

# Defect Detection and Recognition of Mobile Phone Membrane Based on Convolutional Neural Network

Changmao Li<sup>1</sup>, Enbo Zhang<sup>2</sup>, Li Liu<sup>(✉)</sup>  
{[2538001101@qq.com](mailto:2538001101@qq.com)<sup>1</sup>, [248290248@qq.com](mailto:248290248@qq.com)<sup>2</sup>, [link\\_liuli@hotmail.com](mailto:link_liuli@hotmail.com) <sup>(✉)</sup>}

Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

**Abstract.** With the upgrading of mobile phone equipment, automatic detection of mobile phone film defects has been paid more and more attention in industrial production quality. Mobile phone film defect detection is a huge workload and challenging problem. Traditional methods can also detect some industrial identification defects, but these methods can only detect defects under specific conditions, such as obvious defect outline, strong contrast, low noise conditions. The defect detection method of mobile phone film proposed in this paper is to locate the target area with input images obtained from the industrial environment, remove the background, and then classify them into their designated classes through convolutional neural network. Experimental results show that this method can meet the robustness and accuracy of mobile phone film defect detection.

**Keywords:** Mobile phone film; Target location; Convolutional Neural Network; Defect detection.

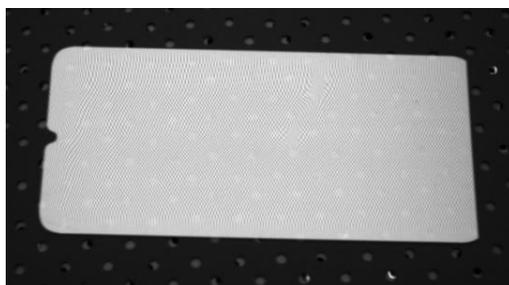
## 1 Introduction

Mobile phone film defect detection is an extremely important part of the operation for manufacturers, of course, manufacturers also make a lot of efforts on defect detection and quality control. In the photos of mobile phone film surface, problems such as strong reflection, background noise and difficult observation of defects are likely to occur, which increase the difficulty of detection. **Figure 1** shows an image of mobile phone film defect.

It can be seen from **Figure 1** that although there are friction trace defects, the background noise is complex and the reflection is strong, which poses a great challenge to the detection of mobile phone film defects. In recent years, machine vision-based methods can overcome many man-made shortcomings, such as low detection accuracy and poor performance, and have become a new research method for trend surface defect detection. Real time, high subjectivity and high work intensity. These inspection systems based on machine vision appear in many industrial applications, such as steel strip inspection [1,2].

---

✉ Corresponding author: Li Liu  
Fund Project: Danlian Science and Technology Innovation Fund (2020JJ26GX029)



**Fig. 1.** Mobile phone film defect image.

Liquid crystal display (LCD) [3], organization inspection [4,5], aluminum profile [6], railway track inspection [7], food inspection [8], optical component inspection [9]. The application of machine vision technology on surface defects has been extensively studied and tested. The traditional image processing method is based on the detection of local abnormal image pixels to solve the problem of prediction defects, which can be divided into structural methods, threshold methods, spectral methods and model-based methods [10]. Structural methods include model matching [11] and morphological processing [12]. Threshold methods include iterative optimization method threshold [13], OTSU method [14], contrast adjustment threshold [15], watershed method [16] and so on. Spectrum methods usually include Fourier transform [17] and wavelet transform [18]. Model-based methods include the Gaussian mixture entropy model [19]. A method based on machine learning usually includes two stages: feature extraction and model classification. Feature extraction includes local binary model (LBP) function [1] and gradient direction histogram (HOG) function [20]. **Figure 1** is the friction defect of the mobile phone film. As can be seen from the figure, the background is mostly black with white spots. These methods usually target specific conditions and lack adaptability and robustness to the aforementioned detection environment. Mobile phone film defect detection is a huge workload and challenging problem. Traditional methods can also detect some industrial identification defects. Although these detection algorithms have obtained good detection results in all aspects of surface defect detection, they cannot be directly applied to the above-mentioned mobile phone film surface detection. In the past decade, many researches have been devoted to the application of machine vision technology in surface defect detection, and The neural network algorithm is classifying the target, classification of natural scenes got good results [21]. A flexible multi-layer deep feature extraction framework based on CNN is proposed to detect anomalies in anomalous data sets [22]. Lin et al. Established a convolutional neural network (CNN) for fault checking of LED chips [23]. Defective areas are located using activation-like mapping technology without the need for annotation at the human area level. Liu et al. A detection system based on neural network detection level (DCNN) has been proposed.

## 2 Image preprocessing

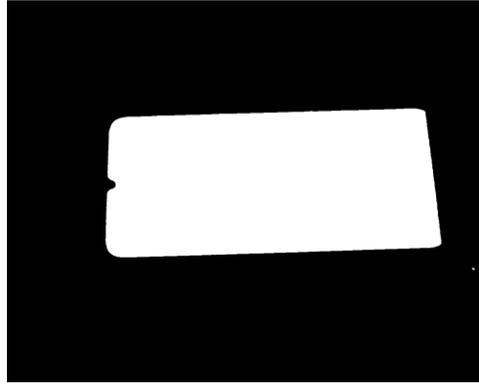
As shown in **Figure 1**, the defect image of mobile phone film was collected in the industrial environment by using an industrial camera. Gaussian smoothing is done to the input image, and the density function of one-dimensional Gaussian normal distribution is used :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (1)$$

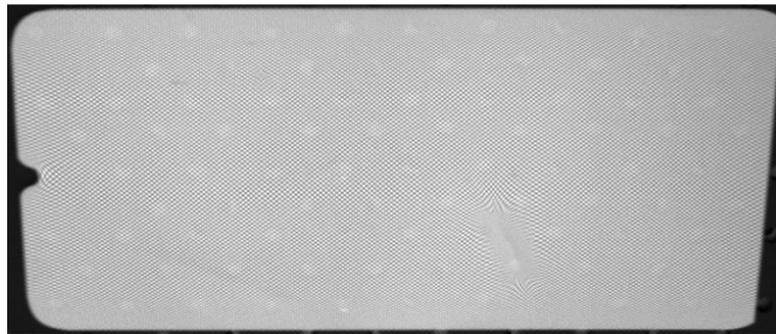
Where  $\mu$  is the mean of  $x$  and  $\sigma$  is the standard deviation of  $x$ . Since each calculation takes the current calculation point as the origin,  $\mu$  is equal to 0. Morphological processing is performed on the image, and interference factors in the background of the image are removed through corrosion and expansion operations. Corrosion is to replace the pixel value of the image under the anchor point overlap with the minimum value, and the expansion is to replace it with the maximum value on the contrary. Finally, through binarization processing, the gray histogram of the image is first calculated. Assuming that the mean value of the gray value is 130, the mean value is called  $M$ . Now arbitrarily select  $A$  gray value  $t$ , then the histogram can be divided into two parts, called  $A$  and  $B$ , respectively, and the average value of these two parts is  $MA$  and  $MB$ . The proportion of the number of pixels in part  $A$  to the total number of pixels is  $PA$ . Same thing with  $PB$ . The inter-class variance given by binarization is defined as:

$$ICA = PA * (MA - M)^2 + PB * (MB - M)^2 \quad (2)$$

Since the mobile phone film is easy to distinguish from the image background, through the above method, this article first uses Gaussian filtering to smooth the image, and then calculates the histogram of the picture to get the pixel threshold of the background and the mobile phone film. Use this critical value to apply binarization processing to the mobile phone. The membrane and the background can be separated, and then the mobile phone membrane can be cut out according to the image. The final picture of the mobile phone membrane pretreatment is as follows:



**Fig. 2.** Binarized mobile phone film image.



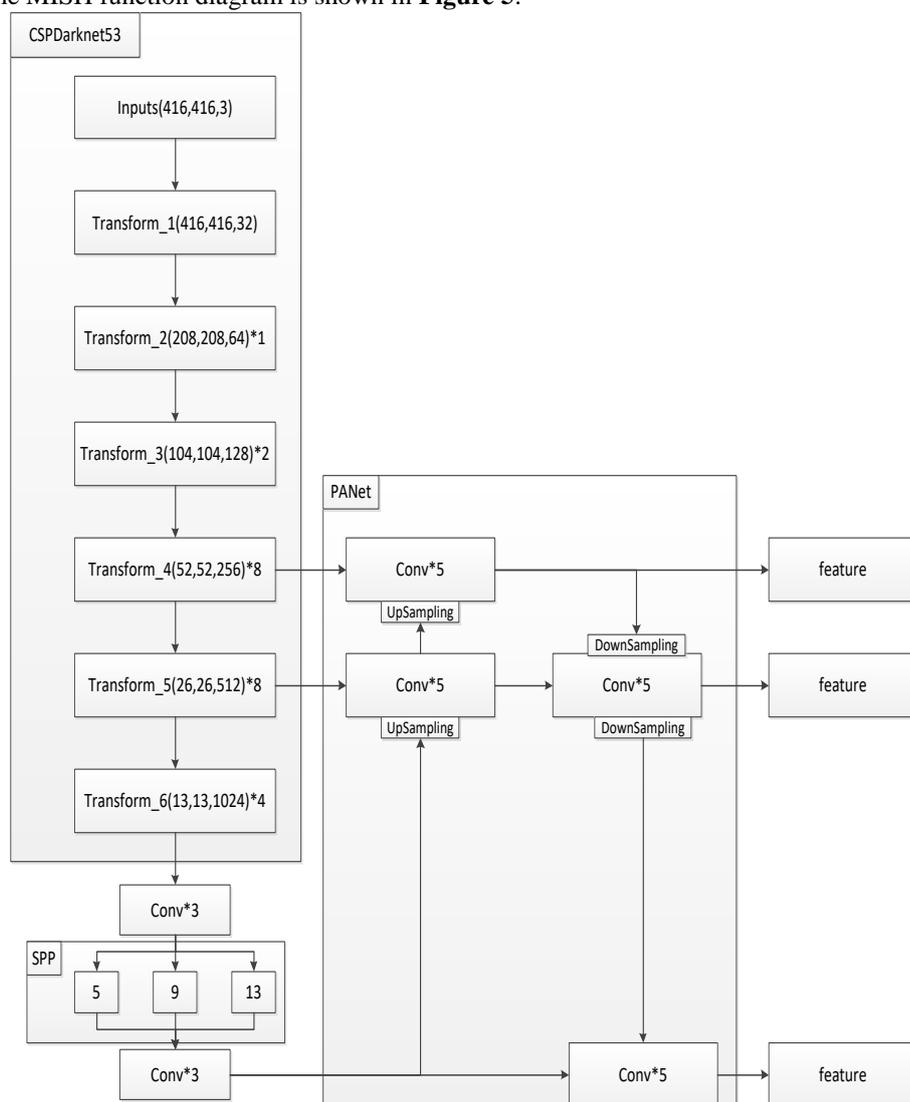
**Fig. 3.** Mobile phone membrane area.

### 3 Establish convolutional neural network

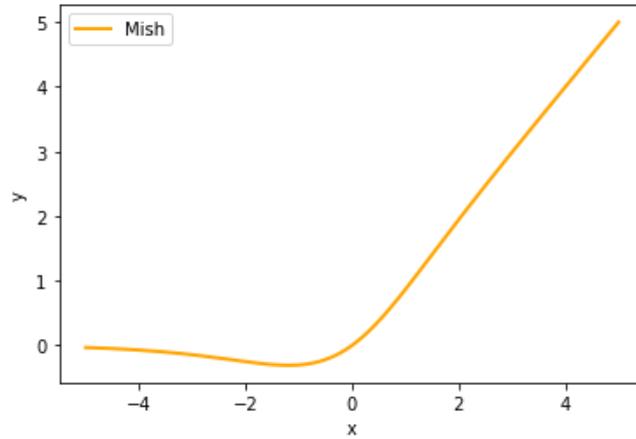
Feature pyramid is added to enhance image features while convolutional neural network is constructed. The overall neural network structure in this paper is divided into CSPDarknet53, SPP and PANet structures. These three parts are combined to finally get the defect characteristics of the mobile phone film image. The flow chart of the convolutional neural network is shown in **Figure 4**. The activation function used in this paper is MISH function, and the formula of MISH function is as follows :

$$Mish = x * \tanh(\ln(1 + e^x)) \quad (3)$$

The MISH function diagram is shown in **Figure 5**:

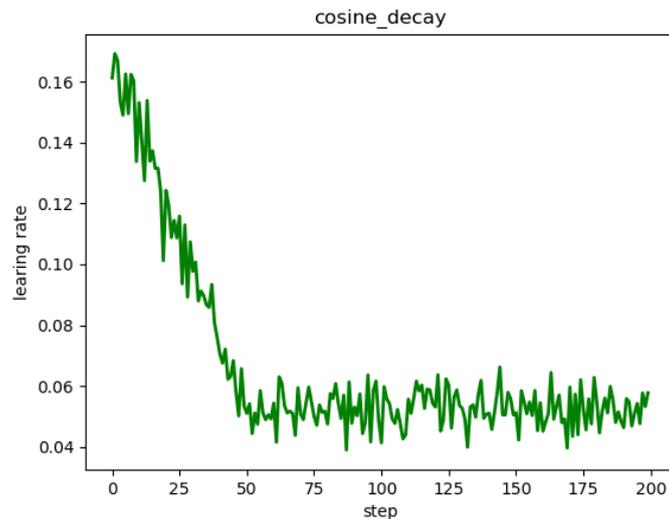


**Fig. 4.** Convolutional Neural Network flow chart.



**Fig. 5.** Coordinate diagram of the Mish function.

In the CSPDarknet53 structure, the convolution kernel is 3\*3 with a step size of 1, and the convolution is carried out successively to obtain the transformation diagram. In the SPP structure, the pooling kernel of 5\*5,9\*9,13\*13 is used to carry out the maximum pooling operation, which can greatly increase the receiver field and separate the most significant context features. PANet structure has the meaning of repeatedly improving features. Upsampling and Downsampling are respectively adopted to realize repeated feature extraction to reduce semantic loss. Finally, the acquired feature is detected and recognized. In terms of data enhancement, this article uses four images to be spliced in a certain direction, through the image flip, color gamut change, and zoom. The great advantage of this is to enrich the background of the detection target and also speed up the calculation. And this article also uses the learning rate cosine annealing decay algorithm. The learning rate cosine annealing decay is shown as follows:



**Fig. 6.** Learn the cosine function of rate decay diagram.

**Table 1.** Convolutional Neural Network Structure Configuration Table.

Network structure	kernel size	Stride
CSPDarkNet53	3*3	1
PANet	1*1	1
SPP	5*5or9*9or13*13	1

The learning rate will first rise and then fall. When it rises, it uses a linear rise. When it falls, the simulated cosine function decreases. Repeatedly execute it many times until the desired effect is achieved.

**Table 1** can show the parameters of the network in detail.

Composition of loss Calculate parameters required by loss When calculating loss, it is actually a comparison between  $Y_{pre}$  and  $Y_{true}$ :  $Y_{pre}$  is the output of an image after passing through the network, which contains contents of three feature layers inside;  $Y_{true}$  is the offset position, length, width, and type of the grid (19, 19), (38, 38), and (76, 76) corresponding to each of its real boxes in a real image. The final output content of the network is the prediction box and its type corresponding to each grid point of the three feature layers.

$Y_{pre}$  is the three feature layers respectively correspond to the position, confidence degree and type corresponding to the three prior boxes on each grid point after the picture is divided into grids of different sizes. For  $y_1$ ,  $y_2$ , and  $y_3$  of the output, [..., : 2] refers to the offset relative to each grid point, [..., 2: 4] refers to the width and height, [..., 4: 5] refers to the confidence of the box, and [..., 5:] refers to the prediction probability of each category. For now,  $Y_{pre}$  is still undecoded. After decoding, it will only look like the real image.  $Y_{true}$  is the offset position, length, width, and type of the grid (19, 19), (38, 38), and (76, 76) corresponding to each real box in a real image. It still needs to be coded to match the structure of  $Y_{pre}$  Loss values need to be processed on three feature layers, taking the smallest feature layer as an example.

The calculation steps of loss are as follows:

- (1). Use  $Y_{true}$  to extract the position of the point (m, 19, 19, 3, 1) and its corresponding category (m, 19, 19, 3, 80) of the real target in the feature layer.
- (2). After the prediction value of prediction is output and processed, the predicted value of reshape is  $Y_{pre}$  and shape is (m, 19, 19, 3, 85). And x y, w h decoded.
- (3). For each picture, calculate the IOU of all the real boxes and prediction boxes. If the coincidence degree of some prediction boxes and real boxes is greater than 0.5, it will be ignored.
- (4). Calculate CIOU as regression loss. Here, only regression loss of positive samples is calculated.
- (5). The loss for calculating confidence is composed of two parts. The first part is that there is actually a target, and the confidence value in the prediction result is compared with 1; In the second part, there is actually no target. In the fourth step, the value of its maximum IOU is compared with 0.
- (6). Calculate the loss of the predicted category, which calculates the gap between the

predicted class and the real class where there is a target in fact.

The actual existing box, the value of the confidence in the prediction result is compared with 1; For boxes that do not actually exist, the confidence value in the prediction result is compared with 0. In this text, the boxes that do not contain the target that are ignored are removed. The actual existing boxes, and the comparison between the predicted results and the actual results.

IOU is the concept of ratio, which is insensitive to the scale of the target object. However, the regression loss optimization of commonly used BBOX and IOU optimization are not completely equivalent, and ordinary IOU cannot directly optimize the non-overlapping part. The CIOU will get the loss value.

Therefore, some people propose to directly use CIOU as regression optimization loss, and CIOU is a very excellent idea. CIOU takes into account the distance between the target and anchor, overlap rate, scale and penalty items, which makes the target box regression more stable and avoids problems such as divergence in the training process like IOU and CIOU. The penalty factor takes into account the comparison between the length and width of the predicted box and the aspect ratio of the target box. Formula is as follows:

$$CIOU = IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (4)$$

$\rho^2(b, b^{gt})$  respectively represent the Euclidean distance of the center point of the prediction box and the real box.  $c$  represents the diagonal distance of the smallest closure region that can contain both the prediction box and the real box.

## 4 Experiment

In actual industrial production line, the images are fewer defects, and the image acquisition and tags require manual operation, so we choose 400 images as the data set, through the image preprocessing operations intercept membrane partial area to keep the mobile phone, then the image according to the ratio of 9:1, 360 images as the training set, the rest for the prediction set. This article uses labeling image annotation software to mark the training set of mobile phone film separately, mark the friction marks in the mobile phone film defect as 01, frame it with a box, and save it in xml format. These images were then input into the established convolutional neural network for training. In order to evaluate the detection results, we used IOU and precision to quantitatively evaluate the performance of the two subtasks.

For the split task, IOU is defined as:

$$IOU(FM, SM) = \frac{Area((FM \cap SM))}{Area(FM \cup SM)} \quad (5)$$

Where FM represents Friction Marks, and SM represents Scratch Marks. The accuracy is used to quantitatively evaluate the performance of classification tasks, and the calculation formula is as follows:

$$Accuracy = \frac{Tp}{Tp + Fp} \quad (6)$$

From the second step, we can obtain the prediction results of the three feature layers, whose shapes are (N, 19, 19, 255), (N, 38, 38, 255), and (N, 76, 76, 255) respectively, corresponding to the positions of three prediction boxes on the grid of 19x19, 38x38, and 76x76 for each graph. However, this prediction result does not correspond to the position of the final prediction box on the picture, which can only be completed by decoding. The three feature layers divide the whole map into grids of 19x19, 38x38 and 76x76 respectively, and

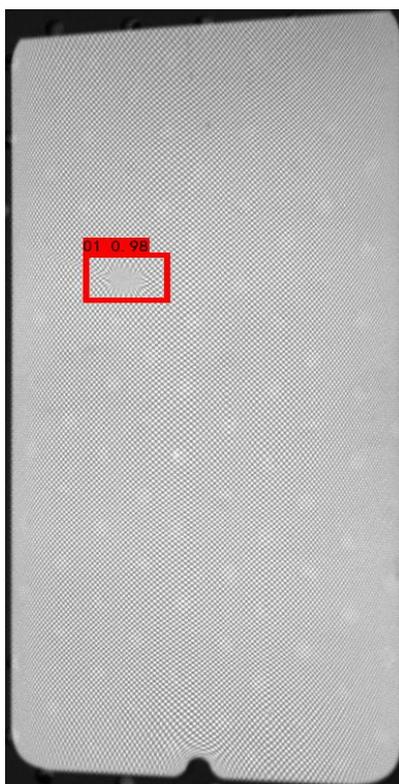
each network point is responsible for the detection of a region. We know that the prediction results of feature layers correspond to the positions of three prediction boxes. We first reshape them, and the results are (N, 19, 19, 3, 85), (N, 38, 38, 3, 85), (N, 76, 76, 3, 85).The 85 in the last dimension contains 4+1+80, representing X\_offset, Y\_offset, H and W, confidence, and classification results, respectively.

The decoding process is to add each grid point with its corresponding X\_offset and Y\_offset, and the result is the center of the prediction box. Then, the length and width of the prediction box are calculated by combining the prior box with H and W. So that gives you the position of the entire prediction box.

Where TP and FP represent the number of defect areas correctly and incorrectly classified into their own categories. Of course, score sorting and non-maximum inhibition screening are also needed to pick out the score of each category that is greater than the set score. Then, non-maximum inhibition is carried out by using the position and score of the box, and the position of the defect can be framed finally.

Both of which are accurately framed, and 01 represents the friction mark defect. Through the overall measurement of the accuracy of defect location and defect recognition, it is found that the accuracy of defect prediction is 98%. And there is only one defect of friction mark in the picture, and there are many interference factors, which have a greater impact, but In general, this method has a good effect in predicting defects.

The result of the test is shown in **Figure 7**:



**Fig. 7.** Detection effect diagram.

## 5 Summary

In this paper, a new convolutional neural network (convolutional neural network) structure, combined with image preprocessing operation, is proposed for defect detection of mobile phone film in industry. By improving the convolutional neural network structure can exactly detect the defects on the membranes of the mobile phone and recognition, using industrial acquisition data sets, forecast data, the test results show that this method can effectively detect the defects, visualization and quantitative experiments have shown that the method can satisfy the requirement of complicated industrial environment. However, one limitation of this method is that it requires a large number of people to manually mark, which will take a lot of time and expense. In the future, we need to improve industrial equipment to automatically mark the defects of mobile phone film images and reduce a lot of unnecessary costs.

### Acknowledgment

The authors acknowledge financial support from the Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks to the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

### References

- [1] Song, K.: noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* pp. 858–864 (2013)
- [2] Wu, Y.: A UAV-based visual inspection method for rail surface defects. *Appl. Sci.* pp. 245–248 (2018)
- [3] Cen, Y. G.: Defect inspection for TFT-LCD images based on the low-rank matrix reconstruction. *Neurocomputing.* pp. 1206–1215 (2015)
- [4] Lei, J.: Scale insensitive and focus driven mobile screen defect detection in industry. *Neurocomputing.* pp. 72–81 (2018)
- [5] Li, Y.: Deformable patterned fabric defect detection with Fisher criterion-based deep learning. *IEEE Trans. Autom. Sci. Eng.* pp. 1256–1264 (2017)
- [6] Chondronasios, A.: Feature selection for surface defect classification of extruded aluminum profiles. *Int. J. Adv. Manuf. Technol.* pp. 33–41 (2016)
- [7] Gibert, X.: Deep multitask learning for railway track inspection. *IEEE Trans. Intell. Transp. Syst.* pp. 153–164 (2017)
- [8] De Araújo, S.A.: Beans quality inspection using correlation-based granulometry. *Eng. Appl. Artif. Intell.* pp. 84–94 (2015)
- [9] Tao, X.: Weak scratch detection and defect classification methods for a large-aperture optical element. *Opt. Commun.* pp. 390–400 (2017)
- [10] Ren, R.: A generic deep-learning-based approach for automated surface inspection. *IEEE Trans. Cybern.* pp. 929–940 (2018)
- [11] Jian, C.: Automatic surface defect detection for mobile phone screen glass based on machine vision. *Appl. Soft Comput.* pp. 348–358 (2017)
- [12] Mak, K. L.: Fabric defect detection using morphological filters. *Image Vis. Comput.* pp. 1585–1592 (2009)
- [13] Li, X.: Quantitative surface crack evaluation based on eddy current pulsed thermography. *IEEE Sens. J.* pp. 412–421 (2017)

- [14] Yuan, X.: An improved Otsu method using the weighted object variance for defect detection. *Appl. Surf. Sci.* pp. 472–484 (2015)
- [15] Win, M.: A contrast adjustment thresholding method for surface defect detection based on mesoscopy. *IEEE Trans. Ind. Inform.* pp. 642–649 (2015)
- [16] Wang, L.: Calculation of flexible printed circuit boards (FPC) global and local defect detection based on computer vision. *Circ.* pp. 49–54 (2016)
- [17] Bai, X.: Saliency-based defect detection in industrial images by using phase spectrum. *IEEE Trans. Ind. Inform.* pp. 2135–2145 (2014)
- [18] Borwankar, R.: An Optical Surface Inspection and Automatic Classification Technique Using the Rotated Wavelet Transform. *IEEE Trans. Instrum. Meas.* pp. 690–697 (2018)
- [19] Susan, S.: Automatic texture defect detection using Gaussian mixture entropy modeling. *Neurocomputing.* pp. 232–237 (2017)
- [20] Shumin, D.: Adaboost learning for fabric defect detection based on hog and SVM. In *Proceedings of the International Conference on Multimedia Technology, Hangzhou, China*, pp. 26–28 (2011)
- [21] Ganovska, B.: Design of the model for the on-line control of the AWJ technology based on neural networks. *Indian J. Eng. Mater. Sci.* pp. 279–287 (2016)
- [22] Natarajan, V.: Convolutional networks for voting-based anomaly classification in metal surface inspection. In *Proceedings of the IEEE International Conference on Industrial Technology, Toronto, ON, Canada*. Pp. 22–25 (2017)
- [23] Lin, H.: Automated defect inspection of LED chip using deep convolutional neural network. *J. Intell. Manuf.* pp. 1–10. (2017)