2020 25th International Conference on Pattern Recognition (ICPR)
Milan, Italy, Jan 10-15, 2021

Mobile Phone Surface Defect Detection Based on Improved Faster R-CNN

Tao Wang, Can Zhang, Runwei Ding, Ge Yang
Key Laboratory of Machine Perception
Shenzhen Graduate School, Peking University
Shenzhen, China
email:taowang@stu.pku.edu.cn, can.zhang@pku.edu.cn, dingrunwei@pku.edu.cn, yangge@szpku.edu.cn

***Abstract*—Various surface defects will inevitably occur in the production process of mobile phones, which have a huge impact on the enterprise. Therefore, precise defect detection is of great significance in the production of mobile phones. However, traditional manual inspection and machine vision inspection have low efficiency and accuracy respectively which cannot meet the rapid production needs of modern enterprises. In this paper, we proposed a mobile phone surface defect (MPSD) detection model based on deep learning, which greatly reduces the requirement of a large dataset and improves detection performance. First, Boundary Equilibrium Generative Adversarial Networks (BEGAN) is used to generate and augment the defect data. Then, based on the Faster R-CNN model, Feature Pyramid Network (FPN) and ResNet 101 are combined as feature extraction network to get more small target defect features. Further, replacing the ROI pooling layer with an ROI Align layer reduces the quantization deviation during the pooling process. Finally, we train and evaluate our model on our own dataset. The experimental results indicate that compared with some traditional methods based on handcrafted feature extraction and the traditional Faster R-CNN, the improved Faster R-CNN achieves 99.43% mAP, which is more effective in the field of MPSD defect detection.**

***Keywords*—Surface defect detection, BEGAN, Faster R-CNN, FPN, RoI Align**

I. Introduction

Mobile phone surface defect (MPSD) is an inevitable factor in the production process of mobile phones. Efficient defect detection can provide enterprises with production information in time to improve production technology. At present, manual detection is still the main method in the production process of mobile phones. Due to human body fatigue, slow detection speed, and other factors of manual detection, it cannot meet the rapid production needs of current enterprises. Therefore, the surface defect detection technology based on machine vision stands out in order to solve the problems of manual inspection. Vision-based detection technology uses a visual sensor instead of the human eye to acquire the detected object image and uses detection algorithms to extract information from the image to determine whether there exists a defect in the collected image.

At present, a large number of defect detection researches are based on traditional machine vision. Jian et al. [1] studied the defect discrimination method and proposed a joint defect discrimination method based on difference projection and an improved fuzzy C-means clustering (IFCM) algorithm. The former eliminates the influence of external light changes on the gray-scale of the image to be measured, and the latter removes the defect of the blurred gray border in the noise image of the mobile phone screen. However, the defect detection is very dependent on the template image. Jie Zhao et al. [2] proposed a new method to detect and identify glass defects in low-resolution images, that is, using binary feature histograms (BFH) to describe the characteristics of glass defects, compared with the local binary pattern (LBP) [3], the accuracy and speed have been improved. Zhang [4] et al. introduced a discrete Fourier transform (DFT) and optimal threshold into defect detection determined the position of the defect and highlighted it through the spectral residual method in DFT, and finally determined the optimal threshold for defect region segmentation through multiple iterations. Huang et al. [5] proposed a complete Mobile phone Panel Surface Defect Detection framework based on machine vision, which is composed of different feature extraction operators and support vector machine (SVM) [6] classifiers. These methods meet the requirements in terms of accuracy and speed, so they have been applied in the actual production process. However, these methods deeply rely on hand-crafted features, which require experienced experts, and these features are highly targeted. The traditional algorithm's feature extraction ability will greatly decrease if the product materials or production environment change.

Over the recent years, deep learning technology has developed rapidly, and the neural network model in deep learning has been successfully applied and achieved good results in the fields of target classification, target recognition, target tracking, and autonomous driving [7], [8], [9], [10]. A significant advantage of deep learning technology is automatic feature extraction, which can eliminate the tedious image processing and overcome the shortcomings of traditional machine vision detection methods. Many researchers have begun to use deep learning methods to solve the problem of defect detection. Zhang et al. [11] proposed a convolutional neural network (CNN)-based method to detect printed circuit board (PCB) defects, which achieves high detection performance compared with traditional detection methods. Xu et al. [12] proposed an improved deep convolutional network to detect railway subgrade defects. Compared with the traditional HOG+SVM method, improved deep convolutional networks can achieve better performance. Titov et al. [13] embedded the YOLO

978-1-7281-8808-9/20/$31.00 ©2020 IEEE

2020 25th International Conference on Pattern Recognition (ICPR) | 978-1-7281-8808-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPR48806.2021.9412119

algorithm into the unmanned aerial vehicle (UAV) to detect the defects of power lines. He et al. [14] proposed a regression and classification based framework for industrial surface defect detection, and finally achieved the state-of-the-art performance. In summary, the deep learning method has been introduced in the defect detection field and it also achieves great performance. However, it is rarely expanded into the field of MPSD detection.

The purpose of this paper is to explore the possibility of deep learning method in surface defects of mobile phones, especially with limited data samples. In this study, we propose a scheme of applying Faster R-CNN [15] for surface defects of mobile phone. The main contributions of this article are as follows:

First, it is difficult to obtain samples of mobile phone surface defects, so we use BEGAN to augment the dataset, making the data more diverse while ensuring quality. Then, in order to solve the problem of small target defect detection on the mobile phone surface, FPN network [16] and ResNet101 [17] feature extraction network are combined to obtain more powerful feature extraction capabilities. Further, ROI Align layer [18] replaces the original ROI Pooling layer to reduce pixel loss due to quantization. At last, a set of comparative experiments are made with traditional Faster R-CNN. In addition, traditional methods based on feature descriptors(e. g. SIFT [19], LBP [3] and HOG [20]) and SVM are also taken as a reference. The experimental results verify the validity of the proposed scheme.

II. MPSD Detection With Improved Faster R-CNN

A. Faster R-CNN

Faster R-CNN is an object detection and recognition algorithm proposed by Ren et al [15]. This algorithm is the final improved version of the region-convolutional neural network (R-CNN) series of algorithms. Its most prominent contribution is the proposal of a regional proposal network (RPN), compared with the selective search algorithm in the previous generations of R-CNN algorithms, the speed has been improved by nearly 200 times, and the accuracy has also been greatly improved. Faster R-CNN is mainly composed of three parts: feature extraction network, regional suggestion network, and classification regression network. These three parts have different functions and cooperate with each other to achieve the detection task.

The main task of the feature extraction network is to use the convolutional neural network to obtain the feature maps of the input image, which contains rich semantic information. Traditional Faster R-CNN uses VGG16 [21] as a feature extraction network. When the picture is input to the network, the 13-layer learnable convolution kernel performs feature extraction on the image.

The RPN network accepts the feature maps obtained by the feature extraction network for preliminary feature selection. The RPN network has two branches (i.e. $rpn-cls$, $rpn-reg$). $rpn-reg$ is the regression branch of the bounding box. The main function of $rpn-reg$ branch is to distribute the generated boxes near the actual target box through the regression and transfer coordinate of generated boxes to the classification regression network. $rpn-cls$ is a branch of the bounding box classification. This branch will perform a preliminary two-class classification through softmax. Because the RPN network involves both classification and regression tasks, its loss function is determined by the classification loss and the regression loss. Loss functions are composed of two parts:

$$L_{cls}(P_i, P_i^*) = -log[P_i P_i^* + (1 - P_i^*)(1 - P_i)] \quad (1)$$

$$L_{reg}(t_i, t_i^*) = \sum_{i \in (x,y,w,h)} Smooth_{L1}(t_i - t_i^*) \quad (2)$$

$$Smooth_{L1}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \quad (3)$$

$$\begin{aligned} L(P_i, P_i^*) = & \frac{1}{N_{cls}} \sum_i L_{cls}(P_i, P_i^*) \\ & + \lambda \frac{1}{N_{reg}} \sum_i P_i^* L_{reg}(t_i, t_i^*) \end{aligned} \quad (4)$$

where N_{cls} is the batch size of the input image, N_{reg} is the total number of anchor boxes participating in the training, λ is defined as the batch size divided by the total number of anchor boxes, which is used to balance the batch size and the number of anchor boxes, t_i is a 4-dimensional vector, which is the location information of the bounding box predicted by the RPN network, t_i^* is the position information of the relevant real target box, P_i is the probability that the initial suggestion box is predicted by the RPN network as a prospect, P_i^* is the label value of the foreground.

Ultimately, the classification regression network performs accurate target classification and target positioning. The structure diagram of Faster R-CNN is shown in Figure 1.

B. The Overall Architecture of the Proposed Method

The Faster R-CNN method directly applied to the detection of surface defects of mobile phones cannot achieve the ideal detection effect. The main reasons are as follow:

- The similarity between the defect target and the detection background is very high
- The size of the defects are often small
- The shape of defects are diverse
- The ratio of positive and negative samples is unbalanced
- The data sample is limited

Therefore, to deal with the problems in the detection of MPSD, this article has improved several parts of the original Faster R-CNN. The architecture diagram of the improved network is shown in Figure 2.

As shown in Figure 2, this paper mainly improves the original Faster R-CNN from the following three aspects:

- The VGG16 will be replaced by the ResNet-101 as a feature extraction network to extract high-quality feature information.
- Feature Pyramid network (FPN) is embedded in Faster R-CNN to enhance the detection ability of small targets.

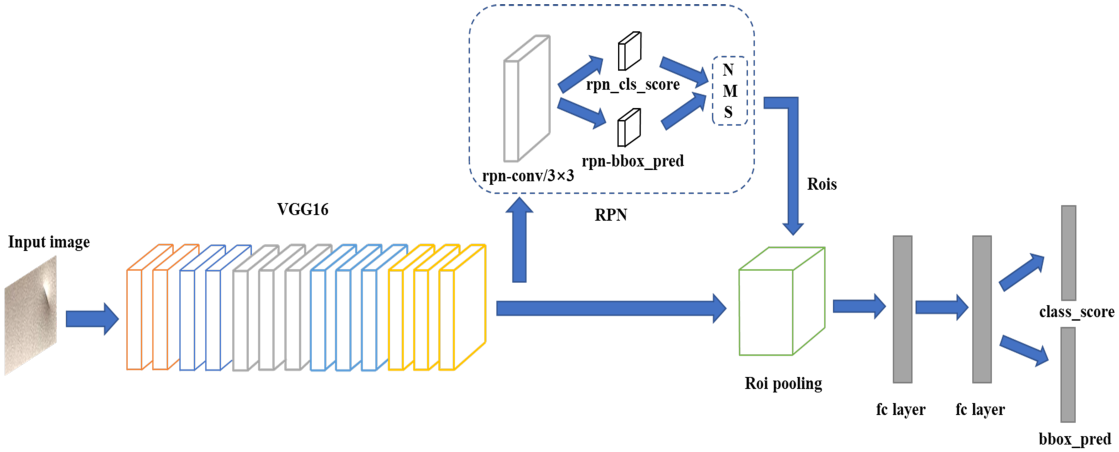


Fig. 1. The architecture of Faster R-CNN.

- The RoI Pooling layer is replaced by the RoI Align layer to reduce the detection capability degradation caused by quantization loss.
- Some common data augmentation measures have been implemented. At the same time, BEGAN is also used to generate high-quality and diverse data to reduce the impact of few samples on training.

C. Base Network

We adopt Resnet101 as a feature extraction network for our model. There are 5 convolutional blocks in total. The first convolution block is composed of a 7×7 convolutional layer with a stride of 2 and a 3×3 maximum pooling layer with a stride of 2. Each remaining convolutional block is composed of several bottlenecks Block composition. The network has a total of 101 layers of convolutions, and the output after each convolutional layer will be activated by batch normalization [22]. The ResNet model is pre-trained on ImageNet to classify 1.2 million pictures. However, there are only a few thousand datasets in this paper. Using such a deep network may lead to overfitting. Therefore, this paper uses the method of transfer learning and data augmentation to avoid overfitting. A pre-trained model is used to extract features from the augmented data, which reduces requirements on image quantity.

D. Feature Pyramid Network

The original Faster R-CNN uses the last single feature map of the feature extraction network for classification and regression, whereas its resolution is greatly lost in the convolution process. Therefore, the original Faster R-CNN detection capability cannot reach the expected performance. We analyzed the defect dataset and calculated the proportion of defects in each defect picture. A statistics of the dataset we made show that 58% of the defects are very small, accounting for only 0.5% of the image size. So we introduce the FPN network to cope with this hard problem. T. Lin [16] et al. proposed a multi-scale feature fusion method called feature pyramid network (FPN). FPN constructed a network structure with strong semantic expression capabilities at all scales, whose specific implementation is shown in Figure 2. FPN consists of three parts, a bottom-up structure, a top-down structure, and a lateral connection structure. This article uses ResNet101 as a feature extraction network. Four output layers in ResNet101 are defined as $C2$, $C3$, $C4$, $C5$, which constitutes the bottom-up network structure. The $P5$ layer is obtained by convolution of $C5$, and the feature map with the same size as the $C4$ layer can be obtained after the nearest neighbor upsampling of the $P5$ layer. Then, the $C4$ layer and the $P5$ layer are added to obtain the $P4$ layer. By analogy, the $P2$, $P3$, $P4$, $P5$ top-down structure will be generated. The function of lateral connection is to change the channels through a 1×1 convolution kernel to match the dimensions of the previous layer.

The embedding of the FPN network makes the output of the RPN network multi-scale, and its ROI area is also multi-scale. Therefore, ROIs of different scales need to use different feature layers as input to the ROI pooling layer. The specific distribution method of the feature layer is as follows:

$$k = k_0 + log_2(\frac{\sqrt{wh}}{224}) \tag{5}$$

where k_0 is the reference value, which is generally set to 5, w and h are the width and height of the ROI area, 244 is the size of the pre-training image based on ImageNet, k is the assigned layer Serial number.

E. RoI Align

ROI Align is a method proposed by the Facebook AI Research Department. Compared with the ROI pooling method, the ROI Align method cancels all quantization operations. Four sampling points are set in each bin, and the pixel value of the feature map is obtained by the method of bilinear interpolation, thereby avoiding the loss of accuracy caused by the quantization process. The operation process of RoI Align is as follows:

- First, each proposal region is traversed, keeping the floating-point boundary not quantized.

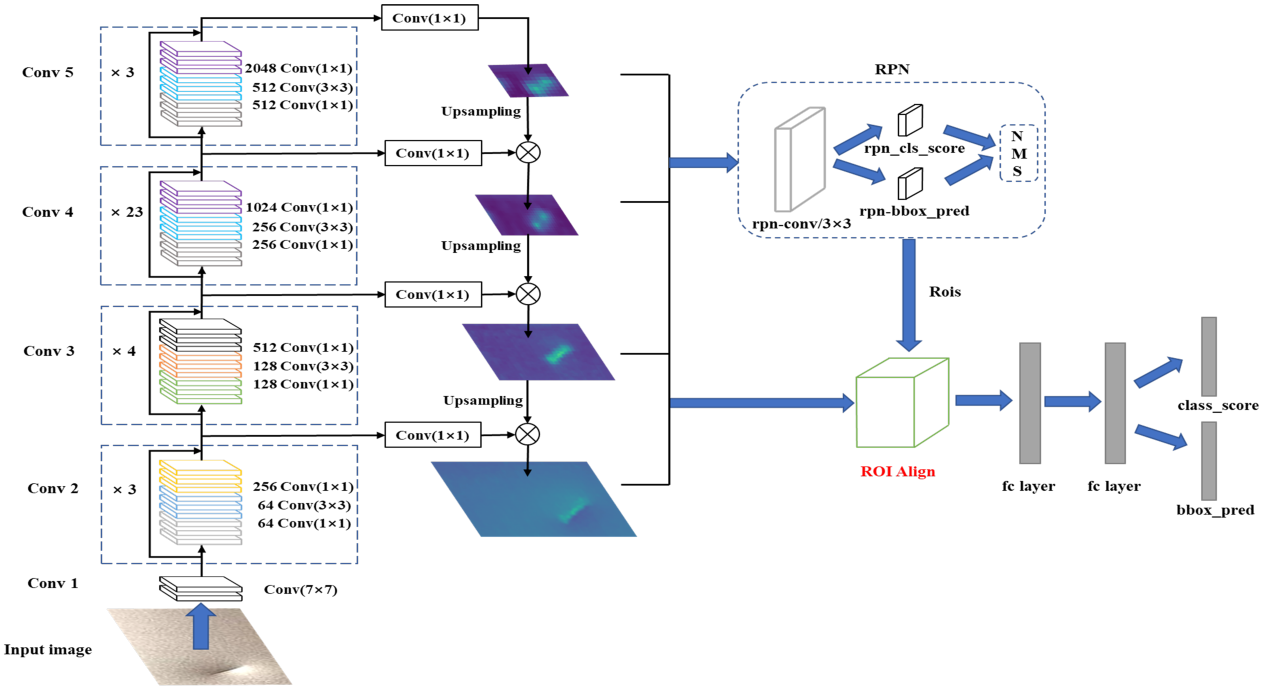
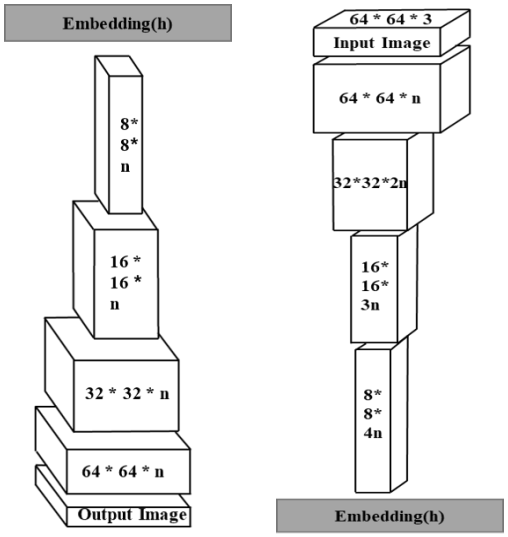


Fig. 2. The architecture of proposed method.

- Then the proposal region is divided into several rectangular bins, and the boundaries of each bin are not quantized.
- Further, four sampling points are set up in each bin and their coordinate positions are obtained. The pixel values of these four sampling points are calculated by bilinear interpolation. Finally, the maximum pooling operation is performed to fix the feature size.

F. BEGAN

Despite the dataset can be increased through image flipping, rotation and color jitter, the data sample is still very small and the defect feature information has not changed much. Therefore, BEGAN [23] network is applied to generate a large number of images, which have completely new defect features. These rich feature information can help the algorithm to adapt to the real production scenario to a certain extent. BEGAN network is different from other GANs, the discriminator here uses the auto-encoder structure, which makes a very simple network. Even without adding some training tricks such as BN, mini-batch, and SELU activation function, it can also achieve very good training results. Unlike typical GANs whose data distribution generated by the generator is as close as possible to the distribution of real data, BEGAN aims to match the auto-encoder loss distribution and the error loss distribution derived from the Wasserstein distance. Besides, A hyper-parameter γ is provided, this hyper-parameter can balance the diversity of the image and the quality of the generation. By controlling the hyper-parameter, a variety of defect features can be obtained, which is why we use BEGAN. The generator and discriminator of BEGAN are shown in Figure 3.



(a) Decoder (b) Encoder

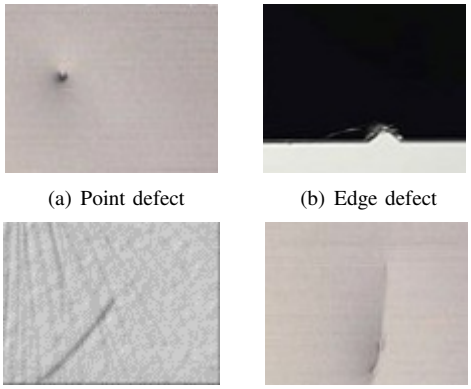
Fig. 3. The architecture of BEGAN

III. EXPERIMENT

A. Building Defect Dataset

Mobile phone surface defects include mobile phone screen defects and mobile phone shell defects. This paper collects 50 defective screens and 30 defective mobile phone cases. The samples are taken at different angles and light sources. A total of 1250 defective pictures are taken. At the same time, in order to quickly process defective pictures, this article fixed the size of the picture to 280×500 or 500×280. This paper

divides the surface defect pictures of mobile phones into four categories, screen scratches, edge defects, point defects, and stripe dents, as shown in Figure 4.



(a) Point defect (b) Edge defect

(c) Screen scratch (d) Stripe dent

Fig. 4. Samples of MPSD

B. Data Augmentation

Obviously limited by objective factors, the number of pictures taken is limited and the number of samples in each category is not balanced. In order to improve the detection and generalization ability of the model, the data of the mobile phone defect image is first augmented. Common methods of data augmentation include flipping, rotation, and color jitter. After the above data augmentation, 2495 samples are obtained. Then BEGAN is used to augment data, As shown in Table I, after expanding the data through BEGAN, the total sample size increased by 2258. Then, the obtained data is divided into training set, validation set, and test set. 70% of data are randomly divided into the training set, 15% are randomly divided into the verification set, and the remaining part is divided into the test set.

C. Evaluation Metrics

In order to effectively evaluate the performance of the algorithm, three evaluation indicators [24] are used, which are average precision (AP), mean average precision (mAP), and inference time. The above indicators are defined as follows:

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$AP = \int_0^1 P(R)dR \tag{8}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{9}$$

where the definition of TP is the number of actual defect samples predicted as defects; FP is the number of is the number of non-defective samples predicted as defects; FN is the number of non-defective samples predicted as non-defects; N is the number of detection categories.

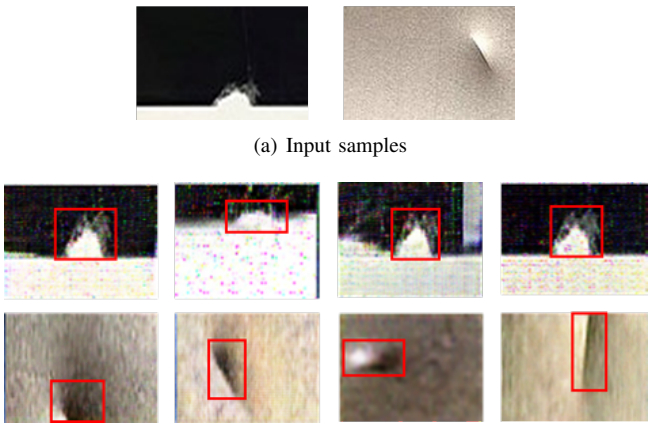
D. Experimental Environment and Training Detail

In order to speed up the training of the network, this paper uses a single GPU (GTX1080Ti) to improve the calculation efficiency. Python is used to implement the training and testing of improved Faster R-CNN. The whole experiment was implemented by using the open-source deep learning framework Tensorflow.

In order to obtain more diverse data, we need to train BEGAN. The size of the collected picture is 280×500, the first is to find the defect in the picture and crop the picture to 64×64. The kernel of each layer is the same size 3×3. Adam algorithm is chosen as the optimization algorithm, the learning rate of generator and discriminator are both set to 0.00008, and the beta1 is 0.5; the hyper-parameter γ is set to 0.5; The number of training steps is set to 10000.

Figure 5 shows some defect image samples generated by BEGAN. It can be seen from the figure that the generated defect image can well represent the real defect, and at the same time it has a certain diversity in geometric shape.

After obtained enough data, the proposed method is trained by using the Adaptive Moment Estimation (Adam) optimization algorithm [25] with momentum of 0.9, weight decay of 0.0001, beta1 of 0.9, and betal2 of 0.999. The number of iterations is set to 40,000. The learning rate is 0.001. The training batch size is set to 16 to avoid local minimum value.



(a) Input samples

(b) Output samples

Fig. 5. Some samples generated by BEGAN

E. Contrast Experiment and Ablation Experiment

Figure 6 shows some samples of MPSD by using the proposed method. Each predicted defect is marked through a rectangular box. The model we proposed successfully detected defects with extremely high accuracy. In order to evaluate the effectiveness of the improved algorithm proposed in this paper, a comparative experiment and an ablation experiment are designed. We list the AP and mAP in Table II to compare our method with other algorithms. From Table II, we can see that the proposed method obtains the highest mAP at

TABLE I
DATASET DETAIL

Dataset	Point defect	Edge defect	Screen scratch	Stripe dent	Total
Original number	560	560	670	705	2495
Augment number	523	536	583	616	2258
Total	1083	1096	1253	1321	4753

TABLE II
COMPARED WITH TRADITIONAL METHODS

Method	mAP	Point defect	Edge defect	Screen scratch	Stripe dent	Inference time
HOG+SVM	63.45%	73.88%	62.33 %	48.52%	69.06%	–
LBP+SVM	72.39%	77.52%	75.21%	61.80%	75.03%	–
Faster R-CNN(VGG16)	90.84%	90.81%	90.96%	91.44%	90.15%	0.111s
Faster R-CNN(ResNet101)	93.92%	95.75%	93.61%	93.97%	92.35%	0.197s
SSD-300	90.47%	90.82%	89.58%	90.17%	91.31%	0.016s
Yolov3(Darknet-53)	92.47%	94.36%	88.63%	92.23%	94.45%	0.029s
Our method	**99.43%**	**99.39%**	**99.99%**	**99.45%**	**98.89%**	0.208s

TABLE III
ABLATION EXPERIMENT

Method	FPN	RoI Align	Data Augmentation	mAP
Faster R-CNN(ResNet101)	×	×	×	93.92%
	✓	×	×	96.43%
	✓	✓	×	97.36%
	✓	✓	✓	99.43%

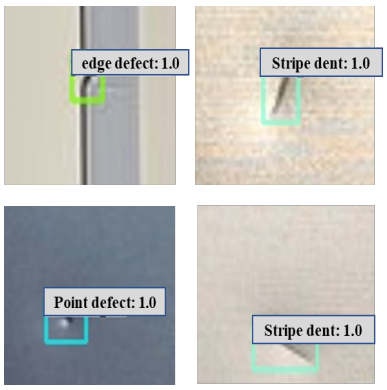


Fig. 6. Samples of detection results.

99.43%. Compared with the previous two traditional Faster R-CNN, the improved algorithm improves the detection accuracy by 8.59% and 5.51% respectively. Compared with one-stage methods(e.g. SSD, Yolov3), the mAP of our method can improve the mAP by 9.04% and 6.96% respectively. Compared with the traditional algorithm combining feature descriptors and SVM, the deep learning-based method has increased at least 18% on mAP. As can be seen from the table, the mAP based on the HOG method is only 63.45%, and the mAP based on the LBP method reaches 72.39%. The deep network in this article can reach 99.43%. The significant performance improvement indicates the greater performance of deep learning.

To verify the effectiveness of the improved module, the ablation experiment was designed with the following experimental schemes:

- Faster R-CNN based on ResNet101 was trained and tested on the dataset without augmentation.
- Faster R-CNN based on FPN+ResNet101 is trained and tested on the dataset without augmentation.
- Faster R-CNN based on FPN+ResNet101+RoI Align is trained and tested on the data set without augmentation.
- Faster R-CNN based on FPN+ResNet101+RoI Align is

trained and tested on the augmented dataset.

As we can see from Table III, the second scheme can increase the mAP of the original Faster R-CNN model by 2.51%, the third scheme can increase the mAP of the original Faster R-CNN model by 3.44%, and the last scheme can increase the mAP of the original Faster R-CNN model by 5.51%.

CONCLUSION

In this study, a mobile phone surface defect detection method based on improved Faster R-CNN is proposed. In order to solve the problem of small samples, we used the BEGAN method to expand the dataset. At the same time, the FPN structure is embedded in the traditional Faster R-CNN structure to achieve high-quality feature extraction for small size mobile phone defects, and the RoI Pooling layer is replaced by the RoI Align layer to prevent the regression of the bounding box of small size defects from being affected by the quantization operation. The contrast experiments and ablation experiments of improved strategies indicate that the BEGAN augmentation, FPN, and RoI Align can effectively improve the performance of the model. Finally, the proposed model can achieve 99.43% mAP on the mobile phone surface defect dataset, which outperforms the traditional Faster R-CNN and other traditional vision methods based on handcrafted feature extraction. The proposed method has two strengths: (i) Small samples can be easily augmented and the augmented data are more diverse. (ii) The method has a complex feature extraction network, so it is able to be used in other diverse small defect detection.

Although the performance of the proposed method is high enough, there are still some aspects that need to be improved: (i) The Inference speed of the proposed method is about 0.208s. Compared with the one-stage algorithm, its speed is not very ideal. Next, we will simplify the model to speed up the model while ensuring performance. (ii) The resolution of the dataset is low and only the local regions of mobile phones are used for training and detection. Maybe we will build a full view of high-resolution mobile phone surface defect dataset. (iii) At present, only the type and location of defects are detected. In production, some defects within a certain error range are acceptable. Therefore, in future work, we will focus on how to quantify and define the severity of defects.

ACKNOWLEDGMENT

We are very grateful that this work was supported by the National Key R&D Program of China (2018YFB1308600, 2018YFB1308602).

REFERENCES

[1] C. Jian, J. Gao, and Y. Ao, "Automatic surface defect detection for mobile phone screen glass based on machine vision," *Applied Soft Computing*, vol. 52, no. 52, pp. 348–358, 2017.

[2] J. Zhao, Q. Kong, X. Zhao, J. Liu, and Y. Liu, "A method for detection and classification of glass defects in low resolution images," pp. 642–647, 2011.

[3] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[4] H. Zhang, Z. Guo, Z. Qi, and J. Wang, "Research of glass defects detection based on dft and optimal threshold method," pp. 1044–1047, 2012.

[5] H. Huang, C. Hu, T. Wang, L. Zhang, F. Li, and P. Guo, "Surface defects detection for mobilephone panel workpieces based on machine vision and machine learning," pp. 370–375, 2017.

[6] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," *Microsoft Research Technical Report*, 1998.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[8] R. Girshick, "Fast R-CNN," pp. 1440–1448, 2015.

[9] B. Han, J. Sim, and H. Adam, "BranchOut: Regularization for online ensemble tracking with convolutional neural networks," pp. 521–530, 2017.

[10] C. Chen, A. Seff, A. L. Kornhauser, and J. Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," pp. 2722–2730, 2015.

[11] C. Zhang, W. Shi, X. Li, H. Zhang, and H. Liu, "Improved bare pcb defect detection approach based on deep feature learning," *The Journal of Engineering*, vol. 2018, no. 16, pp. 1415–1420, 2018.

[12] X. Xu, Y. Lei, and F. Yang, "Railway subgrade defect automatic recognition method based on improved faster r-cnn," *Scientific Programming*, vol. 2018, pp. 1–12, 2018.

[13] E. Titov, O. Limanovskaya, A. Lemekh, and D. Volkova, "The deep learning based power line defect detection system built on data collected by the cablewalker drone," in *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, 2019.

[14] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1493–1504, 2020.

[15] S. Ren, K. He, R. Girshick, and S. Jian, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2015.

[16] T. Y. Lin, P. Dollr, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," 2016.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770–778, 2016.

[18] K. He, G. Georgia, D. Piotr, and G. Ross, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, pp. 1–1, 2017.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," vol. 1, pp. 886–893, 2005.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[22] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating deep network training by reducing internal covariate shift," 2015.

[23] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: Boundary equilibrium generative adversarial networks," *arXiv: Learning*, 2017.

[24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[25] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014.