the pixel expansion is carried out first. The results show that it is effective to expand 200 black pixels in the two dimensions, and 200 pixels are added to the corresponding coordinates in the segmentation process. The size of the subgraph will not change after segmentation, as shown in Figure 11.

According to the center coordinates of each moving contact after processing the target area, the subgraphs are segmented, and the segmented images are shown in Figure 12. It can be seen that all the subgraphs contain the parts needed for contact depth measurement.

4.4. Analysis of Precise Positioning Results. The subgraph obtained by rough positioning is input into the second stage Mask R-CNN to obtain the region of interest of the static contact, which is an important basis for the subsequent calculation of contact depth.

Among them, the segmentation of the moving contact in the original picture is shown in Figure 13(a). Figure 13(b) shows the key areas around the moving contact identified by the precise positioning step, including the tape spring and the static contact. The area of the tape spring and the area of static contact are shown in Figures 13(c) and 13(d). By calculating the distance between the center of the cylinder and the region of interest of the paddle and the static contact, the contact depth of the moving and static contact can be calculated.

The multitask loss function curve of Mask R-CNN is shown in Figure 14.

The change process of the loss function value of the subgraph model is shown in Figure 14. The number of training sessions is represented by *x*-axis, and the loss function value is represented by *y*-axis. Figures 14(a)–14(d) show the loss function, classification branch loss function, mask branch loss function, and positioning branch loss function, respectively. It can be seen from Figure 14 that the losses tend to converge with the increase in training times. There are jump forms in the middle of each graph because there is a mode transition from only training network heads to global training. Although the curve of loss function shows that it converges all the time, sometimes when the training times reach a certain value, the loss function mutation will appear, so we need to refer to the trend chart of loss function when selecting the model.

The accuracy of the positioning of the moving and static contacts is set in this paper. True of moving contact (T_M) indicates the number of samples with correct positioning of moving contact. True of static contact (T_S) indicates the number of samples with correct positioning of static contact. Accuracy of moving contact (A_M) indicates the positioning accuracy of moving contact. Accuracy of static contact (A_S) indicates the positioning accuracy of static contact. Each original image has three static contacts, so in this experiment, only when the three static contacts in one image are positioned can they be regarded as correctly. A total of six sample groups were set up, and the number of samples in each group was 50. The DLM positioning result data are shown in Table 2.

(a)   (b)
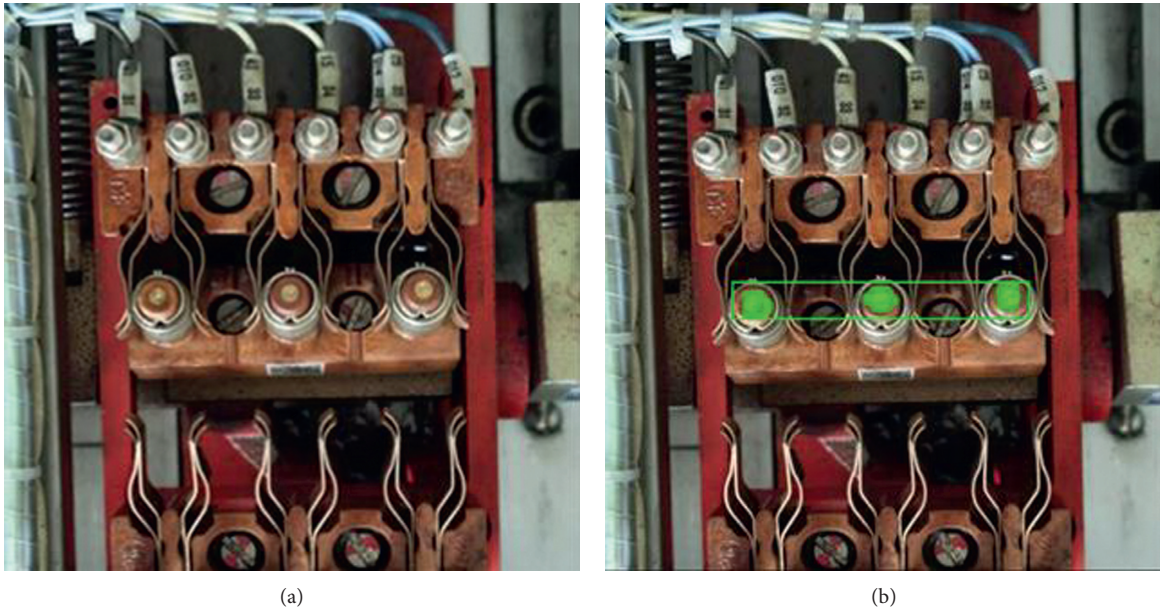
FIGURE 9: The comparison of detection result: (a) the original image; (b) the rough positioning image.

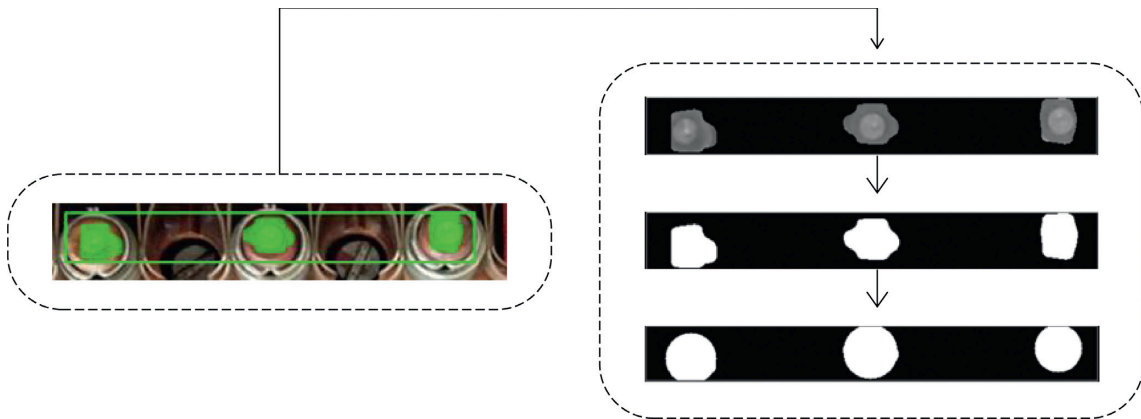FIGURE 10: The treatment process of moving contact.

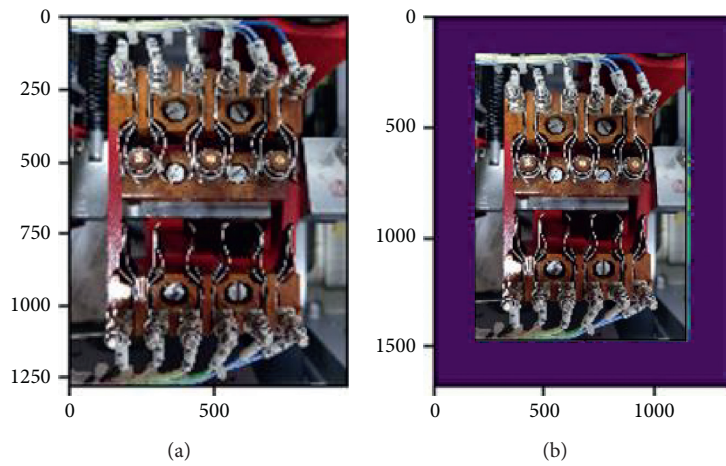(a)   (b)

FIGURE 11: The pixel expansion effect: (a) the original image; (b) the image after expansion.
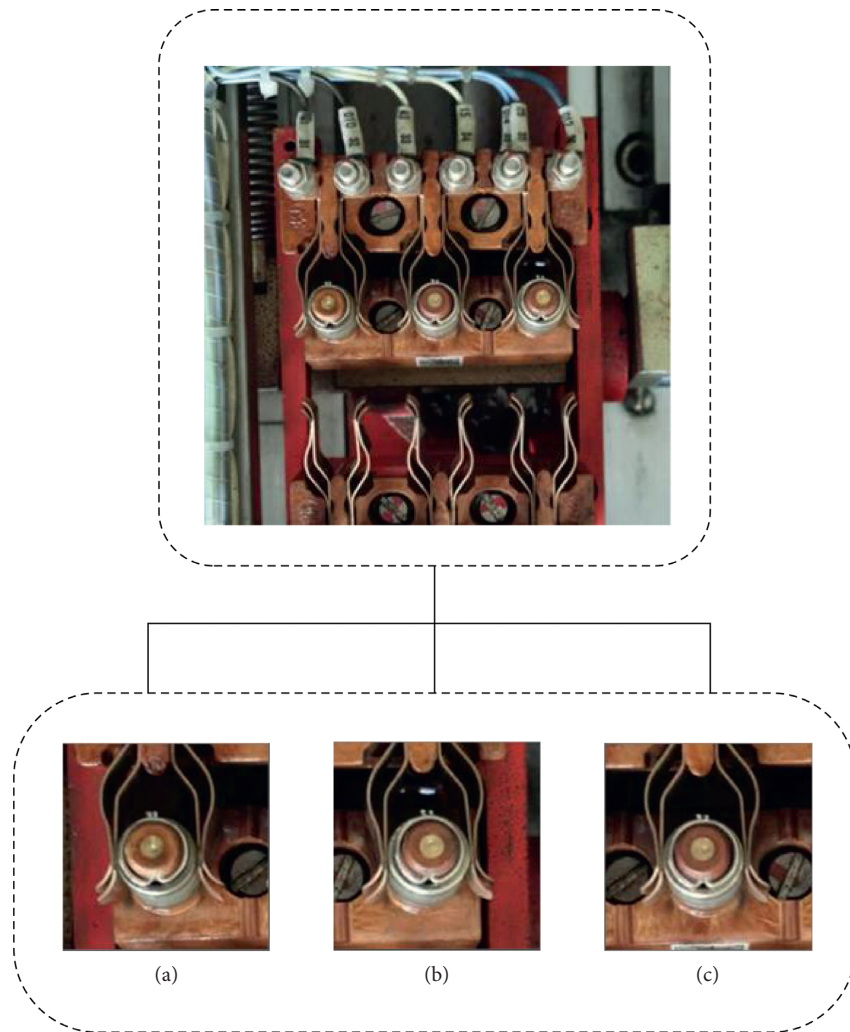
Figure 12: The subgraph segmentation: (a) the first moving contact; (b) the second moving contact; (c) the third moving contact.
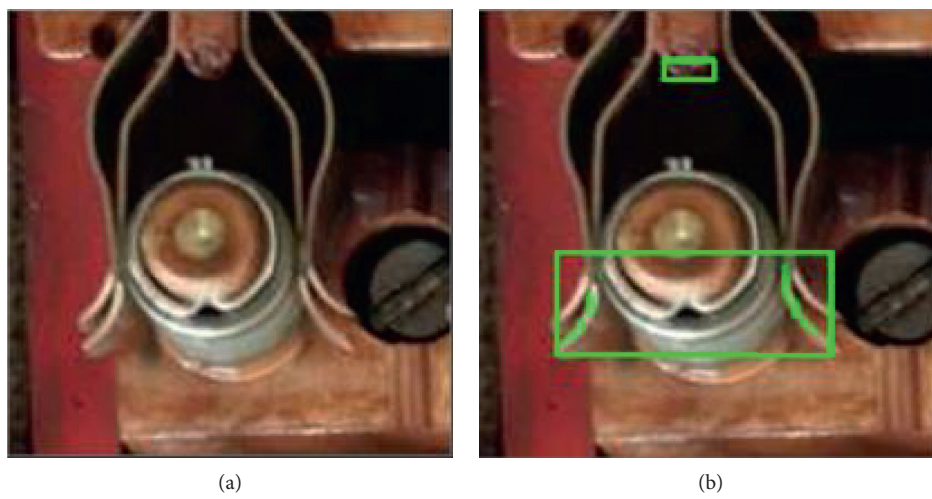

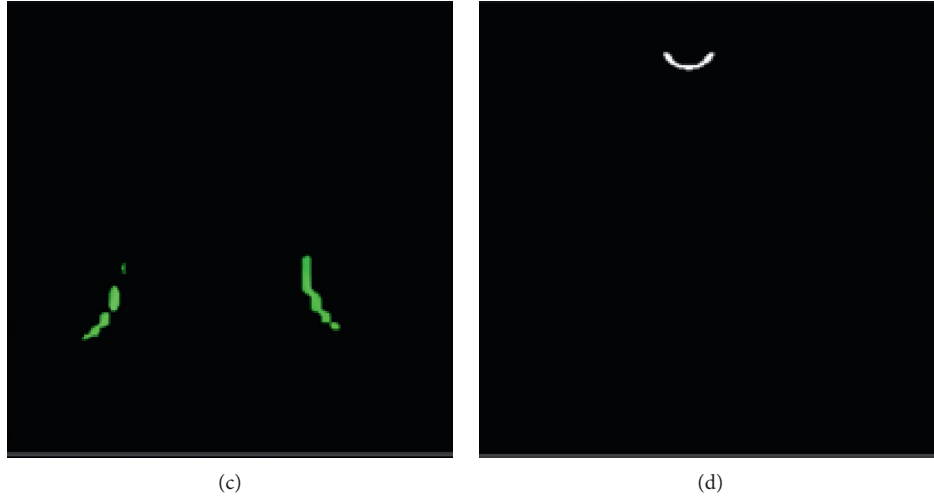(a)   (b)

Figure 13: Continued.

(c)

(d)

FIGURE 13: The precise positioning result.
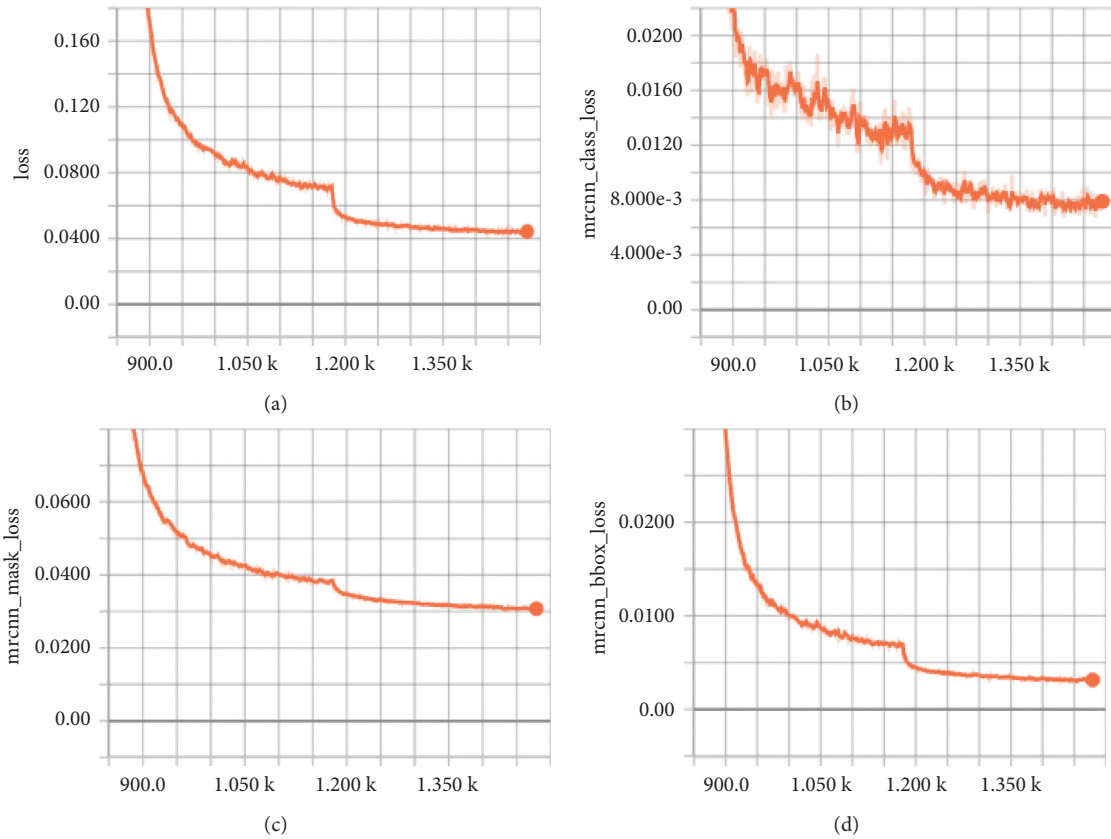

(a)

(b)

(c)

(d)

FIGURE 14: Plot of the loss functions: (a) loss; (b) class_loss; (c) mask_loss; (d) bbox_loss.

$$A_M = \frac{T_M}{50} \times 100\%,$$

$$A_S = \frac{T_S}{50} \times 100\%. \tag{3}$$

And as shown in Figure 15, the number of groups is represented by x-axis, the accuracy is represented by y-axis, the positioning result of moving contact is counted by the red line, and the positioning result of static contact is counted by the blue line. According to the results of DLM, it can be seen that the static contact is more difficult to locate than moving contact, and there is a positive correlation between the positioning accuracy of static contact and that of moving contact. After SPSS software correlation calculation, R value is greater than 0, equal to 0.955 and close to 1, which is means that the two groups of data

TABLE 2: The DLM positioning result data.

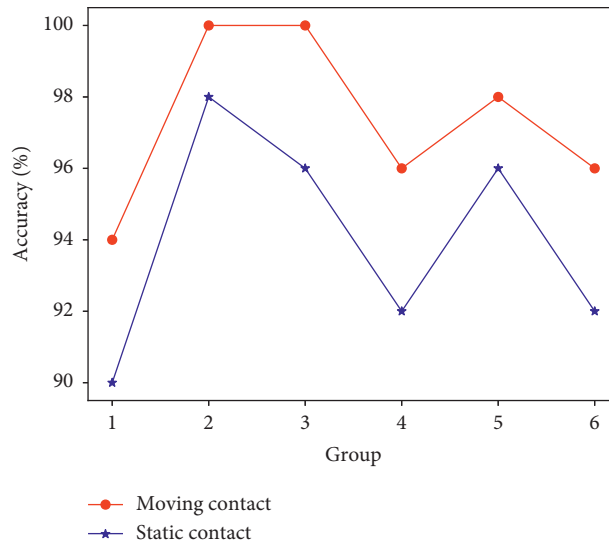| Group	T_M (piece)	T_S (piece)	A_M (%)	A_S (%)
1	47	45	94	90
2	50	49	100	98
3	50	48	100	96
4	48	46	96	92
5	49	48	98	96
6	48	46	96	92
Avg	48.67	47	97.33	94

FIGURE 15: Plot of DLM positioning accuracy.

are positively correlated, and the correlation is strong, P value is 0.02, less than 0.05, so the two groups of data obtained from the experiment have a significant positive correlation. This also shows that in the two-step positioning process, the positioning results of each step will have an important impact on the subsequent experiments.

Table 3 shows the positioning results of the three methods for the moving and static contact of the switch, respectively. Because Yolo and R-CNN are commonly used in the target location algorithm, we compare them to increase the reliability of the article. The number of T_M and T_S is the sum of six groups of results. The data can be seen intuitively in Figure 16.

As can be seen from Figure 16, there is no big difference in the positioning accuracy of moving contacts among DLM, Yolo, and Mask R-CNN, and their accuracy rate is 97.33%, 97.67%, and 95.67%. However, the accuracy of static contacts is quite different, which means that static contacts with smaller pixels in the original image are not easy to be located. The accuracy rate of static contact positioning based on the DLM is 94%, while the Yolo is 62.33% and Mask R-CNN is 5%. The DLM algorithm can be used to solve the problem of static contact positioning accuracy. At present, the algorithm has reached a high detection accuracy, which can reach 94 percent under normal conditions, and it also has a good detection effect on pictures under severe working conditions.

TABLE 3: Comparison of positioning accuracy.

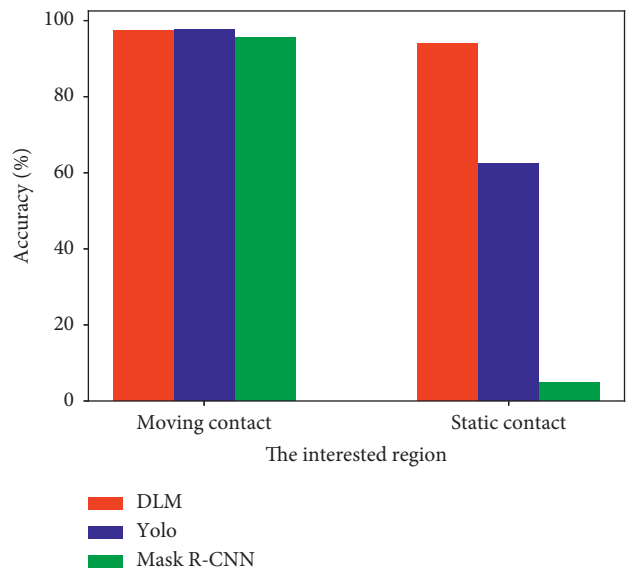| Method	T_M (piece)	T_S (piece)	A_M (%)	A_S (%)
DLM	292	282	97.33	94
Yolo v4	293	187	97.67	62.33
Mask R-CNN	287	15	95.67	5

FIGURE 16: Comparison of positioning accuracy of moving and static contacts by DLM, Yolo, and Mask R-CNN method.

At the same time, it can be found from Figure 16 that Yolo is better than Mask R-CNN in the one-step positioning of static contact. This is because the Yolo positioning result is a rectangular frame, in which there are static points and their adjacent areas. However, Mask R-CNN can accurately locate to the pixel level, so the detection area is much smaller than Yolo, resulting in the difference of positioning effect.

5. Conclusion

In this paper, the DLM is proposed to locate the moving and static contact of switch machine. First, rough positioning is carried out to obtain the moving contact area in the original image for subsequent measurement; second, the precise positioning is carried out. The subgraph obtained from rough positioning is preprocessed, and then secondary positioning is carried out to improve the detection accuracy of the target area. Through two positioning, the static contact area can be accurately obtained, improving the accuracy of automatic contact depth measurement. The experimental results show that DLM can automatically locate the internal parts of switch machine box in batch, the positioning accuracy of static contacts has been greatly improved, and the robustness is good.

The positioning of the switch machine moving and static contact based on DLM better promotes the automation of rail transit maintenance work, reduces the work intensity of maintenance personnel, and provides a reference for further research on future inspection work. The inspection method proposed in this article can also be applied to a variety of industrial scenarios to improve operation efficiency. The DLM method of this paper has the following points to be improved:

(1) The DLM is a two-step positioning method, which takes a long time compared with one-step positioning, so the detection efficiency needs to be improved

(2) Before Mask R-CNN training, a large number of labels need to be produced manually, and the training effect is related to label production

(3) The angle correction algorithm can effectively solve the angle interference caused by the portable device

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Z. Lv, A. Ma, X. Chen et al., "Defect detection of pandrol track fastener based on local depth feature fusion network," *Complexity*, vol. 2021, Article ID 6687146, 9 pages, 2021.

[2] Y. Lei, H. Reza Karimi, L. Cen et al., "Processes soft modeling based on stacked autoencoders and wavelet extreme learning machine for aluminum plant-wide application," *Control Engineering Practice*, vol. 108, 2021.

[3] S. Shin, M.-H. Kim, and S. B. Choi, "Ultrasonic distance measurement method with crosstalk rejection at high measurement rate," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 4, pp. 972–979, 2019.

[4] X. Chen, J. Xu, H. Chen, H. Ding, and J. Xie, "High-accuracy ultrasonic rangefinders via pMUTs arrays using multi-frequency continuous waves," *Journal of Microelectromechanical Systems*, vol. 28, no. 4, pp. 634–642, 2019.

[5] M. Veng, J. Perchoux, and F. Bony, "Fringe disappearance in self-mixing interferometry laser sensors: model and application to the absolute distance measurement scheme," *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5521–5528, 2019.

[6] Z. Zhang, X. Yang, and H. Gao, "Weighted smallest deformation similarity for NN-based template matching," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 6787–6795, 2020.

[7] F. Liu, C. Gong, X. Huang, T. Zhou, J. Yang, and D. Tao, "Robust visual tracking revisited: from correlation filter to template matching," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2777–2790, 2018.

[8] Z. Zhenhua Guo, L. Lei Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

[9] M. A. Naiel, M. O. Ahmad, and M. N. S. Swamy, "A vehicle detection scheme based on two-dimensional HOG features in the DFT and DCT domains," *Multidimensional Systems and Signal Processing*, vol. 30, no. 4, pp. 1697–1729, 2019.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, San Diego, CA, USA, June 2005.

[11] S.-K. Pavani, D. Delgado-Gomez, and A. F. Frangi, "Gaussian weak classifiers based on co-occurring Haar-like features for face detection," *Pattern Analysis and Applications*, vol. 17, no. 2, pp. 431–439, 2014.

[12] C. Lohrmann, P. Luukka, M. Jablonska-Sabuka, and T. Kauranne, "A combination of fuzzy similarity measures and fuzzy entropy measures for supervised feature selection," *Expert Systems with Applications*, vol. 110, pp. 216–236, 2018.

[13] L. P. Valem, C. R. D. Oliveira, D. C. G. Pedronette, and J. Almeida, "Unsupervised similarity learning through rank correlation and kNN sets," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 14, no. 4, pp. 1–23, 2018.

[14] X. Yin, M. Yin, L. Xie, and G. Yin, "Automatic measurement system of section geometry parameters for composite filter rods based on machine vision," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 5037–5050, 2020.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[17] C. Szegedy, "Going deeper with convolutions," in *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.

[18] J. Redmon, S. Divvala, R. Girshick et al., "You only Look once: unified, real-time object detection," in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, June 2016.

[19] R. Girshick, "Fast R-CNN," in *Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, December 2015.

[20] S. Ren, K. He, R. Girshick et al., "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[21] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2020.

[22] I. Ahmed, S. Din, G. Jeon et al., "Exploring deep learning models for overhead view multiple object detection," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5737–5744, 2019.

[23] A. M. M. Sizkouhi, M. Aghaei, S. M. Esmailifar, M. R. Mohammadi, and F. Grimaccia, "Automatic boundary extraction of large-scale photovoltaic plants using a fully convolutional network on aerial imagery," *IEEE Journal of Photovoltaics*, vol. 10, no. 4, pp. 1061–1067, 2020.

[24] L. Xiao, B. Wu, and Y. Hu, "Surface defect detection using image pyramid," *IEEE Sensors Journal*, vol. 20, no. 13, pp. 7181–7188, 2020.