"© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works."

Weak Scratch Detection of Optical Components Using Attention Fusion Network

Xian Tao
Research Center of Precision Sensing and Control, Institute of Automation
Chinese Academy of Sciences
Beijing, China
taoxian2013@ia.ac.cn

Dapeng Zhang
Research Center of Precision Sensing and Control, Institute of Automation
Chinese Academy of Sciences
Beijing, China
dapeng.zhang@ia.ac.cn

Avinash K Singh
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
avinash.singh@uts.edu.au

Mukesh Prasad
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
mukesh.prasad@uts.edu.au

Chin-Teng Lin
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
chin-teng.lin@uts.edu.au

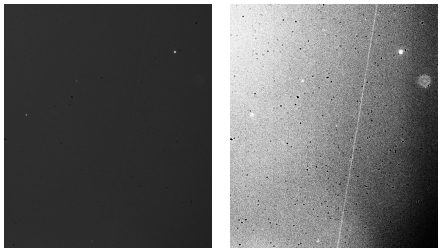
De Xu
Research Center of Precision Sensing and Control, Institute of Automation
Chinese Academy of Sciences
Beijing, China
de.xu@ia.ac.cn

***Abstract*—Scratches on the optical surface can directly affect the reliability of the optical system. Machine vision-based methods have been widely applied in various industrial surface defect inspection scenarios. Since weak scratches imaging in the dark field has an ambiguous edge and low contrast, which brings difficulty in automatic defect detection. To address the problems arising from industry-specific characteristics, this paper proposes "Attention Fusion Network", a convolutional neural network using attention mechanism built by hard and soft attention modules to generate attention-aware features. The hard attention module is implemented by integrating the brightness adjustment operation in the network, and the soft attention module is composed of scale attention and channel attention. The proposed model is trained on a real-world industrial scratch dataset and compared with state-of-art defect inspection methods. The proposed method can achieve the best performance to detect the weak scratch inspection of optical components compared to the state-of-art.**

Keywords—weak scratch inspection, attention fusion networks, machine vision, convolutional neural network (CNN), optical component

I. INTRODUCTION

In modern optical systems, large-aperture optical elements are essential components that are used for optical transmission and energy conversion. The small defects on surface, especially scratches, can lead to dramatic energy loss and system failure [1, 2]. Nowadays, the most common and effective way for surface scratch detection is the machine vision technology, which uses a CCD (Charge Coupled Device) and a light to form a dark field imaging system (DFIS). However, the DFIS possesses two challenges in scratch detection. First, weak scratches are usually at sub-micron in-depth, which causes the brightness of their pixels very similar to the background of the defective image. These pixels cannot be separated by choosing the right threshold only. Second, the defective image contains noise whose influence cannot be eliminated by brightness enhancement operation, as shown in Fig. 1. Fig.1 (a) shows that the intensity difference between the scratch and background is very low. Fig.1 (b) shows the result of the scratch image equalized by the histogram. It can be seen that scratches and background noise are enhanced at the same time, which makes it difficult to detect defects.



(a) (b)

Fig. 1. Dark-field image with weak scratches (a) Original image, (b) Image after histogram equalization

Currently, the detection methods for surface scratches of optical elements are majorly based on filtering algorithms. These filtering algorithms can enhance the image quality for further processing. Li et al. [3] proposed the dual-threshold classification method to extract scratches based on spatial and frequency domain filtering algorithms. In our previous work [4], the Gabor filter-based approaches were proposed, which uses prior knowledge to detect ambiguous scratch. Similarly, Luo et al. [5] proposed an automated scratches detection module. This automated scratch module first filters out the significant scratches and focus on the small scratches. The methods mentioned above simulate a simplified bottom-up visual attention mechanism. However, the filtering rules used in these methods are mostly hand-crafted and easily failed in the new imaging environment.

Recently, convolutional neural networks (CNN) became one of the most attractive methods for surface defect detection due to the end-to-end learning mechanism. CNN has been successfully applied in many defect detection scenarios, such as wafer [6], photovoltaic module cell [7], and weld defect [8]. These models are often built based on existed frameworks, such as VGG [9], ResNet [10], Inception V3 [11], etc. However, the performance of CNN based models is still inadequate for weak scratch defect detection. Moreover, some of defect detection methods based on detection and segmentation networks need to label bounding boxes or pixels. The cost of such industrial data collection and labeling is very high.

This work proposes a novel method by combining the visual attention mechanism together with CNN to deal with

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

the challenges of weak scratch detection. The attention module of the proposed method is a combination of brightness adjustment (BA) modules, scale attention (SA), and channel attention (CA) module.

The proposed method is described in Section II, followed by experimental results in Section III — finally, conclusion and future work in Section IV.

II. Proposed Method

A. Model

The most common approach to improve defect detection is to introduce pre-training weights into CNN. Such an approach does not use any specific network structure or characteristics of the defect. In the proposed model, the attention mechanism from the traditional scratch detection method is integrated into the designed convolutional network structure. The proposed model network structure is based on ResNet18 [10]. BA, SA, and CA modules are integrated into the structure for fine-grained feature extraction, as shown in Fig. 2.

The BA module is at the forefront of the proposed model and utilizes the raw dark-field image as network input. In this paper, the raw dark-field image is processed by the BA module to obtain a brightness-adjusted image, which is concatenated to the raw image and input to the subsequent network. The first convolution block of the model uses a 7×7 kernel to obtain the effect of a large receptive field. A max-pooling layer follows this convolution layer. Compared with the original ResNet18, the proposed network added a SA module behind each Res-block. The model contains four SA modules. A CA module is then added, followed by the last SA module. Finally, the model performs global average pooling, resulting in one output neurons.

B. Brightness Adjustment Module

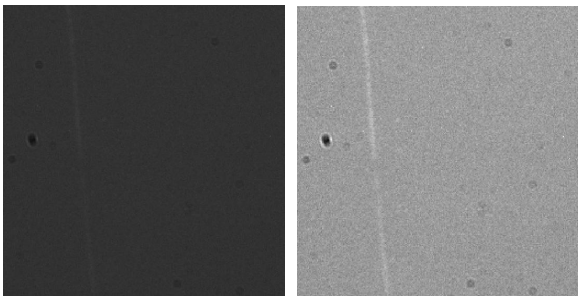
Different from natural images in the Imagenet dataset [12], scratches in dark-field images are characterized by large spans and low contrast. In the lower stages of CNN, the feature maps contain low-level spatial information such as edges and texture, which were very beneficial for scratch representation. To speed up the feature extraction of the low-contrast information in the model, a BA module based on hard attention is proposed. Instead of region proposed [13, 14] methods, we enhance the whole raw dark-field image, then concatenate it with the raw image as a network input.

Inspiring from traditional image processing methods, the BA module is proposed to be embedded in the end-to-end network for training. The operation of the BA module is defined as follows:

$$W_{mult} = 255/(V_{max}-V_{min}) \quad (1)$$

$$W_{add} = -W_{mult}*V_{min} \quad (2)$$

$$I_{BA}= I * W_{mult}+ W_{add} \quad (3)$$



(a) (b)

Fig 3. The input and output of BA, (a) Input dark-field image with weak scratches, (b) output tensor of brightness adjustment module

where I_{BA} denotes the BA module's output tensor, I denote the input image; and W_{mult} and W_{add} represent the adjustment coefficient. V_{max} and V_{min} correspond to the maximum and minimum gray values of the input image I, respectively. Fig. 3(a) is the original dark-field image with weak scratches. Fig. 3(b) shows the output tensor of the BA module. As can be seen from Fig.3, weak scratches are enhanced to the extent that the output of the BA module is visually apparent.

C. Scale Attention Module

A challenge in scratch inspection is the considerable variation of size in the dark-filed image. Some works [15, 16] improved the multi-scale ability by utilizing different-size convolution kernel. Different from above methods, the atrous convolution is an effective way to increase the receptive field, and has a large dilation rate to provide a large receptive field. Inspired by the work in [17], a modified version of the Res2Net module is proposed to fulfill the requirement for SA, named Res2Net-A, shown in Figure 4.

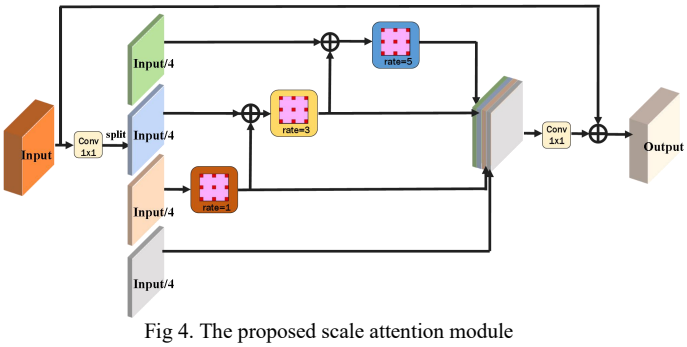


Fig 4. The proposed scale attention module

The SA module mainly uses the structure of the Res2Net model. After the 1×1 convolution, the feature maps are split into four feature map subsets. Each feature subset has the same spatial size, but 1/4 number of channels compared with the input feature maps. Each feature group has its corresponding atrous convolution. The feature group of the current layer adds the result of the next feature group by performing the convolution operation. Finally, the feature maps from all groups are connected and sent to another group of 1 × 1 filters to fuse the information.

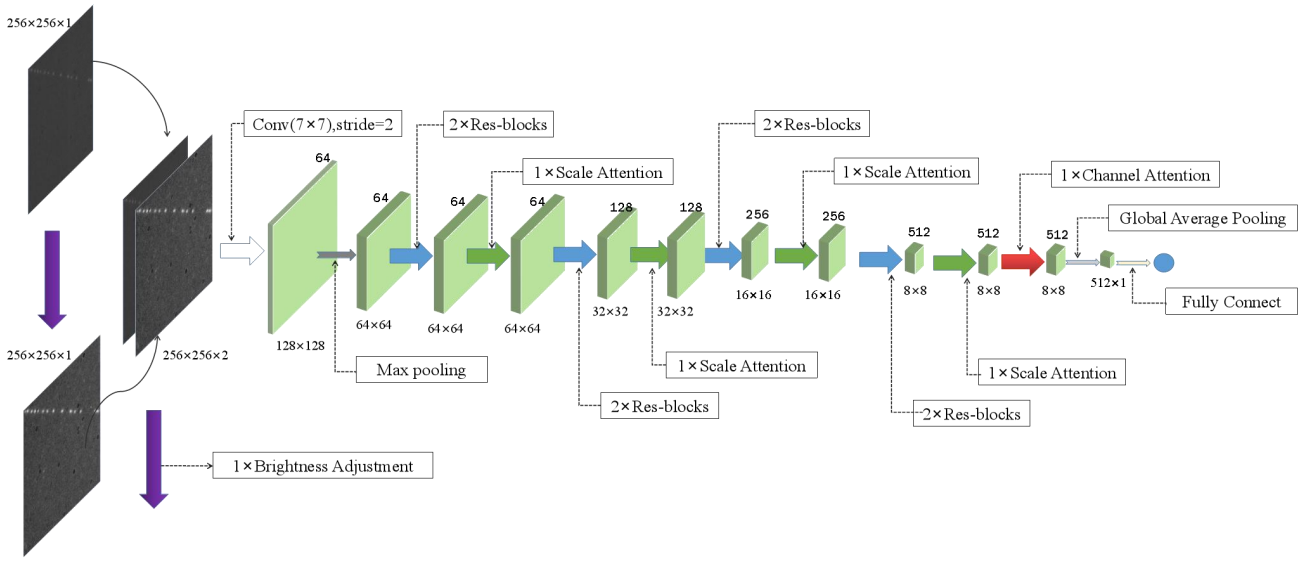


Fig 2.The architecture of the proposed model

The SA module contributes to the proposed model by collecting contextual information at different scales for each feature map and then explores the differences in scale. This mechanism is particularly helpful in improving the ability to discriminate scratch defects in a large spatial range of the proposed model.

D. Channel Attention Module

In the CA module, the SENet [18] is adopted to improve the representational capacity of the network. The shape of input feature maps P is H×W×C. A global average pooling layer is first applied to the CA module, followed by two 1x1 convolution operations. Then, the result is passed through a sigmoid layer to obtain the channel weight σ_c. By multiplying channel weight with P, a new feature map P_c is obtained. The operation is shown in equation (4).

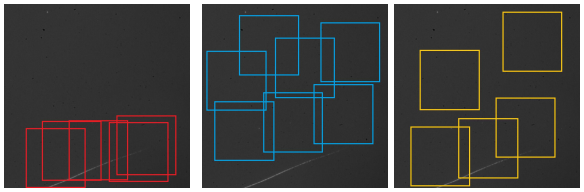
$$P_c = \sigma_c \otimes P \tag{4}$$

By re-calibrating the feature map P, it highlights the scratch features and suppresses background noises. More useful features have been extracted from the feature attention approach. Such information could also help to inspect results.

III. EXPERIMENTS

A. Dataset

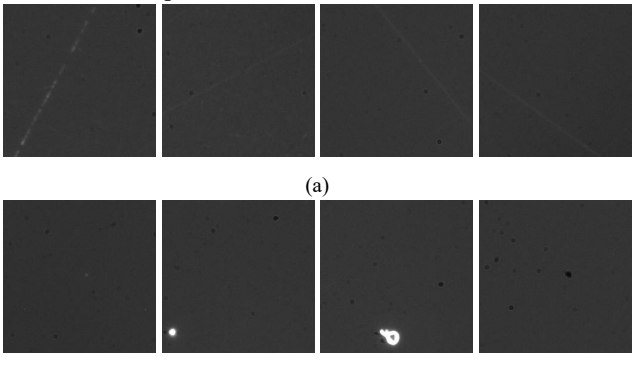
The dataset used in the experiment is collected using an optical component of 810 mm × 460 mm. We designed an inspection instrument for large-aperture optical elements with the DFIS and a line scan camera in our previous work [19]. We scanned the front and back of three large-aperture optical elements and obtained a total of 112 dark-filed defective images, with a resolution of 2048 × 2048 pixels. From the obtained images, we divided them in 80, 12, 20 for training, validation, and testing, respectively. Since the scratches are relatively sparse in the entire image, instead of cropping the big image into a fixed size, we have used three image sampling methods to build the dataset for network training. The first sampling method is Gaussian probability distribution, which is centered on scratches, as shown in Fig.5 (a). This method mainly focuses on scratch samples and can obtain a large number of defect images. The second sampling method is based on uniform distribution, as shown in Fig.5 (b). This methodology can obtain a large number of background images. The third method is random sampling, which is mainly used to supplement the sample.



(a) (b) (c)

Fig 5. Sampling methods, (a)Gaussian probability distribution sampling. (b) uniformly distributed sampling. (c) random sampling

The raw images were sampled into smaller images of 256 ×256-pixel resolutions to build the dataset. Each image is labeled as either a background or a scratch image. The total numbers of training and test samples in each category are listed in Table I. The numbers of training samples, and test samples were 30920 and 1461, respectively. Fig. 6 shows some of the samples from the dataset.



(a)

(b)

Fig 6. Image samples in dataset.(a)scratch. (b) background.

TABLE I. NUMBERS OF TRAINING AND TEST SAMPLES IN EACH CATEGORY OF DATASET

	Train	***Validation***	***Test***
Scratch	16800	582	754
Background (Non-Scratch)	14120	414	707
Sum	30920	996	1461

B. Implementation Details

Our method was implemented using the deep learning platform of Tensorflow [20]. The proposed network uses Adam [21] for optimization training. The initial learning rate sets to 1e-4 and reduced by 10 after three times training while observing the validation loss decreasing slowly. The batch size for training is 32, and a total of 40 iterations of the training network. The following results were obtained using a computing unit with Intel Core i7 and NVIDIA GTX-1070 with 8 GB of video memory.

C. Comparisons with Other Methods

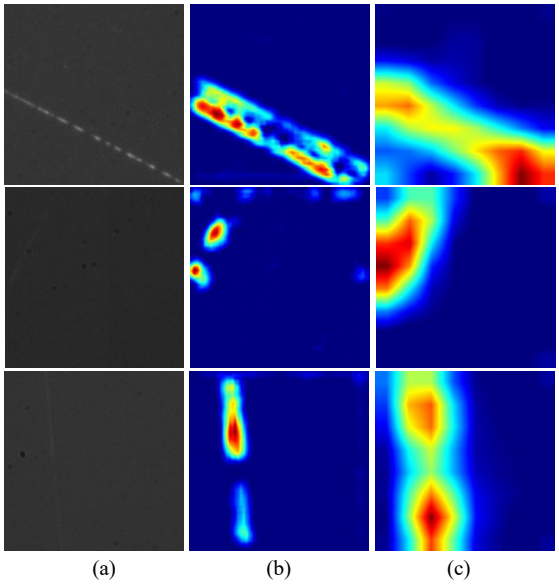
In this experiment, we evaluated the effectiveness of the proposed method, with existing attention mechanisms networks [22] and the state-of-the-art classification methods [11, 23]. The attention network [22] integrates the soft attention mechanism (a bottom-up and top-down mask branch) with residual blocks. We compared our proposed method with the state-of-the-art classification methods on inception structure and separable convolutions, based on Imagenet pre-training weights. The results of a comparison between the proposed method and the state-of-art methods are shown in Table II.

Our proposed method outperforms Attention-56 and Attention-92 with a large margin. The increment on f1-score is 19.23% and 17.59%. Compared with the three state-of-the-art networks, our proposed network can achieve better performance without complex structure and pre-training weights. Such results indicate that the proposed attention fusion network can significantly improve network performance.

TABLE II. COMPARATIVE EXPERIMENT ON TEST DATASET

Model	*Precision*	*Recall*	*f1-score*
Attention-56 [22]	0.8621	0.5285	0.6552
Attention-92[22]	0.9821	0.5106	0.6719
Xception + pretraining [23]	**0.9952**	0.5557	0.7132
InceptionV3 +pretraining [11]	0.9573	**0.7440**	0.8373
Our proposed method	0.9877	0.7427	**0.8478**

D. Visualization of attention maps



(a) (b) (c)

Fig 7. Visualization of attention maps (heatmaps). From left to right column: raw images, attention maps of Attention-92[22] and the proposed method.

We use the grad-CAM method [24] to visualize the attention maps. Figure 7 shows the results of the visualization. Compared with the existing attention mechanisms networks [22], our proposed network can more accurately focus on the scratched area. It indicates the importance of attention fusion and passing the informative features originating from the three attention modules.

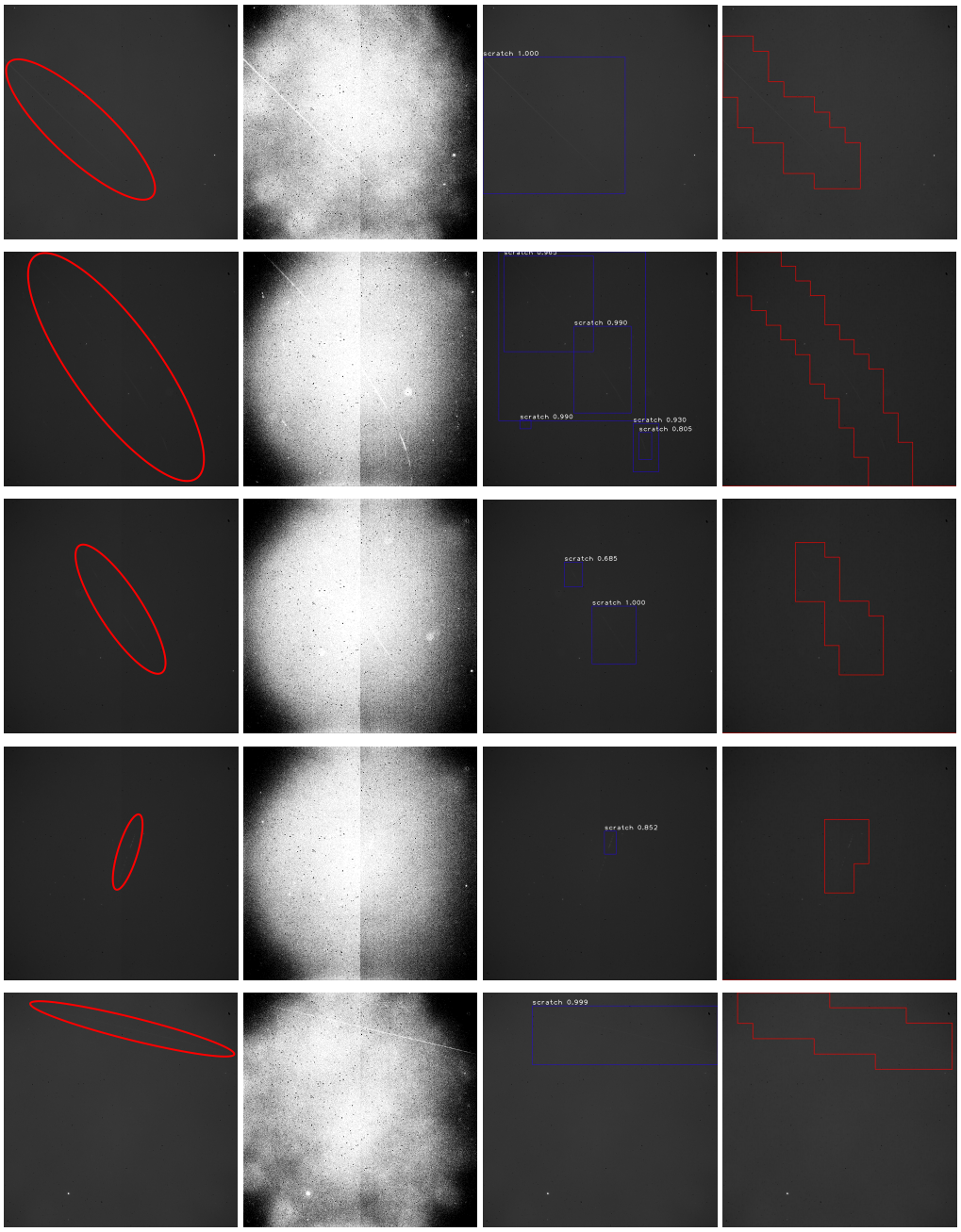
E. Inspection Performance

In this experiment, we compared the performance of the proposed weak scratch inspection method with two other methods, including the well-known traditional method (Ostu [25]) and the object detection network method (Retinanet [26]). In order to locate the position of the scratch, a path scanning method of a redundant sliding window was adopted in the proposed method. The sliding window size was 256×256 pixels, and the redundancy was 128 pixels. The sliding window image was input into the proposed network for binary classification. The contour of all sliding windows containing scratches is the localized defect area. Moreover, 80 images of size 2048×2048 pixels were used for Retinanet training, and the bounding rectangle of the scratch was used as the object detection label. The three inspection methods were conducted on 20 raw images of size 2048×2048 pixels. The scratch inspection results are shown in Figures 8. Although the Ostu detection method provides some scratch information, as shown in Figure 8(b), it does not achieve the positioning of scratches, and it is also a challenge to separate the scratches from the background due to the influence of background noise. It has even appeared that the scratches are submerged in the background noise, for example, the 4th-row scratch sample in Fig. 8.

Figure 8(c) shows the results of the Retinanet. This case shows that Retinanet detects the weak scratch accurately. However, compared to the proposed method, the Retinanet could miss the detection of scratched areas, such as the images of the third and fourth rows in Fig. 8. This happens due to too few training samples of the Retinanet, which are only 80 images. At the same time, the same scratch may be detected multiple times in the Retinanet, such as the image on the second row in Fig. 8. Compared with the proposed method, the Retinanet needs to prepare a large number of sample images of well-labeled regions. In this comparative experiment, the proposed method shows strong performance in various weak scratch detection of the raw image compared with the traditional binarization method and the latest object detection method.

F. Ablation Analysis

Our proposed method is based on resnet18, so resnet18 is the most fundamental baseline model. To improve the learning ability of weak scratch features, we added the blocks after the last residual block of the resnet18 network. Moreover, the BA model is added at the forefront of the network, and the SA model is added after the residual modules. As shown in Table III, our proposed CA module and BA (called resnet18+CA+BA) increased the f1-score from 0.8172 to 0.8347 in the test dataset and greatly improved the recall of scratch detection. This shows that two attention modules help to extract advanced features of scratches. We also compare the proposed SA with the CA module (called resnet18 + CA + SA). The comparison results show that the value of the f1-score is further increased from



(a) (b) (c) (d)

Fig 8. Scratch inspection results. From left to right column: raw images(scratch areas are marked in red), results of Otsu[25], retinanet[26], and the proposed method.

TABLE III. ABLATION EXPERIMENTS ON THE TEST DATASET

Model	*Precision*	*Recall*	*f1-score*
resnet18	0.9943	0.6936	0.8172
resnet18 +CA	**0.9962**	0.6950	0.8187
resnet18 +CA+BA	0.9927	0.7202	0.8347
resnet18 +CA +SA	0.9946	0.7361	0.8460
Our proposed method (resnet18 +CA+SA+BA)	0.9877	**0.7427**	**0.8478**

0.8347 to 0.8460. Finally, our proposed method achieved the highest f1-score. This shows that the proposed three attention modules can further focus on weak scratch feature information, which is useful for our defect detection task. Compared with the high precision, the proposed method does not have high recall in the test dataset. Actually, in the real detection environment, there are many discontinuous point-like scratch defects. The falsely classified samples are shown in Fig. 9.

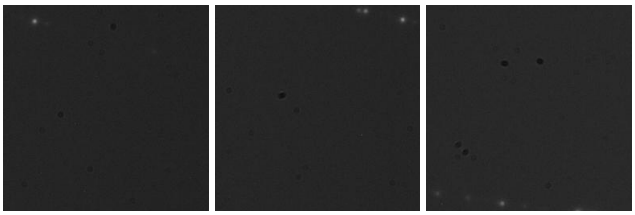


Fig 9. The falsely classified samples in the test dataset

The discontinuous point-like scratch defect is very similar to dust on the background, consisting of even three discontinuous points. At the same time, due to the small proportion in the defective image, it is difficult to identify

directly through the classification network. After removing this special type of scratch, the recall of the test dataset can reach more than 95%.

IV. Conclusion and Future Work

In this paper, a novel deep learning framework is developed for weak scratch inspection. In the framework, an attention fusion network is proposed, which includes hard and soft attention modules. The hard attention module is implemented by integrating the BA operation in the network, and the soft attention module is composed of SA and CA. The proposed method outperforms traditional and other state-of-the-art deep learning-based methods on a real-world weak scratch dataset.

Future extensions of this paper can focus on discontinuous point-like defect and unlabeled weak defect detection tasks. Discontinuous point-like scratch defects can be detected using segmentation networks. Moreover, since manual labeling of defects is a laborious and expensive process, defect detection using only normal samples would be a very promising task. For example, generative adversarial networks (GANs) [27] are used to expand defective samples and inspect defects. Therefore, our future work will focus on the automatic detection of defects without labels.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 61703399, 61973302 and Science Challenge Project(No.TZ2018006 -0204-02),

References

[1] Ota H, Hachiya M, Ichiyasu Y, et al. Scanning surface inspection system with defect-review SEM and analysis system solutions[J]. Hitachi Review, 2006, 55(2): 79.

[2] Ding W D, Zhang Z T, Zhang D P, et al. An effective on-line surface particles inspection instrument for large aperture optical element[J]. International Journal of Automation and Computing, 2017, 14(4): 420-431.

[3] Li C, Yang Y, Xiong H. Dual-threshold algorithm study of weak-scratch extraction based on the filter and difference[J]. High power laser and particle beams, 2015, 27(7).

[4] Tao X, Xu D, Zhang Z T, et al. Weak scratch detection and defect classification methods for a large-aperture optical element[J]. Optics Communications, 2017, 387: 390-400.

[5] Luo Z, Xiao X, Ge S, et al. ScratchNet: Detecting the Scratches on Cellphone Screen[C]//CCF Chinese Conference on Computer Vision. Springer, Singapore, 2017: 178-186.

[6] Kyeong K, Kim H. Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks [J]. IEEE Transactions on Semiconductor Manufacturing, 2018, 31(3): 395-402

[7] Deitsch S, Christlein V, Berger S, et al. Automatic classification of defective photovoltaic module cells in electroluminescence images [J]. Solar Energy, 2019, 185: 455-468.

[8] Zhang Z, Wen G, Chen S. Weld image deep learning-based on-line defects detection using convolutional neural networks for Al alloy in robotic arc welding[J]. Journal of Manufacturing Processes, 2019, 45: 208-216.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[11] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[12] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

[13] Dai J, He K, Sun J. Convolutional feature masking for joint object and stuff segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3992-4000.

[14] Hariharan B, Arbeláez P, Girshick R, et al. Simultaneous detection and segmentation[C]//European Conference on Computer Vision. Springer, Cham, 2014: 297-312.

[15] Gu Z, Cheng J, Fu H, et al. CE-Net: Context Encoder Network for 2D Medical Image Segmentation[J]. IEEE transactions on medical imaging, 2019.

[16] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[17] Gao S H, Cheng M M, Zhao K, et al. Res2Net: A New Multi-scale Backbone Architecture[J]. arXiv preprint arXiv:1904.01169, 2019.

[18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[19] Tao X, Zhang Z, Zhang F, et al. A novel and effective surface flaw inspection instrument for large-aperture optical elements[J]. IEEE Transactions on Instrumentation and Measurement, 2015, 64(9): 2530-2540.

[20] Abadi M, Agarwal A, Barham P, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems[J]. arXiv preprint arXiv:1603.04467, 2016.

[21] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[22] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3156-3164.

[23] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.

[24] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 618-626.

[25] Ostu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man and Cybernetics, 1979, 9(1): 62-66.

[26] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[27] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.