



An Automatic Surface Defect Detection Method with Residual Attention Network

Lei Yang^(✉), Suli Bai, Hanyun Huang, and Shuyi Kong

School of Electrical and Information Engineering, Zhengzhou University,
Henan 450001, China
leiyang2019@zzu.edu.com

Abstract. Automatic defect detection is of great significance to the production process of modern industries, which will affect the product appearance and quality. And a great number of economic loss may be caused due to the defects in the industrial production process. Traditional manual defect inspection method is labor-intensive and time-consuming with subjective factors, and the detection performance is random and uncertain. Machine learning has shown a good detection ability on small-scale samples, but the detect detection task against with poor contrast, weak texture, etc., will affect the effective feature representation. To address the detection task of steel surface defects, combined with the strong context extraction ability of deep learning, an accurate deep defect classification network is proposed in the paper to provide an end-to-end detection scheme. Fused with the residual network (Resnet50) and spatial attention block, a residual attention network is proposed for effective feature representation, which could make the classification network better focus on the defect areas. Meanwhile, due to the scale information change among different defects, a multi-scale context fusion (MCF) block is proposed for effective multi-scale feature extraction, which is conducive for multi-scale object detection. Experimental results on public defect data set show that the proposed defect detection network could acquire a superior classification performance compared with some typical classification networks.

Keywords: Surface defects · Deep learning · Residual network · Multi-scale context fusion · Channel attention

1 Introduction

Surface defect detection is of great significance for the development of modern industrial manufacturing, which plays an important role in many industrial production tasks [13]. Defects may be generated in any production link in the industrial manufacturing process, and sometimes serious consequences will occur when defective products enter the market to bring the economic loss. Traditional manual defect inspection method is labor-intensive and time-consuming [16], which can not fully guarantee detection accuracy and efficiency of intelligent

manufacturing, so it is not suitable for the strict requirements of high precision and real-time detection in modern large-scale industrial production processes [3]. Therefore, an automatic defect detection method with high accuracy and good robustness is still a meaningful but challenging task for the modern industrial manufacturing.

In recent years, due to the success on effective context feature extraction, convolutional neural networks (CNNs) have achieved an excellent performance in various computer vision fields [9], which could directly process the raw images and extract high-level semantic information, and have been widely applied into different tasks, such as object detection [15], object classification [1], image segmentation [17]. And it could acquire a superior detection performance compared traditional image processing methods and machine learning methods. To improve the detection performance, different network models have been proposed, such as ResNet [6], VGG16 [10], AlexNet [8], etc. With the increasing popularity of deep learning, researchers have explored the applications of deep learning on effective and accurate non-destructive testing (NDT) [2, 15, 18]. In order to solve the misclassification problem caused by background chaos and structural interference, Yang *et al.* proposed an improved SqueezeNet network for surface defect inspection, which combined multiple convolution kernels with different kernel sizes to increase the receptive field, thus obtaining multi-scale features [12]. Faced with the fuzzy edges and unfixed shapes of weak scratches, Tao *et al.* proposed a new defect detection network that automatically detected weak scratches by gathering rich multidimensional features, which included a feature fusion block and a context fusion block to perform multi-scale feature fusion of high-level information and low-level information [11]. Experiments results indicated that it could well overcome the phenomenon of the long span and connectivity of weak scratches. To address the accurate detection tasks of small-scale defect detection, Geng *et al.* proposed a deep convolutional generative adversarial network (DCGAN) and seam carving algorithm to achieve intelligent defect detection of water wall in thermal power plants [4]. Faced with the low-quality images DCGAN was proposed to improve the image quality, and the seam carving algorithm was proposed to solve the overfitting issue of DCGAN network. On the basis, a deep CNN network was built for intelligent defect detection. Combined with the pre-trained deep network, Yang *et al.* proposed a welding defect detection method based on multi-scale feature fusion from X-ray welding images [14]. Due to the different feature representation ability of different network layers in the pre-trained AlexNet network, the features from different network layers were acquired as the multi-scale features. On the basis of multi-scale features, a defect detection model was proposed based on the support vector machine (SVM) classifier and Dempster-Shafer evidence theory to realize multi-scale feature fusion. Inspired by the above work, a deep defect classification network is proposed in the paper for accurate detection of steel surface defects.

Based on the above discussion on defect detection, an end-to-end defect classification scheme for steel surface defects is proposed in this paper to provide an end-to-end detection scheme, as shown in Fig. 1. Experiments on public defect

data set show that the proposed defect detection network could acquire a superior detection performance compared with some typical classification networks, which could further prove the feasibility and superiority of the proposed detection scheme. The main contributions in this paper are drawn as follows:

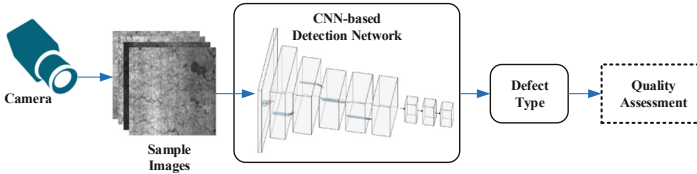


Fig. 1. Diagram of proposed defect detection scheme.

- (1) An end-to-end defect classification network is proposed in this paper for accurate defect detection.
- (2) To realize effective feature representation, a residual attention network is proposed to act as the backbone network to acquire high-level contexts.
- (3) Faced with the multi-scale steel surface defects, a MCF block is proposed for effective multi-scale feature extraction and fusion.

The rest part of this paper is given as follows. Section 2 gives the description of proposed defect classification network. Section 3 introduces the experimental data set and the corresponding image processing method. Section 4 is about the experimental results and analysis. The conclusions of this paper are given in Sect. 5.

2 Proposed Methodology

To address the detection task of steel surface defects, a deep defect classification network is proposed in the paper to provide an end-to-end and accurate detection scheme. This section gives the detailed description about proposed defect classification network and each network block.

2.1 Overview of Network Framework

Faced with the defects of steel surface, an end-to-end defect classification network is proposed in this paper for automatic and accurate defect detection. Figure 2 shows the whole network structure of proposed defect detection network.

As shown in Fig. 2, the whole network framework is composed of three parts: backbone network, MCF block, and classification block. Combined with the Resnet50, a residual attention network is proposed to act as the backbone network for feature representation. Here, a spatial attention block is proposed to embed into each residual block to acquire and learn the importance of feature

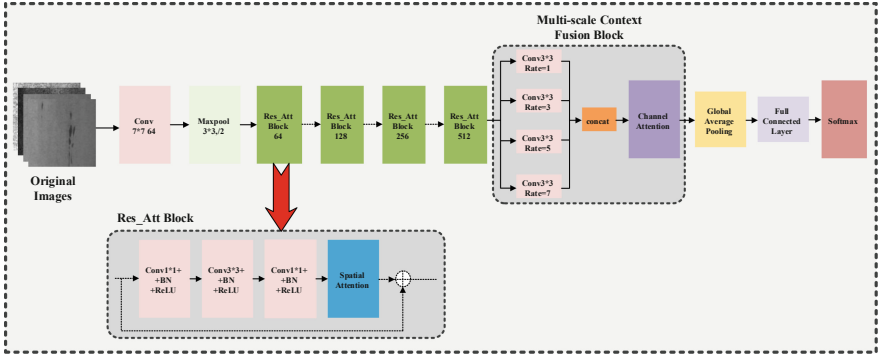


Fig. 2. Network framework of the proposed defect detection network.

maps. Combined with the atrous spatial pyramid pooling block and channel attention block, a MCF block is proposed, which could acquire different receptive fields to obtain multi-scale contexts. And the multi-scale contexts are fed into the classification block, which consists of the global average pooling layer, the fully connection layer and the softmax function.

2.2 Backbone Network

The residual network [6] is a typical deep classification network, which has been widely applied into different deep models for effective feature representation, which could effectively alleviate the problems of gradient disappearance and gradient explosion in neural network and protect the integrity of information. Due to the success of residual network, the Resnet50 network is proposed in this paper to act as the baseline network. As shown in Fig. 2, the basic structure of backbone network is divided into four main units and each unit has a similar structure, which consisting of a series of bottleneck layers contained 3, 4, 6, and 3 residual attention (Res_Attn) blocks separately.

To make the classification network better focus on the defect areas, combined with the advantage of attention block, a spatial attention block [5] (see Fig. 3) is proposed to embed into each residual block to construct the residual attention block, which could capture the weight of different feature channels, thus to focus on the key attention features on defects and suppress irrelevant information.

Combined with the max-pooling layer and average pooling layer, the feature descriptors by these two pooling operations are concatenated together, which are fed into a 7×7 convolution layer followed by a sigmoid function to acquire the final spatial attention weight. And the final attention maps could be acquired through the element-wise multiplication with the raw feature maps.

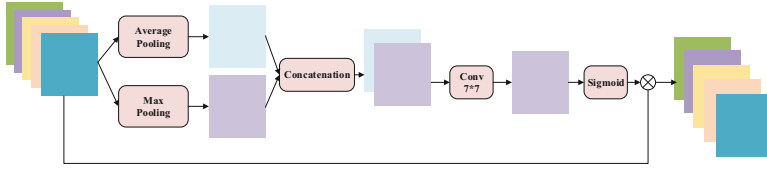


Fig. 3. Network structure of spatial attention block.

2.3 MCF Block

Faced with the defect detection task, except for the shape information, the scale information change among different defects is also a common phenomenon, which puts a strict demand on multi-scale defect detection. For the typical deep network, the high-level network layer can capture a large receptive field and has a strong ability to express abstract and high-level semantic information but a small image resolution. On the contrary, for the shallow network layer, it has a small receptive field and a relatively high resolution but a weak ability to represent effective semantic information. In order to accurately identify multi-scale steel surface defects, a MCF block is proposed in this paper to capture effective multi-scale features (see Fig. 2).

Compared with the standard convolution, the dilated convolutions with different dilated rates could well acquire different receptive fields without the increase of computing cost [19]. Due to this advantage of dilated convolution, as shown in Fig. 2, combined with the feature maps from the residual attention network, these feature maps are processed by four dilated convolutions with the convolution rates of 1, 3, 5, 7 respectively, which is equivalent to the extraction of multi-scale features. And the multi-scale feature maps are concatenated together for feature fusion.

Faced with the multi-channel feature maps generated by multiple network branches of dilated convolutions, a channel attention block is proposed to learn the important information of different channels. The squeeze-and-excitation (SE) block (see Fig. 4) is a typical channel attention block for channel calibration [7]. In order to identify steel defects more accurately and focus on the feature channels with the most information, the SE block is proposed for channel calibration to acquire the channel weights.

3 Experiment Data and Preprocessing

For the effective model evaluation, the suitable data set is the premise for the performance of proposed defect classification network. This section introduces the details about the experimental data set and the corresponding image processing method.

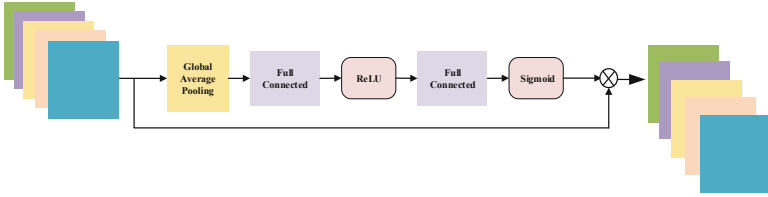


Fig. 4. Network structure of channel attention block.

3.1 Data Set

In order to verify the detection performance of the proposed defect classification network, the NEU metal surface defect data set collected by Northeast University is adopted in this paper for model evaluation, which contains of six types of defects: crazing, inclusion, patches, pitted surface, rolled-in scale and scratches. Figure 5 shows the sample images of these defects. It could be seen that these defect images are against with the poor contrast, weak texture, etc., which will bring a certain effect to accurate defect detection.

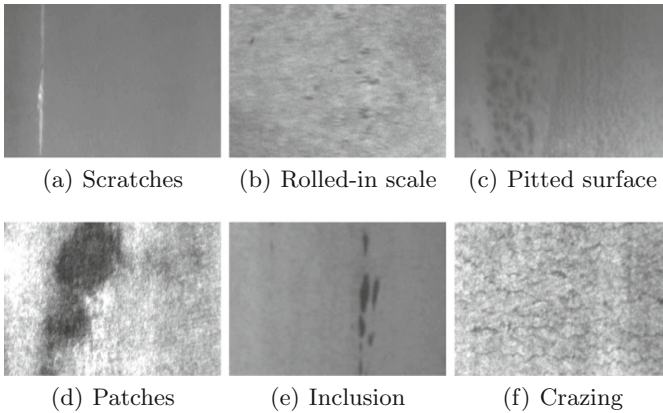


Fig. 5. Sample images of different defects in the NEU metal surface defect data set.

For the NEU metal surface defect data set, each defects contain 300 sample images. The detailed information of various defects in this data set is given in Table 1.

Combined with this defect data set, in order to evaluate the proposed defect classification network more accurately, the data set is divided into training set, validation set, and test set by a ratio of 7:1:2.

Table 1. Parameters about the data set and model training.

Index	Parameter	Value
1	Image size	$200 \times 200 \times 1$
2	Crazing	300
3	Inclusion	300
4	Patches	300
5	Pitted surface	300
6	Rolled-in scale	300
7	Scratches	300

3.2 Image Preprocessing

The effective model training of deep CNNs, it always relies on enough data set to support the detection performance. Data augmentation provides an effective tool for the small-scale data samples and brings great convenience to the deep learning.

In order to improve the detection performance and avoid the overfitting issue, the data augmentation is proposed in this paper to preprocess the raw images to enlarge the data set. For the training set, combined with the traditional machine vision algorithms, several typical image analysis algorithms are proposed in this paper for image preprocessing, including shear transformation, angle rotation, image brightness adjustment, contrast adjustment, saturation adjustment and chroma adjustment.

4 Experiment Results and Analysis

In this section, to effectively evaluate the proposed defect classification network, the comparative experiments and ablation experiments are carried out respectively to further illustrate the advantages of the proposed defect classification network.

Firstly, the implementation details of the proposed defect classification network is described in detail. Secondly, the ablation experiments are given to show the effectiveness of proposed each network block. Finally, the comparison experiments are provided to further illustrate the superiority of the proposed defect classification network.

4.1 Implementation Details

For the proposed defect classification network, it is built with the PaddlePaddle framework¹. To speed up the model training and test, the related experiments are carried out in the NVIDIA Tesla V100 GPU card with 16 GB memory.

¹ <https://www.paddlepaddle.org.cn/>.

In addition, some hyper-parameters of the proposed defect classification network need to be set. Here, the Adam optimizer is adopted, and the initial learning rate is set as 0.01. With the increase of training epoches, the attenuation factor of learning rate decreases gradually. In order to avoid network overfitting, the early stop strategy is proposed. To guide the model training, the cross entropy loss is adopted to act as the loss function. Based on the memory size of GPU card, the batch size is set as 32.

To precisely evaluate the classification ability of different models, some evaluation indicators are introduced for quantitative analysis, including accuracy, precision, recall and F_1 score.

4.2 Ablation Study

In order to further verify the performance of each network block in this paper to the whole detection performance, the ablation study is carried out in this paper to show the effectiveness of proposed each network block. Here, the Resnet50 is taken as the baseline network. On the basis, combined with the spatial attention block and MCF block, the different network configuration schemes are built for the ablation study, including Baseline, Baseline+Spatial Attention, Baseline+MCF Block and Proposed method. Based on the NEU metal surface defect data set, the special experimental results are shown in Table 2. Figure 6 shows the confusion matrix for the ablation experiment.

Table 2. Experiment results of ablation study on NEU data set.

Method	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	F_1 score
Baseline	0.8883	0.8990	0.8862	0.8925
Baseline+Spatial Attention	0.9162	0.9184	0.9182	0.9183
Baseline+MCF Block	0.9353	0.9415	0.9366	0.9390
Proposed	0.9709	0.9738	0.9715	0.9727

As shown in Table 2, compared with the baseline network, due to higher evaluation indicators, it can be clearly seen that the spatial attention block and MCF block both have a positive impact on the performance improvement of the proposed defect classification network, which could prove the effectiveness of proposed each network block. Fused with the spatial attention block and MCF block to the baseline network at the same time, all evaluation indicators reach the best, which also shows the superiority of the proposed defect classification network.

Meanwhile, combined with the confusion matrix in Fig. 6, compared with other network configuration schemes, it could be seen that the proposed defect classification network exists few misclassification cases.

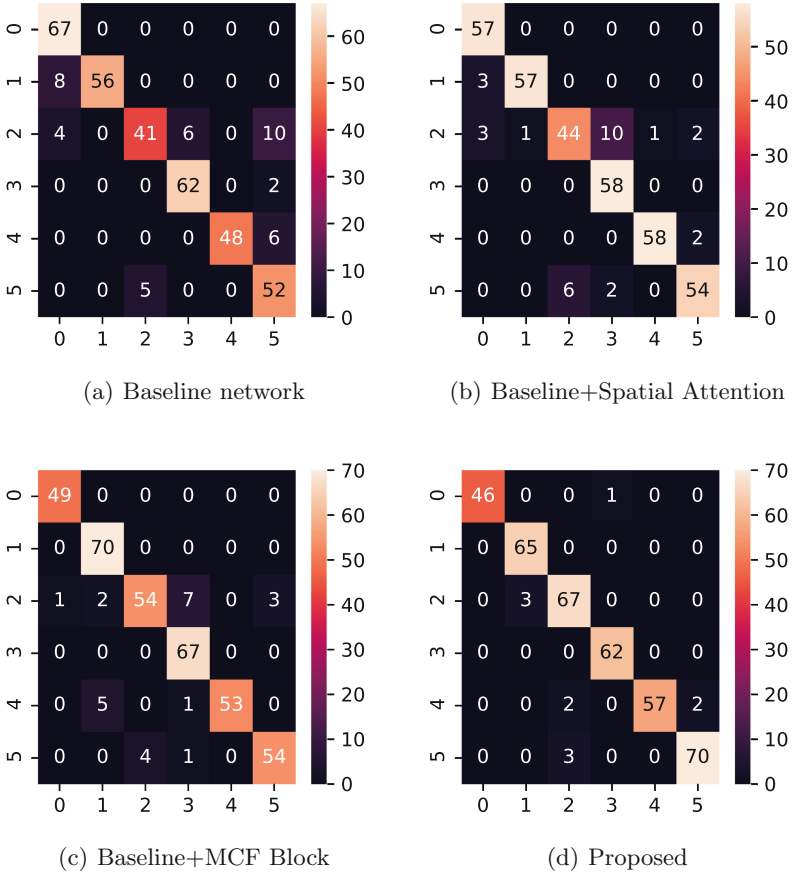


Fig. 6. Confusion matrix for the ablation experiment.

4.3 Performance Comparison

In order to further illustrate the advantages of proposed defect classification network, several typical defect detection networks are selected for performance comparison, including ResNet [6], VGG16 [10], AlexNet [8]. Based on the NEU metal surface defect data set, Table 3 shows the special experimental results of different detection models, and Fig. 7 also gives the confusion matrix on different networks.

As shown in Table 3, compared with other advance detection models, the proposed defect classification network also could acquire the highest detection precision among all the comparison models. Based on the confusion matrix in Fig. 7, the proposed method also has less misclassification defects. Combined with the above results of quantitative analysis, it could further prove the superiority of proposed network.

Table 3. Experiment results of different networks on NEU data set.

Method	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	F_1 score
Resnet50	0.9018	0.9125	0.9009	0.9067
AlexNet	0.9028	0.9144	0.9017	0.9080
VGG16	0.8333	0.8593	0.8139	0.8360
Proposed	0.9709	0.9738	0.9715	0.9727

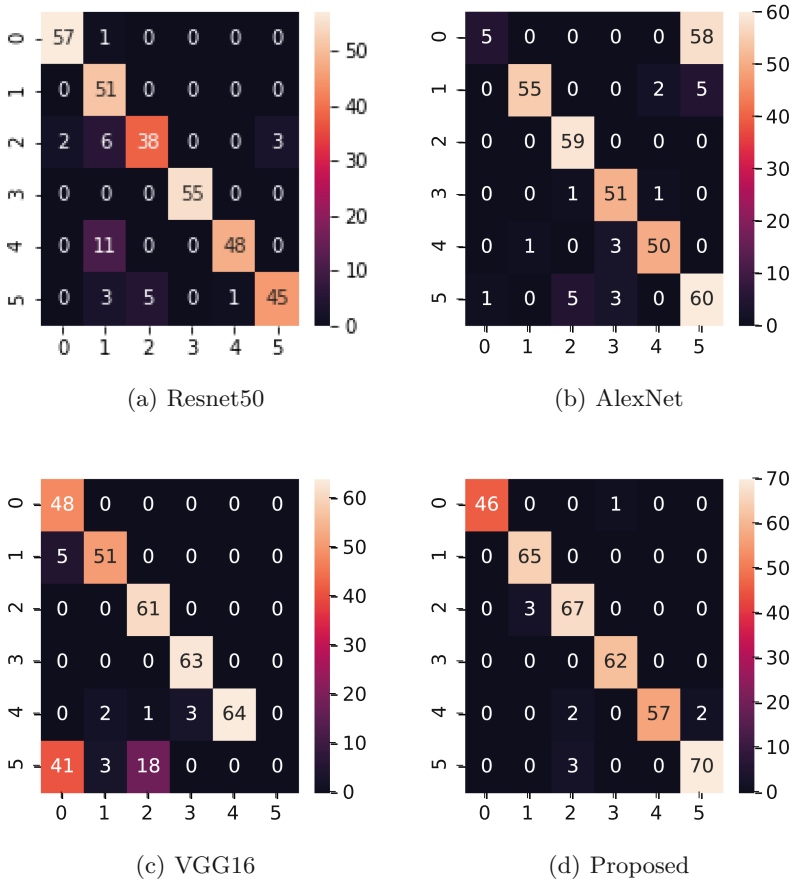


Fig. 7. The confusion matrix for the comparative experiment

5 Conclusion

Faced with the defect detection task against poor contrast and weak texture, a deep defect classification network is proposed in this paper for accurate and automatic defect detection. Combined with the public NEU metal surface defect data set, the proposed defect classification network could acquire a superior

detection performance through the ablation study and comparative experiment. The main work of this paper is drawn as follows.

- (1) An end-to-end deep classification network is proposed for accurate and automatic defect detection.
- (2) A residual attention network is proposed to act as the backbone network for effective feature representation.
- (3) A MCF block is proposed for effective multi-scale feature extraction from local feature maps.

In the future, we will continue to this research work and propose a defect detection network with higher detection precision and efficiency.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 62003309), the National Key Research & Development Project of China (2020YFB1313701) and Outstanding Foreign Scientist Support Project in Henan Province of China (No. GZS2019008).

References

1. Chen, F.C., Jahanshahi, M.R.: NB-CNN: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion. *IEEE Trans. Industr. Electron.* **65**(5), 4392–4400 (2017)
2. Gao, Y., Gao, L., Li, X.: A generative adversarial network based deep learning method for low-quality defect image reconstruction and recognition. *IEEE Trans. Industr. Inf.* **17**(5), 3231–3240 (2020)
3. Gao, Y., Gao, L., Li, X., Wang, X.V.: A multilevel information fusion-based deep learning method for vision-based defect recognition. *IEEE Trans. Instrum. Meas.* **69**(7), 3980–3991 (2019)
4. Geng, Z., Shi, C., Han, Y.: Intelligent small sample defect detection of water walls in power plants using novel deep learning integrating deep convolutional gan. In: *IEEE Transactions on Industrial Informatics* (2022)
5. Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., Fan, C.: Sa-unet: Spatial attention u-net for retinal vessel segmentation. In: *Proceedings of 25th International Conference on Pattern Recognition (ICPR)*, pp. 1236–1242. IEEE (2021)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141 (2018)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, vol. 25 (2012)
9. Qu, Z., Cao, C., Liu, L., Zhou, D.Y.: A deeply supervised convolutional neural network for pavement crack detection with multiscale feature fusion. In: *IEEE Transactions on Neural Networks and Learning Systems* (2021)
10. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint. arXiv:1409.1556* (2014)

11. Tao, X., Zhang, D., Hou, W., Ma, W., Xu, D.: Industrial weak scratches inspection based on multifeature fusion network. *IEEE Trans. Instrum. Meas.* **70**, 1–14 (2020)
12. Yang, J., Fu, G., Zhu, W., Cao, Y., Cao, Y., Yang, M.Y.: A deep learning-based surface defect inspection system using multiscale and channel-compressed features. *IEEE Trans. Instrum. Meas.* **69**(10), 8032–8042 (2020)
13. Yang, L., Fan, J., Huo, B., Li, E., Liu, Y.: A nondestructive automatic defect detection method with pixelwise segmentation. *Knowl.-Based Syst.* **242**, 108338 (2022)
14. Yang, L., Fan, J., Huo, B., Liu, Y.: Inspection of welding defect based on multi-feature fusion and a convolutional network. *J. Nondestr. Eval.* **40**(4), 1–11 (2021). <https://doi.org/10.1007/s10921-021-00823-4>
15. Yang, L., Fan, J., Liu, Y., Li, E., Peng, J., Liang, Z.: Automatic detection and location of weld beads with deep convolutional neural networks. *IEEE Trans. Instrum. Meas.* **70**, 1–12 (2020)
16. Yang, L., Gao, G., Wu, M., Li, J.: Automatic defect recognition method of aluminium profile surface defects. In: 2021 7th International Conference on Robotics and Artificial Intelligence, pp. 21–25 (2021)
17. Yang, L., Song, S., Fan, J., Huo, B., Li, E., Liu, Y.: An automatic deep segmentation network for pixel-level welding defect detection. In: *IEEE Transactions on Instrumentation and Measurement* (2021)
18. Yang, L., Wang, H., Huo, B., Li, F., Liu, Y.: An automatic welding defect location algorithm based on deep learning. *NDT & E Int.* **120**, 102435 (2021)
19. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint. [arXiv:1511.07122](https://arxiv.org/abs/1511.07122) (2015)