



Surface defect detection of voltage-dependent resistors using convolutional neural networks

Tiejun Yang¹ · Shan Peng¹ · Lin Huang¹ 

Received: 5 October 2018 / Revised: 3 September 2019 / Accepted: 1 November 2019 /

Published online: 16 December 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Surface defect detection is an important way to improve the production quality of voltage-dependent resistors (VDRs). To improve the accuracy and efficiency of VDR surface quality detection, an end-to-end surface quality detection method based on deep convolutional neural networks (CNNs) was proposed. The method includes four stages: data preparation, convolution neural network design, CNN training, and testing. First, images of VDRs were acquired from three perspectives, i.e., the front, back, and side, and then training, validation and testing sets were obtained. Second, the proposed CNN models for VDR surface defect detection were constructed. Third, during the training stage, the images with class labels from the established training sets were input to the proposed network for training and validation. Finally, in the testing stage, test images from a total of 408 samples of two VDR models were used to test the trained network. The sensitivity, specificity, accuracy, precision and F measure of the proposed algorithm were compared with those of state-of-the-art methods, and the experimental results showed that the proposed method has a high recognition speed and accuracy and meets the requirements of online real-time detection.

Keywords Surface defect detection · Convolutional neural networks · Voltage-dependent resistors · Image recognition

1 Introduction

A voltage-dependent resistor (VDR) is an electronic device that has nonlinear volt-ampere characteristics (Fig. 1) and is mainly used in voltage clamps in a circuit to absorb the excess current in the case of overvoltage to protect sensitive devices. As the surface quality of VDRs affects their performance, it is necessary to identify any surface defects. For surface detection,

✉ Lin Huang
hlcucu@qq.com

¹ Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin University of Technology, Guangxi 541004, People's Republic of China

machine vision has the characteristics of fast detection, high precision, low noise, strong anti-electromagnetic interference capability, convenience and flexibility, and it is gradually replacing manual detection. However, the accurate detection of minor defects with machine vision systems still faces certain difficulties, as various factors that interfere with the product image acquisition cause the acquired image to not completely and truly reflect the original VDR. The surface defects of VDRs are diverse, e.g., poorly wrapped or overwrapped pins, illegible surface printed information, unacceptable overall dimensions, or various defects on the package surface (such as irregularities, scratches, blemishes, and voids) (Fig. 1). The variable defects make the detection task more challenging, and manual detection may not achieve the desired outcome. Therefore, the computer-aided surface defect detection (SDD) of VDRs is important because it can improve the accuracy and efficiency of the detection.

In recent years, machine vision technology has been widely used in the SDD of electronic components [1, 8, 17]. SDD can be classified into unsupervised SDD and supervised SDD. In unsupervised SDD, the target features are first designed manually, extracted and then classified using classifiers. For example, Lin [14] used the single-level Haar wavelet transform to decompose chip images and extract four wavelet features, based on which ripple defects were identified using a multivariate statistical method. Xi et al. [22] proposed an SDD method of steel billets based on computer vision in which the normal texture was filtered using an isotropic differential filter, features such as the shape factor and ratio of the principal moments were extracted, and then a simple linear regression classifier was applied. Yu et al. [23] presented a coarse-to-fine model to identify defects on rails. Phase-only Fourier transform is used to extract defect regions, and background subtraction is employed to refine the shape of each defect. In these methods, the feature extraction mainly relied on manual design, and no prior knowledge was learned. Due to the small number of manually extracted features, they may only be applicable to specific small data sets and cannot solve more complex SDD problems.

Supervised methods, such as machine learning using prior knowledge, have also been employed to solve the defect detection problem. In this case, a classifier is trained based on

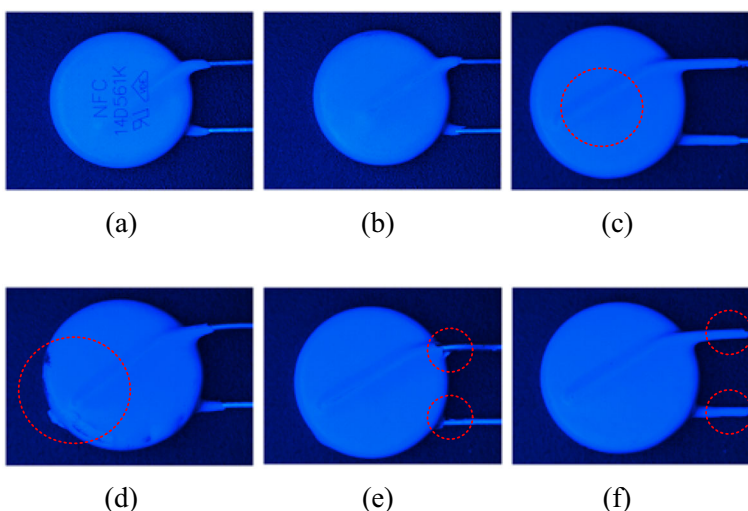


Fig. 1 VDR examples. (a) Normal VDR (front); (b) Normal VDR (back); (c) Missing surface information; (d) Surface damage; (e) Poorly wrapped pins; (f) Excessively wrapped pins

labeled training samples before it is used in the classification. The supervised methods can be further categorized into the manual feature extraction category and the automatic feature extraction category according to the method used to extract the features. In the supervised method with manual feature extraction, the features are manually extracted before the feature learning and training are performed. Shen et al. [18] first divided a bearing image into different regions of interest (ROIs) and further segmented them into candidate defect areas. Then, the features, including the size, location and contrast of each area were extracted manually, and finally, a support vector machine (SVM) classifier was used for training and classification. Kuo et al. [11] used the K-means clustering method to distinguish different features of each part of the chip; ultimately, they classified the features of each part using an effective two-step back-propagation neural network. Liu et al. [16] proposed a classification method for fabric defect images based on an extreme learning machine (ELM), in which geometric and texture features were extracted before the ELM training. Wang et al. [21] presented a method for defect recognition on steel surfaces that used a histogram of oriented gradients (HOG) feature set and a gray-level cooccurrence matrix (GLCM) feature set to train a random forest for defect classification. Since these methods still use manual feature extraction, the recognition results are strongly dependent on the effectiveness of the extracted features.

In recent years, convolutional neural networks (CNNs) [5, 9, 13] have attracted a great deal of attention due to their automatic feature learning and end-to-end high-performance classification capabilities. CNNs were first applied to handwritten character recognition [12] and subsequently extended to other applications, such as object recognition, face detection, image classification and speech recognition. Unlike the traditional recognition methods, CNNs can use large amounts of training data to automatically learn the implicit effective features and achieve end-to-end classification in one network with parallel acceleration through a GPU.

CNNs have also been used in SDD. Cha et al. [3] developed a CNN-based detection method of concrete cracks that was capable of automatically extracting features from a training set of concrete images. Tao et al. [20] proposed a multitask convolutional neural network for detecting the wire of spring-wire socket defects in which the VGG-16 [19] pretrained model is used to initialize the convolutional layers. Chen et al. [4] designed a multispectral CNN network model for solar cell surface defect inspection in which the three spectra in the original image were separated and sent to different CNNs. Using CNNs, it is unnecessary to run an individual feature extraction algorithm for each classification analysis of the data, and high accuracy is achieved in the defect detection. Furthermore, when applying SDD to a specific target, the design of a CNN network structure suitable for the SDD is the key.

To improve the accuracy of the SDD of the VDR, we propose a detection method based on two improved CNN-based network models. One is an eight-layer CNN, which is designed based on the VGG-16 [19] model, called VDR-8-LRN. The network structure of the proposed VDR-8-LRN, containing 5 convolutional layers, 3 full connection layers and an LRN layer, is simpler than that of the VGG-16 while maintaining a high detection accuracy. The other, VDR-FCN, is further improved based on VDR-8-LRN to decrease the number of network parameters required to obtain high efficiency. First, to comprehensively detect VDR surface defects, images of VDRs from three perspectives—the front, back, and side, as shown in Fig. 3—were acquired using an image acquisition device with a coaxial light source. Next, the proposed CNNs were used for training and validation. Finally, the trained CNNs were used for testing.

The remainder of this paper is organized as follows: In Section 2, the method of the VDR image acquisition is described. In Section 3, the proposed method is described in detail. In

Section 4, the experimental results and discussion are presented, and conclusions are provided in the last section.

2 Image acquisition

To obtain high-quality VDR images, the 300,000-pixel MindVision industrial camera was used, and images were acquired at a resolution of 640×480 using a continuous zoom.

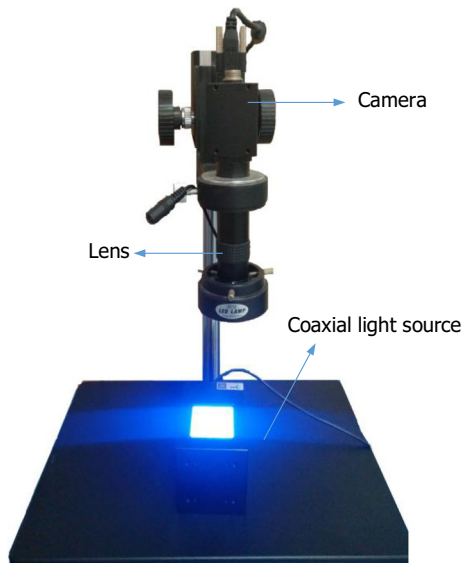
lens with an optical magnification of 0.13–2. Because the smooth surface of the VDR is prone to exhibit reflection under an ordinary light source, which affects the detection outcome, a coaxial LED light source was used to eliminate the reflection, and overexposure and underexposure were avoided by adjusting the illumination intensity of the light source to ensure the VDR image quality. The image acquisition device is shown in Fig. 2.

Two VDR models, i.e., R14 (body diameter: 14 mm) and R10 (body diameter: 10 mm), were used in the experiment. To generate the data sets, 340 VDR samples were randomly selected from each of the two models, of which 240 samples were normal (positive samples) and 100 were defective (negative samples). A total of 680 VDR samples were used in this study. To.

comprehensively detect the defects of the VDR, three pictures were taken on each VDR sample from the perspectives of the front, back and side (as shown in Fig. 3), which generated a total of 2040 experimental sample images.

The data in Fig. 3 show that the VDR images taken under natural lighting conditions had local reflections as well as shadows in the background; however, under the coaxial light source that was used by our image acquisition device, the lighting was uniform, without local reflections, and the background was distinct and noiseless. Overall, the VDR images obtained by our.

Fig. 2 Image acquisition device



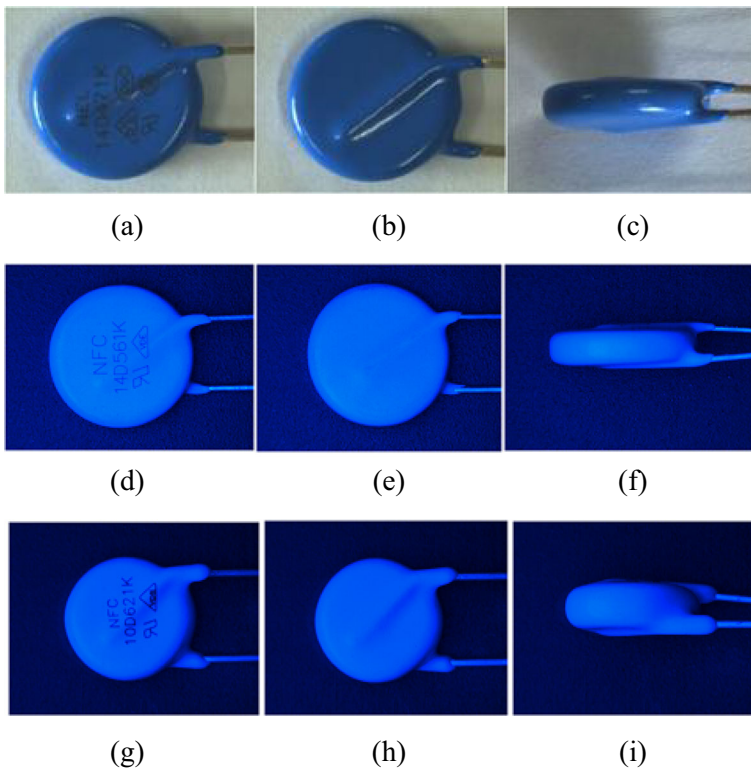


Fig. 3 Examples of VDR image acquisition. Under natural light, the (a) front, (b) back, and (c) side views of the sample; under the coaxial light source, the (d) front, (e) back, and (f) side views of sample R14 and the (g) front, (h) back, and (i) side views of sample R10

acquisition equipment were of good quality, and thus the subsequent experiments on these images were performed without any preprocessing.

3 Method

The deep CNN-based VDR defect detection method proposed in this study includes the following four key steps: data preparation, CNN model design, CNN training, and testing. The specific process of the proposed method (Fig. 4) is described as follows:

- 1) Data preparation. After the VDR images are acquired, the images are manually labeled by experienced professional quality inspectors. Then, the images are randomly divided into a training set, a validation set and a testing set at a number ratio of 7:1:2.
- 2) Design of the CNN model for VDR defect detection. VGG-16 is designed for natural image classification (2000 classes), such as in the ImageNet Challenge, and thus contains many network layers. The target in this study is a VDR, which is simple in shape and does not have complex features such as color or texture, making it a binary problem (normal or defective). Therefore, in this step, we design an appropriate CNN network model specifically for the SDD of VDRs. Given the high multiclassification accuracy of VGG-16, we

simplify its network layer and determine what type of network structure is most suitable for the SDD of VDRs under the premise of ensuring detection accuracy.

- 3) CNN training. Next, training is performed with the prepared training set and validation set based on the proposed CNN model. The process of CNN training includes network initialization, feature learning and classification, and adjustment of network parameters based on stochastic gradient descent (SGD) [2] in an iterative manner until reaching the maximum number of iterations.
- 4) Testing using trained CNNs. Finally, testing is performed on the trained CNNs, and the test result is obtained.

3.1 VGG-based CNN model design

The VGG-16 [19] network model has achieved good results in the image detection of complex targets, exhibiting top-5 classification error rates on the test and validation sets of the ILSVRC-2012 data set at 7.4% and 7.5%, respectively. The model contains 13 convolutional layers, 5 maximum value pooling layers, and 3 full connection (FC) layers. Given that the shape.

“Conv” represents convolutional layer, and “Pool” represents maxpooling layer.

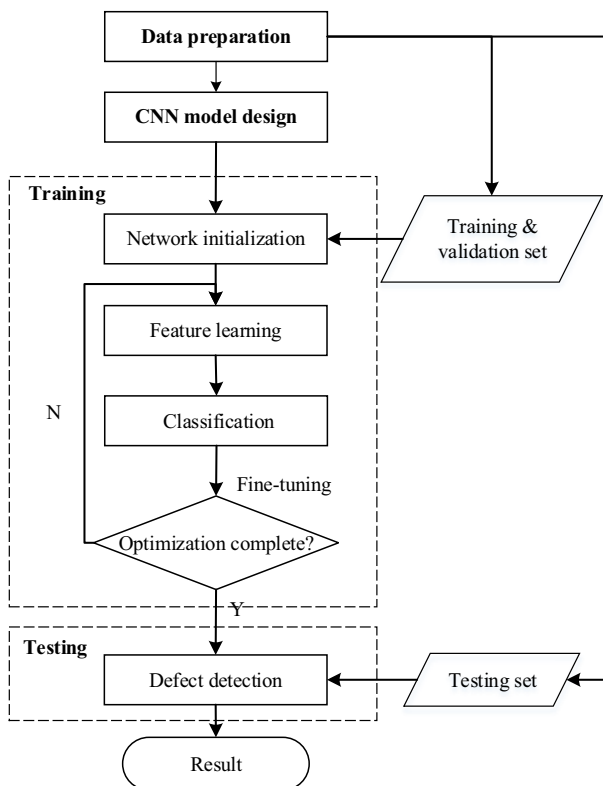


Fig. 4 Overall flow chart of our method, which includes four key steps: data preparation, CNN model design, CNN training, and testing

Table 1 CNN models with different numbers of convolutional layers

Network type	Number of convolutional layers	Convolutional and pooling layers
VGG-16	13	2Conv + Pool +2Conv + Pool +3Conv + Pool +3Conv + Pool +3Conv + Pool
VGG-15	12	2Conv + Pool +2Conv + Pool +2Conv + Pool +3Conv + Pool +3Conv + Pool
VGG-13	10	2Conv + Pool +2Conv + Pool +2Conv + Pool +2Conv + Pool +2Conv + Pool
VGG-11	8	1Conv + Pool +1Conv + Pool +2Conv + Pool +2Conv + Pool +2Conv + Pool
VGG-9	6	1Conv + Pool +1Conv + Pool +1Conv + Pool +1Conv + Pool +2Conv + Pool
VGG-8	5	1Conv + Pool +1Conv + Pool +1Conv + Pool +1Conv + Pool +1Conv + Pool
VDR-8-LRN	5	1Conv + Pool + LRN + 1Conv + Pool +1Conv + Pool +1Conv + Pool +1Conv + Pool

of the VDR is relatively simple and does not have any complex color or texture features, the direct application of a complex model such as VGG-16 for its detection is wasteful. Thus, based on the VGG-16 network model, we kept the pretraining network parameters unchanged on the ILSVRC-2012 data set, and we attempted to improve the detection efficiency by reducing the number of feature extraction convolutional layers and thus the complexity of the network model under the premise of maintaining a high recognition accuracy.

In the process of reducing the convolutional layers, we refer to the VGG network structure design method, in which several networks with 8, 10, 13 or 16 convolutional layers were tested [19]. In the task of VDR defect detection, VGG-16 (with 13 convolutional layers) and similar networks with 12, 10, 8, 6, or 5 convolutional layers (as shown in Table 1, named VGG-15, 13, 11, 9, 8, respectively) were tested. First, the same training samples were used to train these networks, and then some of the test samples were used for testing. Table 2 shows the accuracies of the tested networks and the average time each network took to iterate 100 times. As the number of convolutional layers was reduced from 13 to 6, the accuracy remained constant at 1, while the average computation time decreased continuously; when the number of convolutional layers was reduced further to 5 (i.e., VGG-8), the accuracy was slightly reduced, but the reduction in the computation time was not significant. As the number of

To further improve the generalization ability of the VGG-8 model, a local response normalization (LRN) layer [10] was added to the network. LRN is a normalization technique for improving the accuracy in deep learning training that is performed after the activation and pooling layers to

Table 2 Test results of the accuracies and computation times of the CNN models with different numbers of convolutional layers convolutional layers was further reduced, the detection performance also decreased.

Network type	Number of convolutional layers	Accuracy (%)	Time (s)
VGG-16	13	100	36
VGG-15	12	100	16.5
VGG-13	10	100	11
VGG-11	8	100	11
VGG-9	6	100	9
VGG-8	5	99.51	8.5

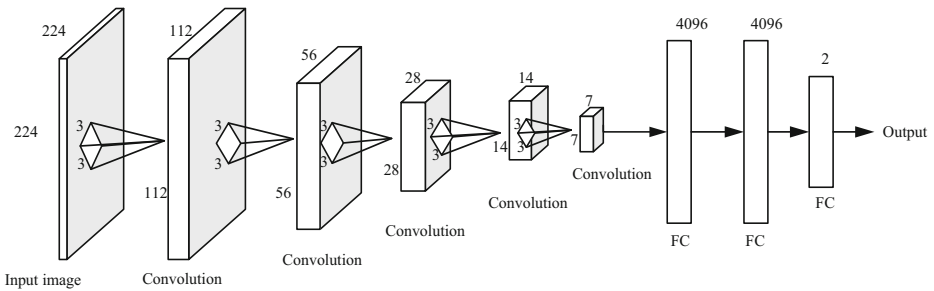


Fig. 5 VDR-8-LRN network model. The network contains 8 layers: 5 convolutional layers and 3 full connection (FC) layers. Each convolutional layer is followed by a maximum pooling layer, and the LRN layer is added after the first pooling layer. The size of the convolution kernel of each convolution layer is 3×3 , with a padding of 1 and a stride of 1, whereas that of the maximum pooling layer is 2×2 , with a stride of 2. The numbers of neurons in the FC layers are 4096, 4096, and 2, in that order

enhance the generalization ability of the network. Under the premise of not changing the original network parameters, the LRN layer was added after the pooling layer of the first layer to generate the improved network model, named VDR-8-LRN (as shown in Table 1). Based on the above analyses, we ultimately chose the VDR-8-LRN model as the defect detection model for the VDRs.

The VDR-8-LRN network model and its parameter settings are described in detail in Fig. 5. The network structure of VDR-8-LRN contains convolutional layers, pooling layers, an LRN layer, and FC layers. The input layer contains images with a size of 224×224 pixels, followed by alternating convolutional layers and maximum pooling layers, with the addition of the LRN layer after the first pooling layer. The size of the feature map after the convolutional layer is identical to that of the previous layer, while that after the pooling layer becomes one-half of its original size. After the fifth pooling layer comes the FC layers, whose output is sent to a 2-way softmax classifier for classification, from which the final result is output. Note that the.

AlexNet network [10], which is similar to the network that we proposed in this study, also has an 8-layer network structure, but the differences are that in the AlexNet network, LRN layers are added after the first and second activation layers and the numbers of pooling layers are decreased for the third and fourth convolutional layers.

3.2 Design of a more efficient CNN model

With five or more convolution layers, the improved CNN model based on VGG-16 performed well in identifying VDR appearance defects. However, it uses three fully connected layers as the classifier, resulting in a high number of network parameters. Table 3 shows a comparison of the number of parameters of VGG-16 and that of the improved VDR-8-LRN. VDR-8-LRN has a similar network structure, but fewer convolutional layers, than VGG-16, and thus uses approximately 27% of the parameters of the convolutional layer of VGG-16. Because both use three FC layers as the classifier and have a large number of parameters in the FC layer, the total number of

Table 3 Comparison of the numbers of parameters in the network models

CNN model	Total number	Number of convolutional layers	Number of classifiers
VGG-16	1.38e8	1.47e7	1.19e8
VDR-8-LRN	1.23e8	3.91e6	1.19e8
VDR-FCN	2.97e6	2.97e6	1026

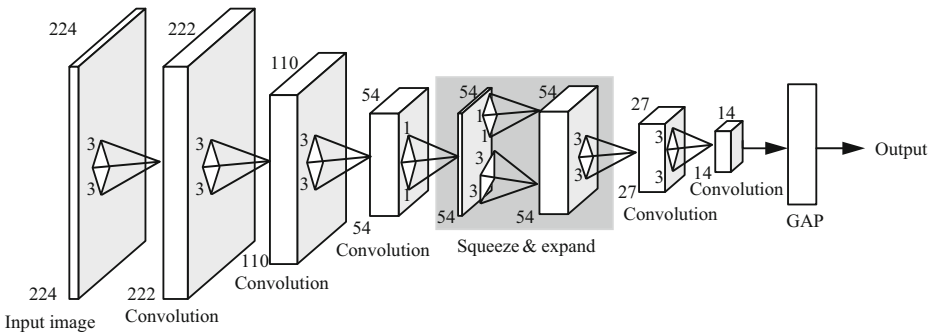


Fig. 6 VDR-FCN network model. The network contains 7 layers: 5 convolutional layers, a squeeze and expand (SE) layer and a GAP layer. The third, fourth and fifth convolutional layers are followed by a maximum pooling layer. The SE layer is added after the first pooling layer and the GAP layer is added after the last pooling layer

parameters of VDR-8-LRN is not reduced significantly (it is approximately 89% of that of VGG-16).

To further reduce the number of parameters of the CNN model needed to improve recognition efficiency, we adopted the global average pooling (GAP) method [15] to replace the FC layer classifier in VDR-8-LRN. Furthermore, we improved the model to enable the continued use of large-sized feature maps as inputs to the GAP classifier. Since the improved network is mainly composed of convolutional layers, we named it the full convolutional neural networks based SDD of VDR (VDR-FCN). Its network structure is shown in Fig. 6.

First, VDR-FCN retains the fourth and fifth convolutional layers and the pooling layer of VDR-8-LRN and removes the pooling layers after the first and second convolutional layers. Second, the pooling layer after the third convolutional layer is retained, and a squeeze and expand (SE) layer [6] was added after it to further optimize the feature maps outputted by the preceding convolutional layers. Lastly, the original FC layer was removed and replaced by the GAP layer. The size of the input classifier's feature map of VDR-FCN is four times that of VDR-8-LRN. The parameter settings of VDR-FCN are described in detail in Table 4. As shown in Tables 3 and 4, the number of parameters of VDR-FCN is only 2.4% that of VDR-8-LRN.

3.3 Training and testing

After designing the CNN models, the next step is the CNN training. The proposed models were trained based on the VGG-16 pretraining model and the ILSVRC-2012 data set. Before the training, to be consistent with the pretraining network, the spatial resolutions of all of the training sample images were adjusted to 224×224 pixels. During the training process, the grayscale mean of the training set sample was subtracted from the input 224×224 RGB images, and then they were sent to the convolutional layer for processing.

The R14 and R10 training sets were each used to train the proposed models. In the training process, since the number of training samples was not very large, the batch size was set to 8, the learning rate to 10^{-4} , the dropout to 0.5 at all dropout layers, the initial momentum to 0.9 with a weight decay of 0.0005, and the maximum number of iterations to 10,000. The stochastic gradient descent method was used in the weight adjustment.

In the testing stage, the two trained VDR-8-LRN and VDR-FCN models were used to test the test sets of their corresponding VDR models, and the final defect detection results were obtained.

Table 4 Parameter settings for the VDR-FCN network model

Layer name/type	Output size	Filter size/stride	Number of params
input image	224*224*3		
conv1	222*222*16	3*3/1	448
conv2	110*110*24	3*3/2	3480
conv3	108*108*32	3*3/1	6944
maxpool1	54*54*32	3*3/2	
SE	squeeze 54*54*16	1*1/1	528
	expand1 54*54*64	1*1/1	1088
	expand2 54*54*64	3*3/1	9280
	concat 54*54*128		
conv4	54*54*512	3*3/1	590,336
maxpool2	27*27*512	2*2/2	
conv5	27*27*512	3*3/1	2,359,808
maxpool3	14*14*512	2*2/2	
GAP	conv 16*16*2	1*1/1	1026
	avepool 1*1*2		

4 Experiments and discussion

The experimental environment was as follows: The Ubuntu16.04 operating system, NVIDIA GPU 1070Ti, and DDR2 6 GB memory. The proposed algorithm and the ones evaluated in this study for comparison were run according to Caffe 1.0 [7]. To verify the effectiveness of the proposed models (VDR-8-LRN and VDR-FCN) in VDR defect detection, the detection accuracy of the proposed models were compared using SVM with HOG (HOG+SVM) [21], VGG-16, VGG-8, and AlexNet. The evaluation indicators of the sensitivity, specificity, accuracy, precision and F measure were used to compare the detection accuracy between the algorithms. In addition, the “training error-number of iterations” curves of VGG-8 and AlexNet were compared with that of the proposed models to evaluate the training error convergence of each model.

4.1 Dataset

The two VDR models, i.e., R14 and R10, were tested in separate experiments, in which a total of 340 samples were tested for each VDR model; from each sample, images from the perspectives of the front, back and side of the VDR were acquired, generating a total of 1020 images for each VDR model. The ground truths of the VDRs are obtained by experienced professional quality inspectors. The acquired samples were randomly divided into a training set, a validation set, and a test set according to the ratio of 7:1:2, which generated 714 training samples, 102 validation samples, and 204 test samples for each of the two VDR models, in which both defective samples (positive samples) and non-defective samples (negative samples) were included, as shown in Table 5.

4.2 Evaluation indicators

To compare the accuracy of the proposed method with that of the other models, five evaluation indicators, i.e., the sensitivity (*SE*), specificity (*SP*), accuracy (*ACC*), precision (*PR*), and F measure (*F*), were used, which are defined as follows:

Table 5 Acquired samples and their divisions

VDR model	Number of training sets		Number of validation sets		Number of test sets	
	N	P	N	P	N	P
R14	504	210	72	30	144	60
R10	504	210	72	30	144	60

P: positive sample; N: negative sample.

$$SE = TP/(TP + FN), \quad (1)$$

$$SP = TN/(FP + TN), \quad (2)$$

$$ACC = (TP + TN)/(TP + FP + TN + FN), \quad (3)$$

$$PR = TP/(TP + FP), \quad (4)$$

$$F = 2(PR \cdot SE)/(PR + SE), \quad (5)$$

where TP represents the true positives, TN represents the true negatives, FP represents the false positives, and FN represents the false negatives.

4.3 Experimental results

During the experiment, the proposed methods were compared with VGG-16, VGG-8 and AlexNet. HOG+SVM is a commonly used supervised method based on manual feature extraction, which first extracts the HOG feature set and then uses SVM for training and testing. The training of VGG-16, VGG-8 and AlexNet uses the traditional method of putting together two types of training samples for training. Table 6 shows the defect detection results

Table 6 VDR defect detection results

Network model	VDR model	SE (%)	SP (%)	ACC (%)	PR (%)	F (%)
VGG-16	R10	88.33	100	96.57	100	93.80
	R14	71.67	100	91.67	100	83.50
VGG-8	R10	90.00	100	97.06	100	94.74
	R14	100	90.97	93.63	82.19	90.22
AlexNet	R10	100	97.92	98.53	95.24	97.56
	R14	68.33	93.06	85.78	80.39	73.87
HOG+SVM	R10	100	95.14	96.57	89.55	94.49
	R14	96.67	100	99.02	100	98.31
VDR-8-LRN	R10	100	100	100	100	100
	R14	100	100	100	100	100
VDR-FCN	R10	100	99.31	99.51	98.36	99.17
	R14	100	100	100	100	100

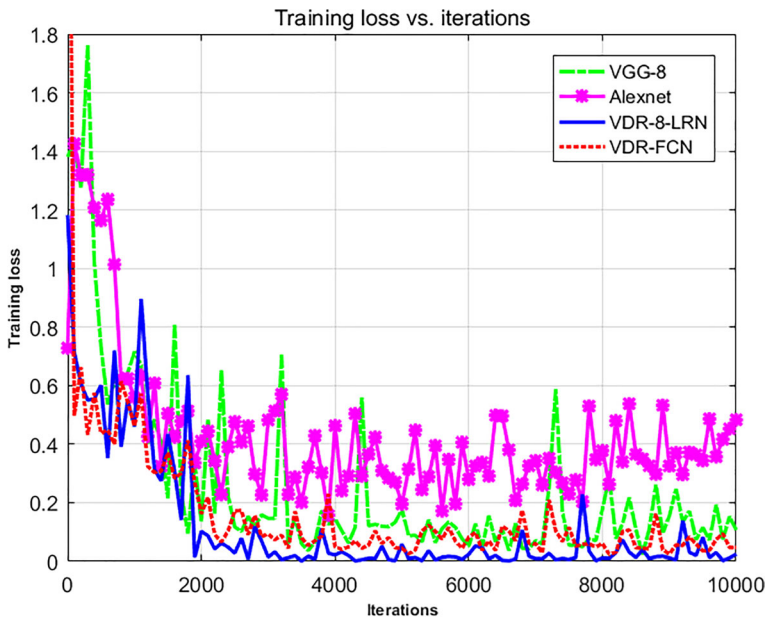


Fig. 7 Relationship between the training loss value and the number of iterations in the four models

on the two VDR models. In the testing stage, 204 test samples were used for each VDR model. When detecting on the R10 model, AlexNet ranked third in the *ACC* (98.53%) and *F* (97.56%), with only false positives a factor, and the *SE*, *SP* and *PR* were 100%, 97.92% and 95.24%, respectively. The *ACC* of VGG-8 was very close to that of AlexNet, 97.06%, with only false negative detection a factor. HOG+SVM and VGG-16 performed the worst, both with an *ACC* of 96.57%.

When detecting with the R14 model, HOG+SVM ranked third in the *ACC* indicator (99.02%), but manual feature extraction was required. VGG-8 ranked fourth in the *ACC* indicator (93.63%). Though VGG-16 had a more complex model structure than VGG-8, the *ACC* of VGG-16 was slightly lower than that of VGG-8, which shows that the more complex the network, the VDR classification is not necessarily more accurate. The AlexNet network showed more false positives and false negatives in the R14 model detection and it performed the worst in the *ACC* (85.78%) and *F* (73.87%).

In comparison, all defects were accurately detected by the proposed VDR-8-LRN on both the R10 and R14 models. It took an average of approximately 83 ms for the VDR-8-LRN model to analyze each sample image. VDR-FCN produced only one FP sample, and its *ACC* indicator was similar to that of VDR-8-LRN. It was faster than VDR-8-LRN, taking an average time of 5 ms to inspect one sample. They both meet the needs of real-time detection. To summarize, the VDR-8-LRN and VDR-FCN proposed in this study had the best performance in detecting the defective VDR samples.

Furthermore, from a training perspective, we compared AlexNet, VGG-8, VDR-8-LRN and VDR-FCN. Given that the relationship between the model training loss and the number of

iterations largely reflects the robustness of the model, we examined the relationship in each of the four models (as shown in Fig. 7). The results show that the training loss

values of the four models were small in the very beginning. The training loss value of the VDR-8-LRN model rose sharply at the beginning, peaked at approximately 1.2, then dropped quickly in a short time, with only a small fluctuation, and rapidly reached a stable value, which was very small, ultimately reaching nearly zero. VDR-FCN also quickly converged to a small value (at approximately 0.04). In contrast, the training loss values of both AlexNet and VGG-8 first declined and then rose, with a maximum value less than 0.9 and a rather large fluctuation. However, AlexNet and VGG-8 were also able to reach stable values quickly and had final loss values close to 0.4 and 0.1, respectively. These results indicate that the proposed VDR-8-LRN and VDR-FCN are more stable.

5 Conclusion

In this study, a CNN model was designed for the SDD of VDRs and achieved high accuracy and efficiency. Using the CNN's advantage of automatic feature learning and the VDR's unique characteristics, we further optimize the network structure of the model on the basis of VGG-16 and constructed a CNN model, i.e., VDR-8-LRN, more suitable for the SDD of VDRs through experimental analyses on networks of different depths. To further improve recognition efficiency, we designed the VDR-FCN, which needs only 2.4% of the number of parameters needed by VDR-8-LRN and took an average time of 5 ms to inspect one sample at a recognition accuracy similar to that of VDR-8-LRN. In the experiment, we compared four different models and assessed the performances of the algorithms from the perspectives of the detection accuracy and training error convergence. These results indicate that VDR-8-LRN and VDR-FCN are effective and stable. In future work, we will attempt to further streamline the model structure while increasing the size of the data set in order to achieve even faster and more accurate defect detection.

Acknowledgements This research was partly supported by the Guangxi Natural Science Foundation (2018GXNSFBA281081), the Guangxi Basic Ability Promotion Project for Young and Middle-aged Teachers (2017KY0247), the Project of Cultivating a Thousand Young and Middle-aged Teachers in Guangxi Universities, the Guangxi Key Laboratory Fund of Embedded Technology and Intelligent System (2018A-07) and the Guangxi Universities Key Laboratory Fund of Embedded Technology and Intelligent Information Processing (2017-1-1, 2017-2-4).

References

1. Asiegbu GO, Haidar AMA, Hawari K (2012) A review of defect detection on electrical components using image processing technology. Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012), pp. 247–255. Springer India India
2. Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In: Lechevallier Y, Saporta G (eds) Proceedings of the 19th international conference on computational statistics (COMPSTAT'2010). Springer Paris, France, pp 177–187 <http://leon.bottou.org/papers/bottou-2010>
3. Cha Y-J, Choi W, Büyüköztürk O (2017) Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil Infrastruct Eng* 32:361–378
4. Chen H, Pang Y, Hu Q, Liu K (2018) Solar cell surface defect inspection based on multispectral convolutional neural network. *J Intell Manuf.* <https://doi.org/10.1007/s10845-10018-11458-z>
5. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. *J Mach Learn Res* 9:249–256

6. Iandola FN, Moskewicz MW, Ashraf K, Han S, Dally WJ, Keutzer K (2017) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. CoRR, abs/1602.07360
7. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J (2014) Caffe: convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM International Conference on Multimedia, Vol. arXiv: 1408.5093, pp. 675–678. Orlando, FL
8. Koch C, Georgieva K, Kasireddy V, Akinci B, Fieguth PW (2015) A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv Eng Inform* 29: 196–210
9. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Proces Syst* 25:1106–1114
10. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*: 1097–1105
11. Kuo C-FJ, Hsu C-TM, Liu Z-X, Wu H-C (2013) Automatic inspection system of LED chip using two-stages back-propagation neural network. *J Intell Manuf* 25:1235–1243
12. LéCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86:2278–2324
13. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
14. Lin H-D (2007) Computer-aided visual inspection of surface defects in ceramic capacitor chips. *J Mater Process Technol* 189:19–25
15. Lin M, Chen Q, Yan S (2014) Network in network. CoRR, abs/1312.4400
16. Liu L, Zhang J, Fu X, Liu L, Huang Q (2018) Unsupervised segmentation and elm for fabric defect image classification. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-11018-16786-11047>
17. Ngan HYT, Pang G, Yung NHC (2011) Automated fabric defect detection - a review. *Image Vis Comput* 29: 442–458
18. Shen H, Li S, Gu D, Chang H (2012) Bearing defect inspection based on machine vision. *Measurement* 45: 719–733
19. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *Comput Sci*
20. Tao X, Wang Z, Zhang Z, Zhang D, Xu D, Gong X, Zhang L (2018) Wire defect recognition of spring-wire socket using multitask convolutional neural networks. *IEEE Trans Compon Packag Manuf Technol* 8:689–698
21. Wang Y, Xia H, Yuan X, Sun B (2018) Distributed defect recognition on steel surfaces using an improved random forest algorithm with optimal multi-feature-set fusion. *Multimed Tools Appl* 77:16741–16770
22. Xi J, Shentu L, Hu J, Li M (2017) Automated surface inspection for steel products using computer vision approach. *Appl Opt* 56:184–192
23. Yu H, Li Q, Tan Y, Gan J, Wang J, Geng Y, Jia L (2019) A coarse-to-fine model for rail surface defect detection. *IEEE Trans Instrum Meas* 68:656–666

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Tiejun Yang is current a professor at the School of Information Science and Engineering, Guilin University of Technology, China. His research interest is image processing and computer vision.



Shan Peng is currently pursuing the Master degree at the School of Information Science and Engineering, Guilin University of Technology, China. Her research interests include computer vision and pattern recognition.



Lin Huang is currently pursuing the Ph.D degree at the Department of Electrical Engineering, Xidian University, China. Her research interests include computer vision and pattern recognition.