

# Surface Defect Detection using Hierarchical Features

Ling Xiao, Tao Huang, Bo Wu, Youmin Hu, Jiehan Zhou,

**Abstract**—In this paper, we propose an instance level hierarchical features based convolution neural network model(H-CNN) for detecting surface defects. The H-CNN uses different convolutional layers' extracted features to generate defect masks. The H-CNN first generates proposal regions. Then, it proposes a fully convolutional neural network to extract different level's convolutional features and detect instance level defects. We applied the H-CNN model in freight train detection system for detecting oil-leaks, and the results demonstrate that the H-CNN can effectively identify and generate defect masks. It achieves 92% accuracy on the large reflective oil-leak stain, 86% on the large non-reflective oil-leak stain, 89% on the small reflective oil-leak stain and 74% on the small non-reflective oil-leak stain. Its image process speed is 0.467s per frame.

## I. INTRODUCTION

Surface defect detection is an important task [1]–[4]. It has two main challenges. First, the defect image dataset is small, due to low occurrence. Second, defect detection requires precisely detecting defect mask.

Most surface defect detection algorithms use the low-level image hues [5]–[10]. These methods use local and global color, gradient, and texture properties to detect surface anomalies. Wu et al. [8] presented a technique that uses multi-feature fusion to detect surface defects. Shi et al. [9] improved the Sobel algorithm for rail surface defect detection. Li et al. [10] proposed a local neighborhood-based surface defect detection algorithm for improving the surface defect detection accuracy in automation industry. Truong et al. [11] suggested an automatic thresholding technique, which uses an entropy weighting scheme to detect defect regions. These algorithms are useful in some cases but have low robustness. They will be easily affected by hand-crafted parameters. Also, most of them aim at the specific defect due to extracted features' low representative ability, which has low generalization.

\*The research work was funded by the National Key Research and Development Program of China (No. 2017YFD0400400), the scholarship from China Scholarship Council (No. 201806160119).

<sup>1</sup>Ling Xiao is with School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, PR China [lingxiao@hust.edu.cn](mailto:lingxiao@hust.edu.cn)

<sup>1</sup>Tao Huang is with School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, PR China [huangtao@hust.edu.cn](mailto:huangtao@hust.edu.cn)

<sup>1</sup>Bo Wu is with School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, PR China [bowu@hust.edu.cn](mailto:bowu@hust.edu.cn)

<sup>1</sup>Youmin Hu is with School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, PR China [youthwh@hust.edu.cn](mailto:youthwh@hust.edu.cn)

<sup>2</sup>Jiehan Zhou is with ITEE -Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland [jiehan.zhou@gmail.com](mailto:jiehan.zhou@gmail.com)

In recent years, deep learning has been applied in detecting objects [12], including instance level object segmentation [13] and surface defect detection [14]–[16]. Li et al. [17] proposed an improved YOLO(you only look once) detection network for real-time steel strip surface defect detection. Zhu et al. [18] took advantage of the tire feature similarity to capture the anomalies. Tao et al. [19] designed a novel cascaded auto-encoder architecture for segmenting and localizing metallic surface defect. Kai et al. [20] proposed a general framework for object instance segmentation. This framework can efficiently detect objects while simultaneously generating a high-quality segmentation mask for each instance. Jia et al. [21] presented a multiple-kernel learning feature fusion method for surface defect detection. Ren et al. [22] proposed a generic approach that requires small training data for automated surface inspection. Instance segmentation technique shows its power in detection and segmentation. It extracts discriminative features, and separates the defects and backgrounds.

This paper proposes a hierarchical features based convolution neural network (H-CNN) model to detect surface defect. The H-CNN adopts the instance segmentation technique. It adds a CNN module to generate more precise defect mask. It first uses the region proposal network(RPN) [23] to generate proposal regions. Then, it suggests a CNN module to learn more defect information, and create more accurate defect masks.

The remainder of the paper is organized as follows. Section 2 presents the work related to the H-CNN, including the region-based convolution neural network approach (R-CNN) and the instance segmentation technique. Section 3 illustrates the H-CNN framework. Section 4 applies the H-CNN to detect oil-leaks for freight trains, and demonstrates that the H-CNN outperforms other methods. Section 5 draws a conclusion.

## II. RELATED WORK

### A. R-CNN

Girshick et al. [24] developed the R-CNN, which applies high-capacity CNNs(convolutional neural networks) to generate bottom-up region proposals for localizing and segmenting objects. The R-CNN has been improved for higher efficiency and accuracy [23], [25]–[27]. The faster R-CNN [23] improved R-CNN by learning the attention mechanism with a RPN. It becomes the leading framework in several benchmarks [28]–[30].

### B. Instance segmentation

There are many instance segmentation methods proposed for the effectiveness of R-CNN. Dai et al. [31] suggested a complex multiple-stage cascade that predicts segment proposals from bounding-box proposals, followed by classification. Li et al. [32] combined the segment proposal system and the object detection system for fully convolutional instance segmentation (FCIS). The FCIS simultaneously addresses object classes, boxes, and masks, making the system operate fast [33], [34]. However, it exhibits systematic errors on overlapping instances and creates spurious edges, which makes it difficult to segment instances.

### III. H-CNN ARCHITECTURE

The H-CNN consists of two phases. In the phase 1, the H-CNN adopts faster R-CNN model to generate regions of interest (ROIs). Then the H-CNN produces a fully convolutional neural network (F-CNN) to extract different hierarchical image features to predict the defect mask. Figure. 1 illustrates the H-CNN architecture.

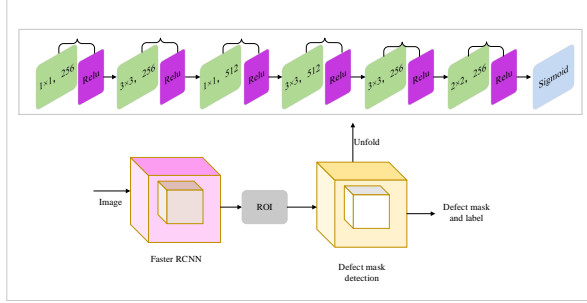


Fig. 1. The H-CNN architecture

#### A. Generate ROIs

The Faster R-CNN detector [23] consists of two stages (Fig. 2). The first stage suggests candidate object bounding boxes. The second stage extracts features using RoIPool from each candidate box and executes classification and bounding-box regression.

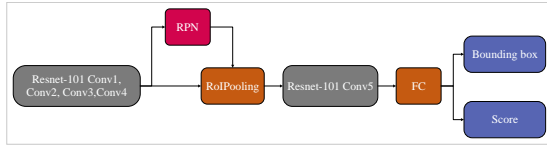


Fig. 2. The faster R-CNN architecture

#### B. Generate defect mask

We propose a CNN model to obtain the hierarchical image features, including both specific and general characteristics. Therefore, it could extract more information from a small dataset.

We use different kernel size on the image with different dimensionality. Suppose the kernel size is  $3 \times 3$ , the padding size is  $1 \times 1$ , the stride is  $2 \times 2$ . When the input image size

is  $5 \times 5$ , the output feature map is  $3 \times 3$ . So, different kernel size on each dimension can output different size feature map. The loss function is defined as (Equation 1),

$$L_{mask}(n_i^l, n_i^{l*}) = \frac{1}{m} \sum_{l=1}^m (-[n_i^{l*} \log(n_i^l) + (1-n_i^{l*}) \log(1-n_i^l)]) \quad (1)$$

where  $n_i^l$  denotes the probability of true detection of  $l_{th}$  pixel of class  $i$ ,  $n_i^{l*}$  equals 1 when predicted label is a real label, and  $m$  denotes pixel numbers of class  $i$ .

### IV. EXPERIMENTS

#### A. Experimental settings

In recent years, freight train detection system (FTDS) is installed on the railway to monitor train safety. FTDS uses high-speed camera to catch all parts of the running train; then, it compresses the captured image and saves them in the computer; subsequently, it selects images of key parts of the train that affects the safety of the train, including gear, bearing, and engine, etc.; The selected images are then transmitted to the data server in monitor room for anomaly detection by indoor inspectors. Once the defect is observed, the indoor inspector will inform outdoor inspectors to confirm and handle the problem [35]. Although effective, the FTDS can only capture, transmit, and store images, leaving anomaly inspection for indoor inspectors. Meanwhile, erroneous detection is prone to occur during human detection due to vision fatigue.

This study focuses on oil-leak detection in FTDS, which has never been studied. Oil-leak detection in FTDS has two main difficulties, namely the similarity to dark backgrounds and small datasets.

Oil-leak has four major types, namely the large reflective, large non-reflective, small reflective and small non-reflective oil-leak stains (Fig. 3). When the oil-leak stain area is larger than  $250 \times 250$  pixels, we define it as large oil-leak stain. The parameter *ratio* is used to differentiate reflective and non-reflective oil-leak stain (Equation 2), if *ratio* is less than 1%, we define it non-reflective oil-leak stain.

$$\begin{cases} ratio = \frac{n}{N} \\ n : \text{number of points that } I > 250 \text{ and } S < 5 \\ N : \text{number of all detected points} \\ I = \frac{R + G + B}{3} \\ S = abs(\frac{\min(R, G, B) * 765}{R + G + B} - 255) \end{cases} \quad (2)$$

We obtain 28 images with oil-leaks in the experiment. We make data augmentations for them. We use 18 raw images to create 500 training samples through flipping, rotation and scaling transformations. And we use another 10 images to generate 150 validation samples and 150 test samples using the same transformations, respectively.

We also introduce an evaluation rule to calculate detection accuracy. Suppose the area of a detected region is  $A$ , the area

of a real oil-leak region is  $B$ . Then, if  $A \cap B \leq 0.8 * B$ , it will be classified as a wrong detection. Then we use  $F_1$ -measure as Formula (3) to evaluate detection accuracy,

$$\begin{aligned} F_1 - measure &= \frac{2 * P * R}{P + R} \\ P &= \frac{TP}{TP + FP} \\ R &= \frac{TP}{TP + FN} \end{aligned} \quad (3)$$

Where  $P$  denotes the precision,  $R$  denotes the recall.  $TP$  means true positive,  $FP$  means false positive,  $FN$  means negative positive.

### B. Implementation

The H-CNN is implemented on a laptop with GPU 1060 MAX-Q, which has four cores with 6GB memory.

The hyper-parameters of the faster R-CNN module are the same as the one in [23]. We resize images as  $512 \times 448$  pixels. We use three scales with box areas of  $64 \times 64$ ,  $128 \times 128$  and  $256 \times 256$  pixels, and three aspect ratios of 1:1, 1:2, and 2:1. Some RPN proposals highly overlap each other. To reduce redundancy, we adopt non-maximum suppression (NMS) on the proposal regions based on their class scores. We fix the IoU (intersection over union) threshold for NMS at 0.7. After NMS, we use the top- $N$  ranked proposal regions as detection results. The learning rate is 0.01, the weight decay is 0.0001, the momentum is 0.9 in the defect mask detection module. We run the mask generation only on proposal regions. Each mini-batch has two images per GPU, and each image has  $N$  sampled RoIs, with a ratio of 1:2 of defects to non-defects. We can predict  $K$  masks per RoI, but we only use the  $k$ -th mask, where  $k$  is the predicted class by the classification branch. Finally, the mask is binarized at a threshold of 0.5.

### C. Results and discussions

We use the obtained 150 images to test the H-CNN detection accuracy on the four type oil-leak stains. To calculate the accuracy, we separate the 150 test images into 10 test sets, and 10 mean accuracy values are obtained from these test sets. Experimental results are shown in Fig. 4, Fig. 5 and Tab. I. It achieves 92% accuracy on the large reflective oil-leak stain, 86% on the large non-reflective oil-leak stain, 89% on the small reflective oil-leak stain and 74% on the small non-reflective oil-leak stain.

TABLE I

$F_1 - measure$  OF THE H-CNN DETECTION RESULTS ON DIFFERENT OIL-LEAK STAINS.

| Rule                | The large reflective | The large non-reflective |
|---------------------|----------------------|--------------------------|
| $F_1 - measure(\%)$ | 92                   | 86                       |
| Rule                | The small reflective | The small non-reflective |
| $F_1 - measure(\%)$ | 89                   | 74                       |

To further evaluate the H-CNN performance, we compare it with the LBP(local binary pattern) [5], the Sobel [9] and

the DNN(deep neural networks) [16]. In the experiment, we make modifications for the three methods shown in Fig. 6.

We apply the four models to detect the four oil-leaks. Results are shown in Fig. 7, Tab. II and Tab. III. The H-CNN achieves higher accuracy than others. Moreover, it consumes 0.467s per frame, which is faster than other methods.

TABLE II

DETECTION RESULT WITH THE FOUR MODELS(%)

| Method    | The large reflective | The large non-reflective |
|-----------|----------------------|--------------------------|
| H-CNN     | 92                   | 86                       |
| The LBP   | 64                   | 56                       |
| The sobel | 58                   | 51                       |
| DNN       | 73                   | 72                       |
| Method    | The small reflective | The small non-reflective |
| H-CNN     | 89                   | 74                       |
| The LBP   | 57                   | 56                       |
| The sobel | 50                   | 51                       |
| DNN       | 72                   | 68                       |

TABLE III

IMAGE PROCESS SPEED OF THE FOUR MODELS(%)

| Method  | H-CNN | The LBP | The sobel | DNN   |
|---------|-------|---------|-----------|-------|
| Time(s) | 0.467 | 0.852   | 1.074     | 0.879 |

## V. CONCLUSIONS

This paper introduces a novel hierarchical features based convolution neural network for surface defect detection. The model has been applied to detect four types of oil-leaks in FTDS. And the experiments show that the H-CNN can detect different oil-leak stains accurately and efficiently. H-CNN shows better results compared with the state-of-the-art methods. The H-CNN also has high robustness to the scaling, the flipping and the rotation transformations.

In the future, the H-CNN can be applied to detect cracks, scratches, foreign matters et al. We will further improve the H-CNN accuracy and efficiency.

## ACKNOWLEDGMENT

The research work was funded by the National Key Research and Development Program of China (No. 2017YFD0400400), the scholarship from China Scholarship Council(No. 201806160119).

## REFERENCES

- [1] Santos, T. G., Miranda, R. M., de Carvalho, C. C., A new NDT technique based on bacterial cells to detect micro surface defects. NDT & E International, vol. 63, 2014, pp. 43-49.
- [2] Huang, DP, Liao, SP, Sunny AL, Yu SD., A novel automatic surface scratch defect detection for fluid-conveying tube of Coriolis mass flow-meter based on 2D-direction filter. MEASUREMENT, vol. 126, 2018, pp. 332-341.
- [3] Yao, MH., Gu, QL., A sparse representation method for image-based surface defect detection. OPTOELECTRONICS LETTERS, vol. 14, No. 6, 2018, pp. 476-480.
- [4] Li, Y., Huang, H., Xie, Q., Yao, L., Chen, Q., Research on a Surface Defect Detection Algorithm Based on MobileNet-SSD. Applied Sciences, vol. 8, No. 9, 2018, pp. 1678.

- [5] Liu, XJ., Xue, F., Teng, L., Surface Defect Detection Based on Gradient LBP. In 2018 IEEE 3RD INTERNATIONAL CONFERENCE ON IMAGE, VISION AND COMPUTING (ICIVC), Chongqing, PEOPLES R CHINA, 2018, pp. 133-137.
- [6] Rong, D., Rao, XQ., Ying, YB., Computer vision detection of surface defect on oranges by means of a sliding comparison window local segmentation algorithm. COMPUTERS AND ELECTRONICS IN AGRICULTURE, vol. 137, 2017, pp.59-68.
- [7] Jian, C.X., Gao, J. Ao., Y.H., Automatic surface defect detection for mobile phone screen glass based on machine vision. Applied Soft Computing, vol. 52, 2017, pp. 348-358.
- [8] Wu, XJ., Xiong, HJ., Yu, ZY., Wen, PZ., A Surface Defect Detection Method Based on Multi-Feature Fusion. In 9th International Conference on Digital Image Processing (ICDIP), Hong Kong, PEOPLES R CHINA, 2017.
- [9] Shi, T., Kong, JY., Wang, XD., Liu, Z., Zheng, G., Improved Sobel algorithm for defect detection of rail surfaces with enhanced efficiency and accuracy. JOURNAL OF CENTRAL SOUTH UNIVERSITY, vol. 23, No. 11, 2015, pp. 2867-2875.
- [10] Li, CF., Chen, XH., Surface Defect Detection Algorithm Based on Local Neighborhood Analysis. In International Conference on Information Technology and Intelligent Transportation Systems (ITITS), Xian, PEOPLES R CHINA, vol. 454, pp. 307-315, Nov. 2016.
- [11] Truong, M. T. N., Kim, S., Automatic image thresholding using Otsu's method and entropy weighting scheme for surface defect detection. Soft Computing, vol. 22, No. 13, pp. 4197-4203, Nov. 2018.
- [12] Ghadiri, F., Bergevin, R., Bilodeau, G. A., From superpixel to human shape modelling for carried object detection. Pattern Recognition, vol. 89, 2019, pp. 134-150.
- [13] Dakhia, A., Wang, T., Lu, H., Multi-scale Pyramid Pooling Network for salient object detection. Neurocomputing, vol. 333, 2019, pp. 211-220.
- [14] Shi, D., Zhou, J., Xu, J., Yang, J., Li, X., Zhao, Z., et al, Machine Vision-Based Segmentation and Classification Method for Intelligent Roller Surface Monitoring. In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, pp. 1811-1817, Oct. 2018.
- [15] Jia, H., Lian, XF., Product Surface Defects Detection Based on Multiple-Kernel Learning Feature Fusion Method. In 2018 IEEE INTERNATIONAL CONFERENCE OF INTELLIGENT ROBOTICS AND CONTROL ENGINEERING, Lanzhou Jiaotong Univ, Lanzhou, PEOPLES R CHINA, pp. 247-251, Aug. 2018.
- [16] Jung, SY., Tsai, YH., Chiu, WY., Hu, JS., Sun, CT., Defect Detection on Randomly Textured Surfaces by Convolutional Neural Networks. In 2018 IEEE/ASME INTERNATIONAL CONFERENCE ON ADVANCED INTELLIGENT MECHATRONICS (AIM), Auckland, NEW ZEALAND, pp. 1456-1461. Jul. 2018
- [17] Li, J.Y., Su, Z.F., Geng, J.H., Yin, Y.X., Real-time Detection of Steel Strip Surface Defects Based on Improved YOLO Detection Network. IFAC-PapersOnLine, vol. 51, No. 21, 2018, pp. 76-81.
- [18] Zhu, QD., Ai, XT., The Defect Detection Algorithm for Tire X-ray Images Based on Deep Learning. In 3rd IEEE International Conference on Image, Vision and Computing (ICIVC), Chongqing, PEOPLES R CHINA, pp. 138-142, Jun. 2018.
- [19] Tao, X., Zhang, DP., Ma, WZ., Liu, XL., Xu, D., Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks. APPLIED SCIENCES-BASEL, vol. 8, 2018, No. 9.
- [20] Kaiming H., Georgia Gkioxari, Piotr Dollr, Ross Girshick, Mask R-CNN. Computer Vision and Pattern Recognition, 2018.
- [21] Jia, H., Lian, X., Product Surface Defects Detection Based on Multiple-Kernel Learning Feature Fusion Method. In 2018 IEEE International Conference of Intelligent Robotic and Control Engineering, Lanzhou, China, pp. 247-251. Aug. 2018.
- [22] Ren, R., Hung, T., Tan, K. C., A generic deep-learning-based approach for automated surface inspection. IEEE transactions on cybernetics, vol. 48, No. 3, 2018, pp. 929-940.
- [23] Ren, S., He, K., Girshick, R., Sun, J., Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015, pp. 91-99.
- [24] Girshick, R., Donahue, J., Darrell, T., Malik, J., 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.
- [25] He, K., Zhang, X., Ren, S., Sun, J., Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, vol. 37, No. 9, 2015, pp. 1904-1916.
- [26] Girshick, R., Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.
- [27] Sun, H., Liu, X., Min, G., Zhou, J., Zhang, W., Fully Convolutional Network Based Ship Plate Recognition. In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1803-1808, Oct. 2018.
- [28] Jiang, H., Learned-Miller, E., Face detection with the faster R-CNN. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 650-657, May 2017.
- [29] Sun, X., Wu, P., Hoi, S. C., Face detection using deep learning: An improved faster rcnn approach. Neurocomputing, vol. 299, 2018, pp. 42-50.
- [30] Li, J., Liang, X., Shen, S., Xu, T., Feng, J., Yan, S., Scale-aware fast R-CNN for pedestrian detection. IEEE transactions on Multimedia, vol. 20, No. 4, 2018, pp. 985-996.
- [31] Dai, J., He, K., Sun, J., Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3150-3158.
- [32] Li, Y., Qi, H., Dai, J., Ji, X., Wei, Y., Fully convolutional instance-aware semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2359-2367.
- [33] Dai, J., He, K., Li, Y., Ren, S., Sun, J., Instance-sensitive fully convolutional networks. In European Conference on Computer Vision, Springer, Cham, pp. 534-549, Oct. 2016.
- [34] Dai, J., Li, Y., He, K., Sun, J., R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems, 2016, pp. 379-387.
- [35] R. Liu, Principle and application of TFDS. Chin. Railw, vol. 5, 2005, pp. 26-27.

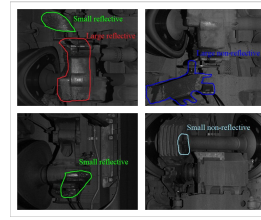


Fig. 3. Different kinds of oil-leak, We use red color denotes the large reflective, blue is the large non-reflective, green is the small reflective and cyan is the small non-reflective.

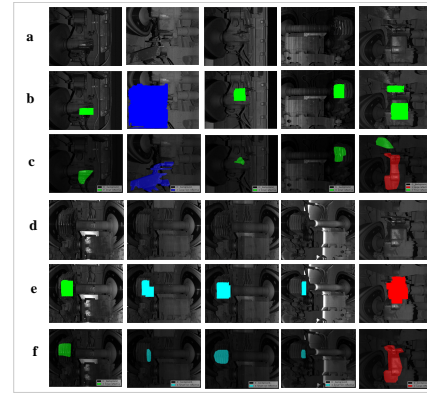


Fig. 4. The H-CNN detection results on different oil-leak stains. Row *a* and *d* are raw images, row *b* and *e* are H-CNN detection results, and row *c* and *f* are benchmarks. We use red color denotes the large reflective, blue is the large non-reflective, green is the small reflective and cyan is the small non-reflective. The benchmarks are labeled using "labelme".

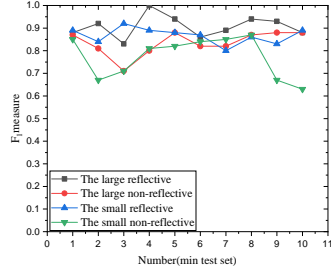


Fig. 5.  $F_1$  –measure of the H-CNN detection results on different oil-leak stains using different test sets.

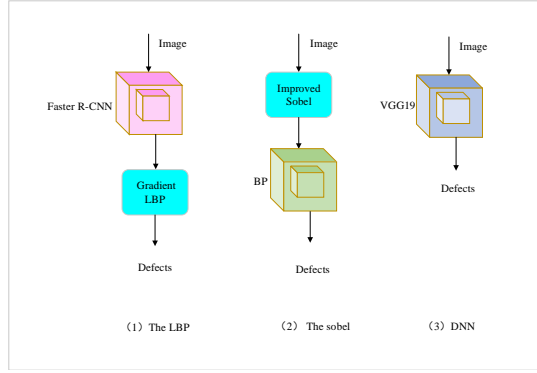
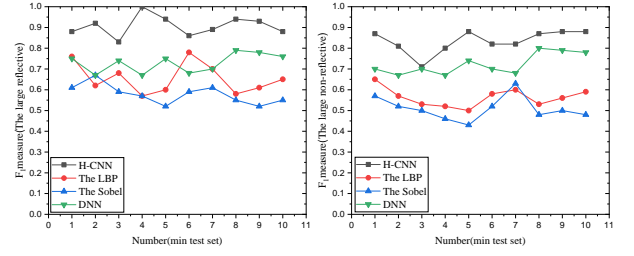
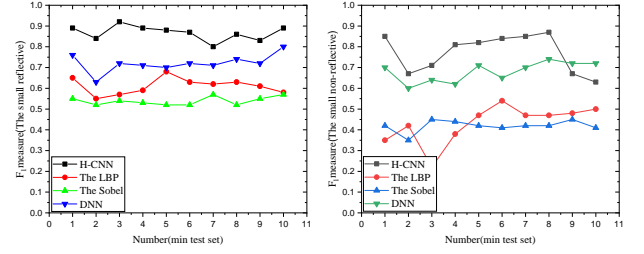


Fig. 6. The compared three defect detection models.



(a)

(b)



(c)

(d)

Fig. 7. Four methods' detection results on four oil-leaks. (a)The large reflective (b)The large non-reflective (c)The small reflective (d)The small non-reflective.