Artificial Life and Robotics
https://doi.org/10.1007/s10015-019-00545-x

ORIGINAL ARTICLE

Defect detection method using deep convolutional neural network, support vector machine and template matching techniques

Fusaomi Nagata[1] · Kenta Tokuno[1] · Kazuki Mitarai[1] · Akimasa Otsuka[1] · Takeshi Ikeda[1] · Hiroaki Ochi[1] · Keigo Watanabe[2] · Maki K. Habib[3]

Received: 15 March 2019 / Accepted: 26 June 2019

Abstract

In this paper, a defect detection method using deep convolutional neural network (DCNN), support vector machine (SVM) and template matching techniques is introduced. First, a DCNN for visual inspection is designed and trained using a large number of images to inspect undesirable defects such as crack, burr, protrusion, chipping, spot and fracture phenomena which appear in the manufacturing process of resin molded articles. Then the trained DCNN named sssNet and well-known AlexNet are, respectively, incorporated with two SVMs to classify sample images with high recognition rate into accept as OK category or reject as NG one, in which compressed feature vectors obtained from the DCNNs are used as inputs for the SVMs. The performances of the two types of SVMs with the DCNNs are compared and evaluated through training and classification experiments. Finally, a template matching technique is further proposed to efficiently extract important target areas from original training and test images. This will be able to enhance the reliability and accuracy for defect detection.

Keywords Deep convolutional neural network (DCNN) · Support vector machine (SVM) · Template matching · Defect detection system

1 Introduction

Recently, DNNs, DCNNs and SVMs are gathering attention in various industrial fields due to their high classification abilities. For example, Guo et al. designed a DCNN and evaluated it based on benchmarking datasets MNIST and CIFAR10, in which different methods of learning rate set and different optimization algorithm of solving the optimal parameters were analyzed [1]. Deng et al. proposed an automatic defect verification system called Auto-VRS to decrease the false alarm rate and reduce operator's workload. The system consisted of two subsystems for fast circuit comparison and DNN-based defect classification [2]. As for the research about comparison of DCNN and SVM, for example, Wan et al. proposed a fiber classification method based on DCNN which was designed to classify grayscale fiber images including 7 different kinds of shapes [3]. It was reported that the comparison results indicated that the proposed DCNN approach achieved higher accuracy than a conventional shallow NN with a hidden layer and a SVM in fiber recognition.

As for the research combining DCNN and SVM, for example, Shao et al. presented an effective method based on SVM concept to detect weld defect in X-ray images taken in real time. After all potential defects were segmented by a background subtraction algorithm, three features including defect area, average grayscale difference and grayscale standard deviation were extracted. The extracted features were used as input to the SVM classifier to distinguish nondefects from defects. It was reported that the proposed automatic defect detection method could reduce the undetected rate and false alarm rate effectively [4]. Niu and Suen presented their hybrid model, in which the DCNN and SVM worked as a trainable feature extractor and a classifier,

This work was presented in part at the 24th International Symposium on Artificial Life and Robotics, Beppu, Oita, January 23–25, 2019.

✉ Fusaomi Nagata
nagata@rs.socu.ac.jp

1 Department of Mechanical Engineering, Faculty of Engineering, Sanyo-Onoda City University, 1-1-1 Daigaku-Dori, Sanyo-Onoda 756-0884, Japan

2 Okayama University, Okayama, Japan

3 American University, Cairo, Egypt

respectively. The hybrid model was evaluated using the well-known MNIST handwritten digit dataset [5]. Sun et al. also proposed a novel hybrid model that integrated the synergy of two superior classifiers DCNN and SVM for functional magnetic resonance imaging (fMRI) recognition [6]. It was reported that when the CNN was used as a feature extractor, the SVM classifier was demonstrated to be the best combining counterpart compared with other classifiers such as decision tree, neural network, K-nearest neighbor, random forest, and AdaBoost. In the effort by Chan et al., images were first preprocessed, i.e., denoised by a filtering technique and cropped by an image boundary extraction. Features in the images were extracted using a DCNN named AlexNet [7] and then images were classified using a SVM classifier based on the features [8]. In the field of kinesiology, Wu et.al proposed a CNN–SVM combined model for the pattern recognition of knee motion [9]. It was reported that the CNN–SVM combined model could automatically extract features using the CNN part, and improve the classification accuracy by means of the SVM part.

Besides these, many papers that combine DCNN and SVM have been published up to now, in which the powerful AlexNet trained by the developer is frequently applied to versatile feature extractors for SVMs. The DCNN named AlexNet is originally designed and trained using images consisting of 1,000 kinds of various categorizations for general classification. However, it seems that the combination of a SVM and a DCNN trained using images including actual different kinds of defects and the comparison with another SVM combined with the AlexNet have not been reported yet in detail.

The authors already proposed a user-friendly design and training application for DCNNs and applied it to make three types of DCNNs for classifying images of resin molded articles into two, five and six categorizations [10]. DCNNs generally have several blocks consisting of convolutional, ReLU and pooling layers to accept image files in the former hidden layers, which lead to fully connected layers and a softmax function layer. The design application allows students and novice engineers to design and train DCNNs even if they are not familiar with the software development using C++ or Python. Then the application was extended to be able to design and obtain SVMs by one class unsupervised learning [11]. In this paper, a defect detection method using DCNN, SVM and template matching techniques is proposed for visual inspection. Two types of SVMs are first designed using the application, then they are learned using typical OK images without any defect to be able to distinguish images including defects from all images. It is assumed that the defects are crack, burr, protrusion, chipping, spot and fracture which appear in the manufacturing process of resin molded articles. Figure 1 shows examples of the typical defects. Two types of pretrained DCNNs, i.e., our designed

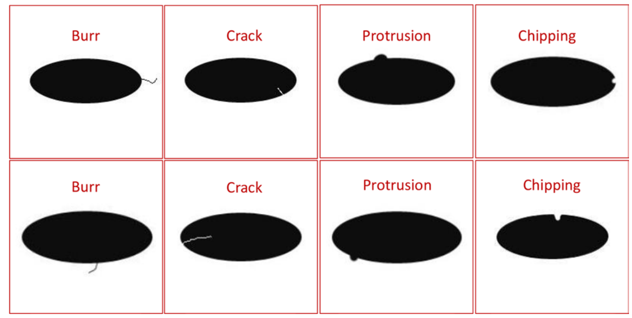

Fig. 1 Examples of four kinds of defects which appear in production process of resin molded articles, whose horizontal length is around 30 mm

sssNet and well-known AlexNet, are severally incorporated into the foreparts of two SVMs as feature extractors, in which compressed feature vectors extracted by the DCNNs are used as input vectors for the SVMs. The performance of the SVMs incorporated with the two types of DCNNs are compared and evaluated through training and classification experiments. Finally, a template matching technique is proposed and applied to the SVM using AlexNet to narrow important target areas from original training and test images. This will be able to enhance the reliability and accuracy for binary classification using the SVM.

2 SVMs incorporated with two kinds of DCNNs

The binary classification using the DCNN alone named sssNet was evaluated in [10], however, in which four drawbacks were recognized. The first one is that the structure and parameters of DCNNs were more complex than those of SVMs. The second one is that images of non-defective and defective articles must have been prepared for training DCNNs. The third one is that much more time was required for the training of DCNNS than SVMs. The last one is that some additional training was needed to grow satisfactory generalization ability to test data.

In this section, two types of SVMs for binary classification are designed using the already developed DCNN & SVM design tool. Actually, the most important function which is required to a defect inspection system is to remove defective products from all products. It is not allowed that any defective product is mixed into lots of non-defective products. To cope with this need, two types of SVMs are tried to be designed and trained using the DCNN & SVM design tool. It is expected that the trained SVMs will be able to classify input images into OK or NG category including crack, burr, protrusion, chipping, spot and fracture.

2.1 SVM incorporated with our designed DCNN named sssNet

As for the first SVM, our designed DCNN named sssNet is used to extract multidimensional feature vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_{32}]^{\mathrm{T}}$ from each inputted image. Figure 2 illustrates the designed binary class SVM whose input is the feature vector $\boldsymbol{x}$, generated from the 1st fully connected layer (11th layer) in the sssNet. Gaussian kernel function is used for one class training of the SVM, in which feature vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{5100} \in \Re^{32\times1}$, extracted from 5100 OK images are only used for unsupervised learning of the SVM with the sssNet. NG images with defects are not needed for training the SVM. Sequential minimal optimization (SMO) algorithm [12] is applied to solve the quadratic programming (QP) of the SVM. It took about several minutes for training the SVM.

If a feature vector $\boldsymbol{x} \in \Re^{32\times1}$ extracted from a test image using the sssNet is given to the trained SVM, the corresponding output value $f(\boldsymbol{x})$ called the score is obtained by:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i \exp\left(-\left\|\frac{\boldsymbol{x}_i^* - \boldsymbol{x}_s}{k}\right\|^2\right) + b, \tag{1}$$

where k and $\boldsymbol{x}_s$ are, respectively, the kernel scale and the standardized input vector calculated by:

$$\boldsymbol{x}_s = (\boldsymbol{x} - \boldsymbol{x}_\mu) \oslash \boldsymbol{x}_\sigma \tag{2}$$

with

$$\boldsymbol{x}_\mu = \frac{\sum_{j=1}^{5100} \boldsymbol{x}_j}{5100} \tag{3}$$

$$\boldsymbol{x}_\sigma = \left[\frac{1}{5100}\sum_{j=1}^{5100}(\boldsymbol{x}_j - \boldsymbol{x}_\mu)^{\circ 2}\right]^{\circ\frac{1}{2}}, \tag{4}$$

where $\oslash, \circ 2, \circ\frac{1}{2}$ are, respectively, the Hadamard operators for element-wise division, power and root; N is the number of support vectors determined in the training process using the training data set of OK category; $\boldsymbol{x}_i^* \in \Re^{1\times32}$ $(i = 1, 2, \ldots, N)$ is the determined support vectors; α_i $(i = 1, 2, \ldots, N)$ and b are the Lagrange multipliers and the bias, respectively, which are the SVM parameters estimated through the training. In this experiment, k, N and b were obtained as 1.1875, 2, 621 and -1.0639 from the training, respectively.

Binary classification can be estimated by checking the sign of $f(\boldsymbol{x})$ such that $f(\boldsymbol{x}) > 0$ and $f(\boldsymbol{x}) < 0$ mean OK category and NG category, respectively.

2.2 SVM incorporated with well-known AlexNet

As for the second SVM, a well-known DCNN called AlexNet is used to extract the feature vector $\boldsymbol{x} \in \Re^{4096\times1}$ from each inputted image. Note that the feature vector has 4,096 elements to cope with one thousand classification task. The AlexNet is one of DCNNs trained using over one million images in ImageNet database [13], which has totally 25 layers including five convolution layers and three fully connected layers. The AlexNet can classify inputted images into 1,000 object categories including keyboards, mouses, pencils, and many kinds of animals. That is the reason why it is supposed that the AlexNet has learned and obtained plenty of feature representations.

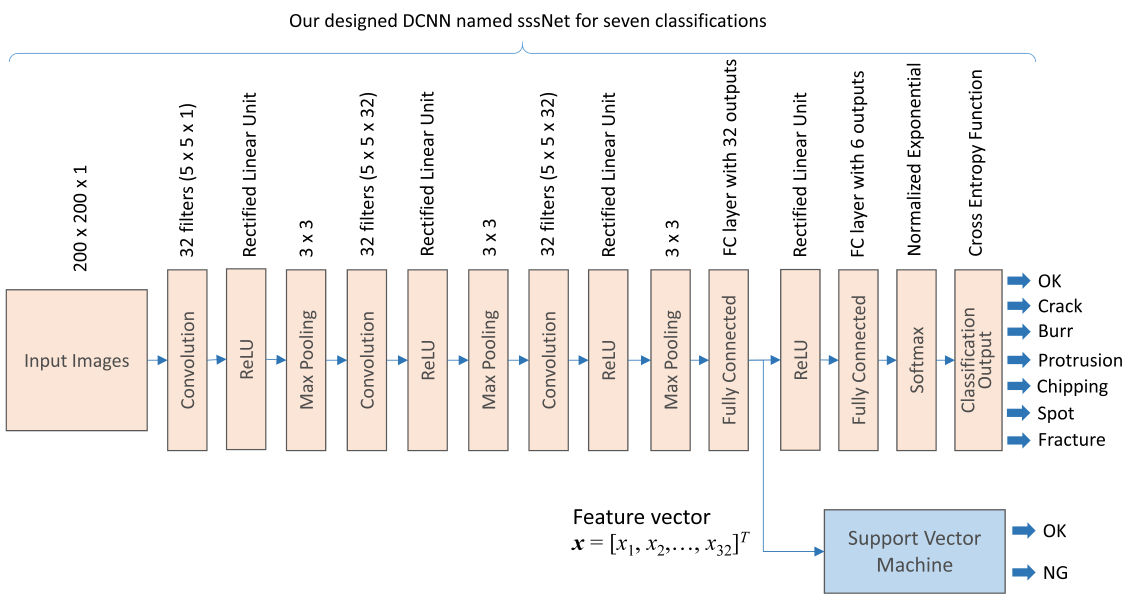

Fig. 2 The proposed binary class SVM whose input is the feature vector generated from our designed DCNN named sssNet

Figure 3 illustrates the binary class SVM whose input is the feature vector generated from the 2nd fully connected layer (20th layer) in the AlexNet. Similarly, 5,100 OK images $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{5100} \in \Re^{4096 \times 1}$ were used for unsupervised learning of the SVM with the Alexnet. It also took about several minutes for training. The same conditions in the case of the SVM with the sssNet were applied to this training. After the training, k, N and b were obtained as 26.7690, 2, 667 and -1.0635, respectively.

3 Classification experiments using the two kinds of trained SVMs

After training the two kinds of SVMs, classification experiments were conducted to check the generalization ability to test OK and NG images. The test images were not used in the training process. Figure 4 shows the classification results, i.e., histograms, using the SVM shown in Fig. 2. The horizontal and vertical axes denote the output values from the SVM trained with our designed sssNet and the number of image samples, respectively. It is observed from the histogram given by Fig. 4 that the SVM can discriminate NG images from OK ones. As can be also seen from Fig. 4, the values of the output from the SVM is much larger than 1 when non-defective images are classified. This is the cause that the score given by Eq. (1) depends on the standard deviation $\boldsymbol{x}_s$ of feature vectors extracted by the DCNN, the support vectors $\boldsymbol{x}_i^*$ determined through learning and their

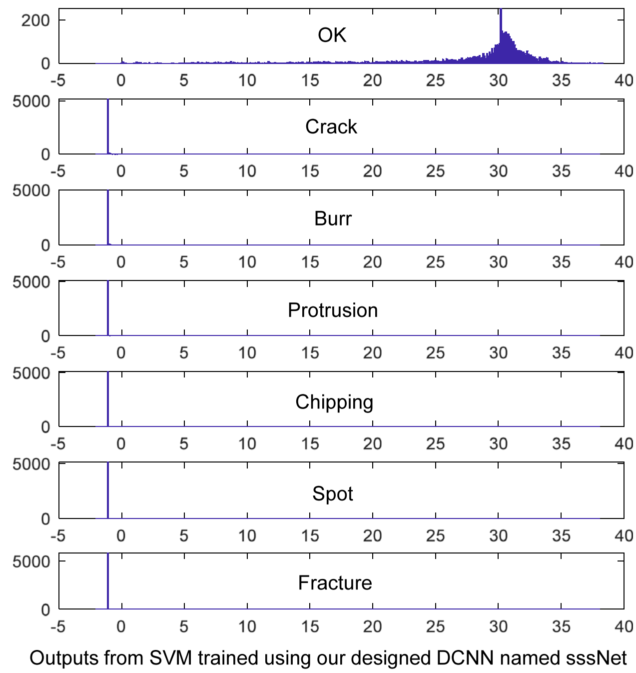

Fig. 4 Classification results using the SVM shown in Fig. 2, in which horizontal and vertical axes denote the output from the SVM trained with our designed sssNet and the number of image samples, respectively

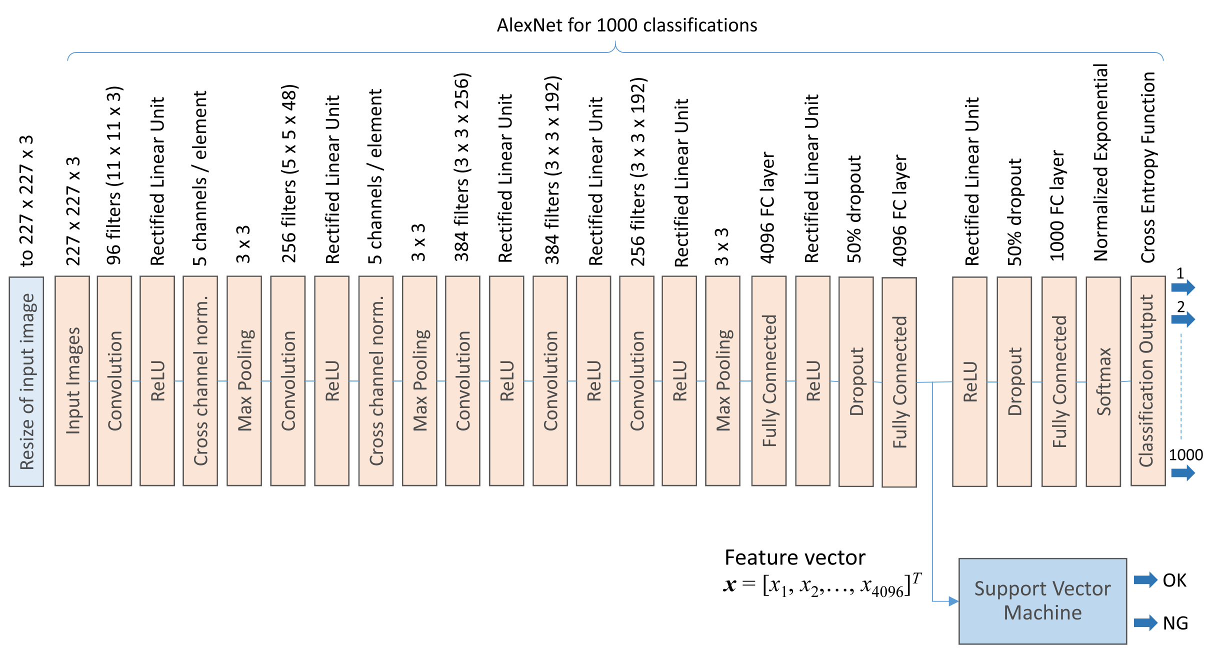

Fig. 3 The proposed binary class SVM whose input is the feature vector generated from AlexNet

number N. The authors have confirmed the tendency that the score becomes larger with the increase of the number.

Also, Fig. 5 shows the classification results using the SVM shown in Fig. 3. It is observed from Fig. 5 that the SVM with AlexNet can also discriminate NG images from OK ones with the almost same reliability as the SVM with sssNet. Actually, lengths of feature vectors generated from sssNet and AlexNet are quite different as 32 and 4,096; however, almost the same discrimination ability can be obtained. In the case of the target images given by $200 \times 200 \times 1$ resolution as shown in Fig. 1, the feature vector with 4,096 components given to SVM seems to be redundant.

The detailed comparison of the number of misclassified images with defects are shown in Table 1, in which it is observed that the sssNet could perform some superiority. Actually, the format of images given to the input layer of the AlexNet is fixed to $227 \times 227 \times 3$ (RGB 3 channels), so that all images with the format $200 \times 200 \times 1$ prepared in this experiment must have been superfluously converted to that as shown at the first layer in Fig. 3. On the other hand, the sssNet has been originally designed and trained to classify only the images of resin molded articles into non-defective one or seven kinds of defects. That is the reason why the superiority of sssNet to AlexNet is supposed to be the ability to extract more characterized feature vectors from the target images including the defects such as crack, burr, protrusion, chipping, spot and fracture in spite of the shorter feature vector.

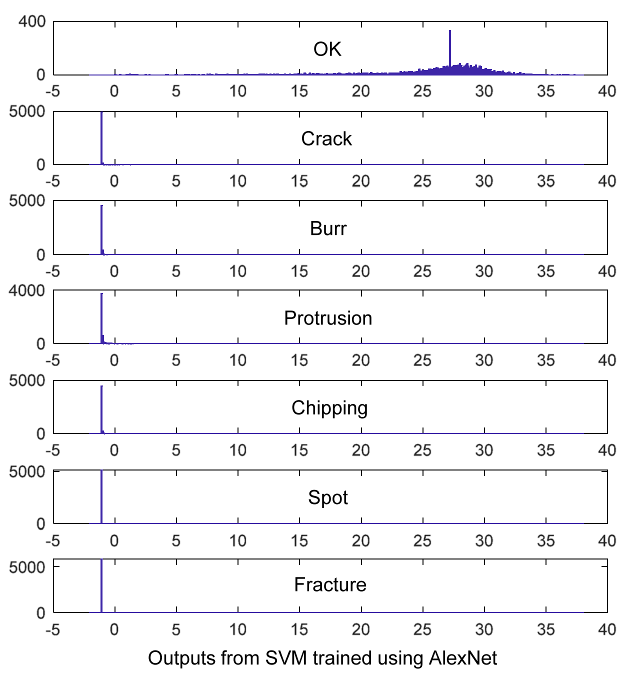

Fig. 5 Classification results using the SVM shown in Fig. 3, in which horizontal and vertical axes denote the output from the SVM trained using AlexNet and the number of image samples, respectively

4 Extraction of rectangle area including features

4.1 Image extraction using normalized cross-correlation

In this section, a template matching operation is introduced. When a template image, whose size is (M, N), is applied, then a padding operation makes the original target image larger as shown in Fig. 6. The original target image and the padded area is together called expanded target image. The correlation coefficient $\alpha(u, v)$ between the template image and an area with the same size in the target image expanded with the padding operation is calculated by [14]:

$$\alpha(u, v) = \frac{s_{it}(u, v)}{s_i(u, v)s_t(u, v)}, \tag{5}$$

$$s_{it}(u, v) = \sum_{y=v}^{v+N-1} \sum_{x=u}^{u+M-1} \{f(x, y) - \bar{f}_{u,v}\}\{t(x - u, y - v) - \bar{t}\}, \tag{6}$$

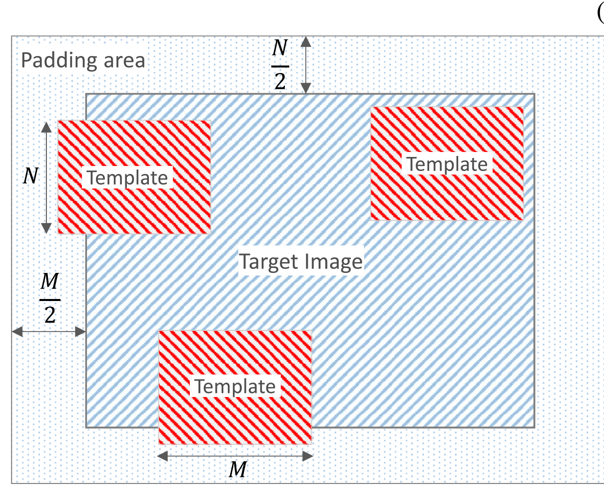

Fig. 6 Configuration among target image, padding area and template image whose size is (M, N)

Table 1 Detailed comparison of the number of misclassified images with defects

SVM	Burr	Crack	Chip.	Knob	Spot	Frac.
sssNet	13	4	1	0	0	0
AlexNet	167	20	298	127	0	0

$$s_i(u,v) = \sqrt{\sum_{y=v}^{v+N-1} \sum_{x=u}^{u+M-1} \{f(x,y) - \bar{f}_{u,v}\}^2}, \tag{7}$$

$$s_t(u,v) = \sqrt{\sum_{y=v}^{v+N-1} \sum_{x=u}^{u+M-1} \{t(x-u,y-v) - \bar{t}\}^2}, \tag{8}$$

where (u, v) is the left upper position of the template image in the expanded target image coordinate; $s_{it}(u, v)$ is the covariance; $s_i(u, v)$ and $s_t(u, v)$ are the standard deviations; $f(x, y)$ is the normalized value of grayscale at the position (x, y) in the expanded target image coordinate; $t(x-u, y-v)$ is the normalized value of grayscale at the position $(x-u, y-v)$ in the template image coordinate; M and N are the width and height of the template image, respectively; $\bar{t}$ is the mean value of grayscale in the template; $\bar{f}(u, v)$ is also the mean value of grayscale in an area just below the template.

The correlation coefficients $\alpha(u, v)$ based on Eq. (5) are sequentially obtained by raster scanning the template image from top left to bottom right in the expanded target image. After the raster scanning, an area best matched to the template image can be extracted by checking the maximum value of $\alpha(u, v)$. Figure 7 shows a template image and examples of extracted images using the template matching function. In learning of SVM and classifying by SVM in the next subsection, the proposed template matching is applied to all of the raw images of $200 \times 200 \times 1$ so that the resolution and the target area can be reduced and narrowed to $61 \times 59 \times 1$, respectively. This means that the template matching technique enables to reduce the overall calculation load concerning the image processing in DCNNs and SVMs.

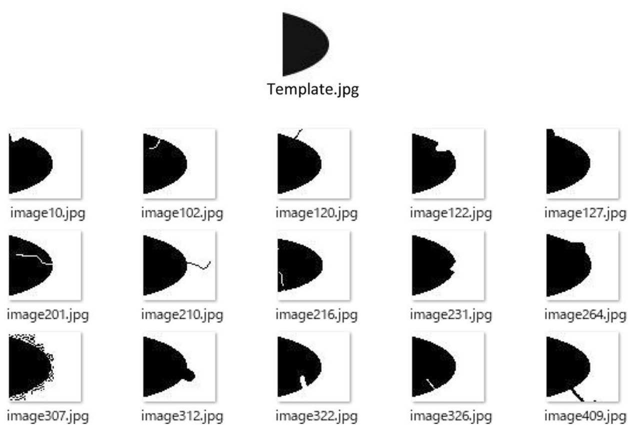

Fig. 7 Examples of extracted images using the template matching

4.2 Experiment

To evaluate the effectiveness of the template matching, 3,000 OK images as shown in Fig. 8 are trimmed from original OK ones. Then a SVM is designed based on the block diagram as shown in Fig. 3 and trained using the 3,000 trimmed OK images. After the training, the SVM is evaluated using 120 test images as shown in Fig. 9. The test images consist of 20 OK ones without a defect and 100 NG ones with one of the defects as shown in Fig. 1. Figure 10 shows the binary classification result of 120 test images shown in Fig. 9, in which images with minus and plus scores are estimated as NG and OK, respectively. It is observed from the classification experiment that all the test images are satisfactorily classified into OK or NG category.

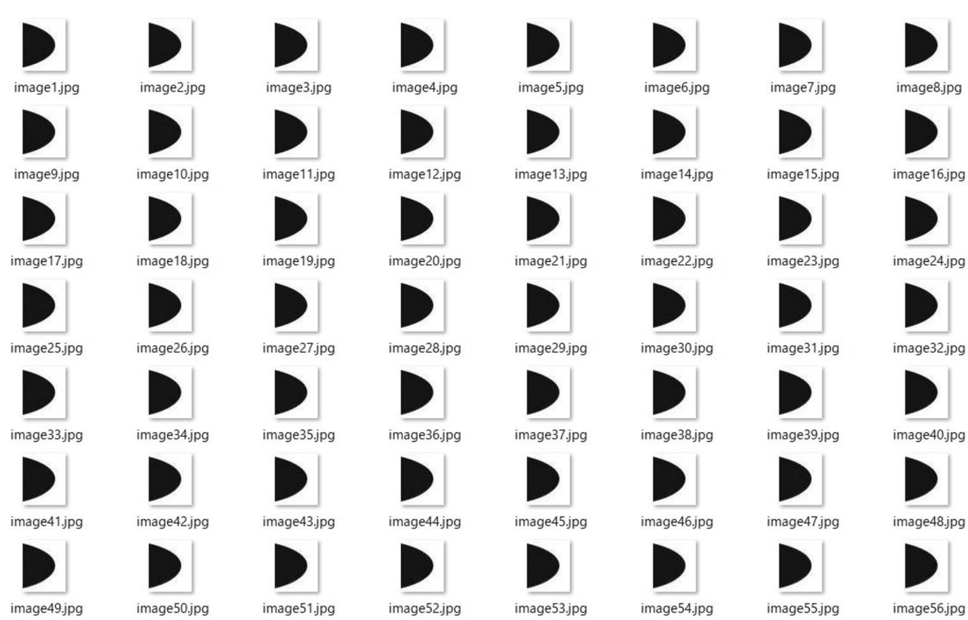

Fig. 8 Some of 3,000 OK images for training the SVM using AlexNet

Fig. 9 Some of 120 test images for evaluating the SVM trained using images shown in Fig. 8

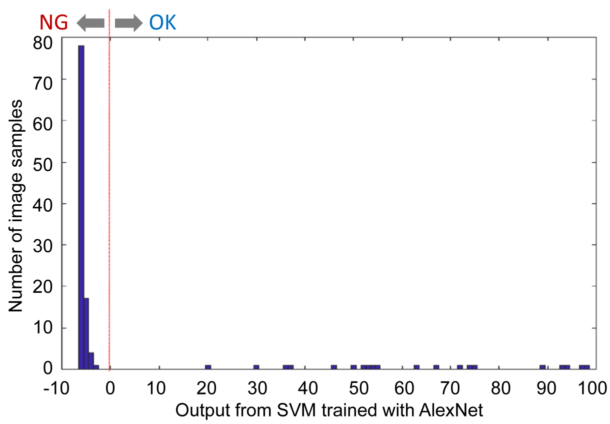

Fig. 10 Binary classification result of 120 test images shown in Fig. 9

5 Conclusions

In this paper, two types of SVMs for the defect detection of resin molded articles are designed and trained using the developed DCNN&SVM design tool, and evaluated to discriminate NG sample images including typical defects from OK ones seen in the manufacturing process of resin molded articles so that it is confirmed that the SVM with our designed sssNet can perform almost the same recognition ability as that with AlexNet in spite of the shorter feature vector. In addition, a template matching technique is successfully applied together with the SVM using AlexNet to narrow the important featured areas in training and test images so that the overall calculation load to deal with the many images in training and processing can be reduced.

Acknowledgements This work was supported by JSPS KAKENHI Grant number 16K06203 and MITSUBISHIPENCIL CO., LTD.

References

1. Guo T, Dong J, Li H, Gao Y (2017) Simple convolutional neural network on image classification. In: Proceedings of 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), pp 721–724, Beijing
2. Deng YS, Luo AC, Dai MJ (2018) Building an automatic defect verification system using deep neural network for PCB defect classification. In: Proceedings of 2018 4th International Conference on Frontiers of Signal Processing (ICFSP), pp 145–149, Poitiers, France, Sept 2018
3. Wang X, Chen Z, Liu G, Wan Y (2017) Fiber image classification using convolutional neural networks. In: Proceedings of 2017 4th International Conference on Systems and Informatics (ICSAI), pp 1214–1218
4. Shao J, Shi H, Du D, Wang L, Cao H (2011) Automatic weld defect detection in real-time X-ray images based on support vector machine. In: Proceedings of 2011 4th International Congress on Image and Signal Processing, pp 1842–1846, Shanghai, China, Oct. 2011
5. Niu XX, Suen C (2012) A novel hybrid CNN-SVM classifier for recognizing handwritten digits. Pattern Recognit 45(4):1318–1325
6. Sun X, Park J, Kang K, Hur J (2017) Novel hybrid CNN-SVM model for recognition of functional magnetic resonance images. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp 1001–1006
7. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
8. Chan GCY, Muhammad A, Shah SAA, Tang TB, Lu CK, Meriaudeau F (2017) Transfer learning for diabetic macular edema (DME) detection on optical coherence tomography (OCT) images. 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp 493–496

9. Wu H, Huang Q, Wang D, Gao L (2018) A CNN-SVM combined model for pattern recognition of knee motion using mechanomyography signals. J Electromyogr Kinesiol 42:136–142
10. Nagata F, Tokuno K, Watanabe K, Habib MK (2018) Design application of deep convolutional neural network for vision-based defect inspection. In: Proceedings of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp 1701–1706
11. Nagata F, Tokuno K, Ochi H, Otsuka A, Ikeda T, Watanabe K, Habib MK (2018) A design and training application for deep convolutional neural networks and support vector machines developed on MATLAB. Lecture Notes in Mechanical engineering (LNME) and communications in computer and information science (CCIS), Springer, Berlin, p 7
12. Platt J (1998) Sequential minimal optimization: a fast algorithm for training support vector machines. Tech Rep MSR 98:1–24
13. ImageNet. http://www.image-net.org
14. Lewis JP (2001) Fast normalized cross-correlation. Ind Light Magic, p 7