Specular reflection Surface Defects Detection by using Deep Learning

Zhong Zhang
Dep. of Mechanical Engineering
Toyohashi University of Technology
1-1 Hibarigaoka Tenpaku-cho,
Toyohashi 441-8580, Japan
zhang@me.tut.ac.jp

Borui Zhang
Dep. of Mechanical Engineering
Northeastern University
No. 3-11 Wenhua Road, Heping
District Shenyang, P. R. China
18640573525@163.com

Takuma Akiduki
Dep. of Mechanical Engineering
Toyohashi University of Technology
1-1 Hibarigaoka Tenpaku-cho,
Toyohashi 441-8580, Japan
akiduki@me.tut.ac.jp

ABSTRACT
As you know that defects inspection of specular surface is very difficult because its specular reflection is very strong and defects' reflection is weaker. And the existing computer vision-based industrial parts surface defect detection methods are limited by environmental factors, and the image preprocessing process is complex. On the other hand, with the rapid development of Convolutional Neural Networks (CNN) that is one type of deep learning and has excellent performance for image processing, has led to the rapid development of computer vision research based on deep learning. In this paper, we proposed an ensemble CNN in which integrated two convolutional neural network models for surface defect detection, and obtained better results.

CCS Concepts
• Networks → Network economics.

Keywords
Specular surface; Defect detection; Convolutional neural network; Machine vision；Networks

1. INTRODUCTION
The surface defects of industrial products have adverse effects on the aesthetics, comfort, and usability of the products. Machine vision-based surface defect inspection [1,2,3] can largely overcome the shortcomings of manual inspection methods, such as Changes in detection accuracy, reduction in efficiency due to fatigue and so on, are increasingly widespread in modern industry.

Machine vision-based surface defect detection has the advantages of non-contact, high speed, and high accuracy, but it also has particular drawbacks, such as image ambiguity, interference and influence caused by environmental factors and image registration algorithm [4]. In order to overcome the above problems, in this study based on the Convolutional Neural Networks (CNN)[5,6,7], we propose an ensemble CNN in which integrated two convolutional neural network models for surface defect detection. The ensemble CNN does not require manually designed machine vision systems and does not care much about the features of the image. Compared with the image registration algorithm, the image preprocessing process is simple and the generality is strong.

The structure of this paper is as follows: Section 2 describes the machine vision system based on the characteristics of the surface. The third section explains the image preprocessing process. In section 4, the CNN models used in deep learning and their principles are explained. Section 5 describes the ensemble model and compares the learning performance of ensemble model with separated CNN model.

2. REVIEW OF THE CNNs
In this study, a multi-layer convolutional neural network was used to identify and detect pre-processed images, and the parameters were continuously adjusted according to convolutional neural network feedback to achieve optimal results.

2.1 Traditional CNNs
Convolutional neural network (CNN) is a neural network that processes and extracts data in a grid structure. In the name of convolution, it refers to a neural network that uses a convolution operation at least in one layer of the network instead of a general matrix multiplication operation, that is, at least one convolution layer is used in the entire network structure. Convolutional neural networks have been developed in early deep learning domains such as LeNet [5] and Alexnet [6]. Neural networks can have a wide variety of topologies. The most common of these is the "Multilayer Fully Connected Feedforward Neural Network" [7]. A basic convolutional network consists of three basic components: convolutional layer, pooled layer and output layer.

(1) Convolution layer. A weight matrix is defined in the convolutional layer to extract certain features from the image. This weight is combined with the image, and all pixels are covered at least once, thereby producing a convoluted output. The weight matrix appears in the image like a filter that extracts specific information from the original image matrix. The convolutional neural network first learns the weights and then minimizes the loss function, similar to a multilayer perceptron (MLP) [7]. As the network structure becomes deeper, the features extracted by the weight matrix become more and more complex.

(2) Pooling layer. The sole purpose of pooling is to reduce the size of the image space. The pooling is done on its own in each depth dimension, so the depth of the image remains the same. The most common form of pooling layer is the largest pooling. The maximum pooled image still retains the information of the image and the size of the image has been halved. This can greatly reduce the parameters.

(3) Output layer. After multi-layer convolution and pooling, it needs to be output as a class. Convolution and pooling layers only extract features and reduce the parameters that the original image

ICISDM 2019, April 6–8, 2019, Houston, TX, USA

ACM ISBN 978-1-4503-6635-9/19/04…$15.00
https://doi.org/10.1145/3325917.3325930

brings. The output layer has a loss function similar to the classification cross-entropy [7] used to calculate the prediction error. Once forward propagation is complete, back propagation starts to update weights and deviations to reduce errors and losses. For the depth characteristics, we still use the loss function of Softmax 7] as the network objective function to guide the learning process. Replacing a fully connected layer network with global average pooling generally has better predictive performance.

2.2 Different Two Independent CNNs Models

The convolutional neural network used in this study refers to ConvPool-CNN-C and ALL-CNN-C model convolutional neural networks [8]. The above two models have been modified in light of the actual situation and are shown in Figure 1 and Figure 2. Compared with the traditional neural network, a global average pooling layer is used instead of the full connected layer. The bottom is the input to the first convolutional layer, containing the original content, the surface picture of the object to be detected, that the computer is trying to interpret. The top layer is the output, that is, the computer's final conclusion, as to whether the surface is defective. The middle of the two is the mathematical functional layer, in which it compresses the most important identification information and conducts it to the next layer.

The input in this study is a single-channel gray scale image of 64×64 pixels and can be considered as a 2d matrix. Each pixel value in the matrix is 0 to 255, where 0 is for black and 255 is for white. According to the surface defect condition, the image is marked in advance, and the image is divided into two types: perfect and defective. After the pre-processed image is created, the training set is input into the convolutional neural network for learning.

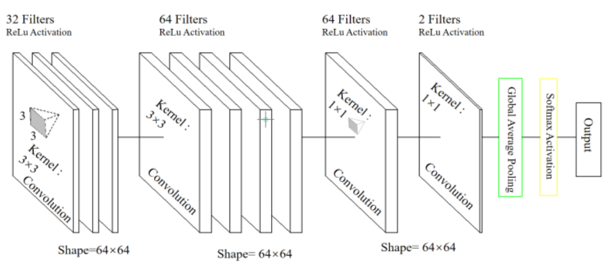

Figure 1. ConvPool-CNN-C-based model

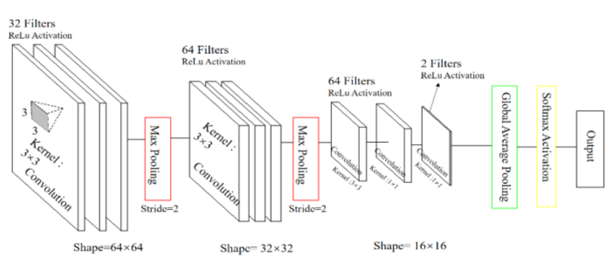

Figure 2. ALL-CNN-C-based model

The data set consists of 30000 images, two-category 64×64 RGB images. Of these, 24000 images were for training/verification and the other 6000 images were for testing. The convolutional neural network used in this study inputs an image of a 64×64 grayscale image. Convolution is performed on the first convolution layer using a 3×3 collation input picture, and the convolutional feature matrix is output. Product, nonlinearity, pooling, and final classification output neural network judgment results. After each convolution operation, a processing called Rectified Linear Unit (ReLU) [5,6] is performed

3. NEW ENSEMBLE CNN AND MACHINE VISION-BASED DEFECTS DETECTING SYSTEM

3.1 New Ensemble CNN Proposed

Ensemble learning refers to the ability to combine multiple classifiers and then combine them to achieve better predictive performance. The key to ensemble learning is to allow each individual learner to have its own differences, while at the same time having a certain error rate upper bound to integrate all the individual learners. The key to integrated learning is to allow each individual learner to have its own differences, while at the same time having a certain error rate upper bound, integrate all the individual learners. In the results of a large machine learning competition, the best result is usually the integration of the model rather than a single model. For example, a highest scoring single model structure achieved the 13th place in ILSVRC2015, while the 1st to 12th have used different types of model integration.

The definition of an ensemble model is straightforward [9]. It uses the input layer shared by all models. In the top layer, the integration calculates the average of the two models outputs by using the average merge layer. Furthermore, we combine the two trained models into one new ensemble model and average the integrated model outputs. This study intends to obtain better detection results by ensemble two different types of convolutional neural networks. At the output, the two models are grouped together and evaluated. There are many different types of ensemble model, this study uses stacking algorithm. The stacking algorithm involves the training of a learning algorithm combined with predictions from a variety of other learning algorithms. The schematic diagram of the proposed ensemble model is shown in Figure 3.

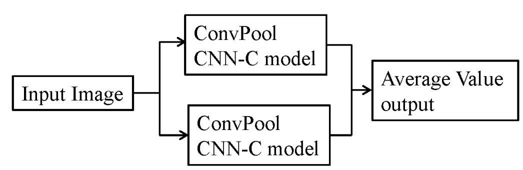

Figure 3. Ensemble Model schematic

3.2 Machine Vision-Based Defect Detecting System by using Ensemble CNN

In this study, the object to be tested is an irregular curved surface part, and the surface is sprayed with a highly reflective paint. There are the following two detection difficulties: (1) The surface defects of the pixels are too tiny to be directly observed with the naked eye; (2) The angle of the surface reflected light changes with the position of the curved surface. Therefore, a machine vision-based defects detecting system is introduced to solve the above problems.

The machine vision-based defects detecting system refers to the use of computers to achieve human visual functions, that is, using computers instead of humans to do image identification, measurement and judgment [10]. This system is mainly composed of three parts: image acquisition, image pre-processing and defects identification.

3.2.1 Image acquisition device

The image acquisition device is shown in Figure 2. Lighting is an important factor affecting the input of the machine vision system and directly affects the quality of the image. The light source used in this study was an LED light source. The light source and the camera are located on the same side of the part under test and at a certain angle. The advantage of this illumination method is that it can obtain a good reflected light image and facilitate adjustment and installation. A computer-connected camera is used to acquire the images from the target parts. The optical parameters used for the lens in the shot are as follows: focal length 30mm] (35mm] equivalent focal length), aperture F8, ISO sensitivity 800. The camera is fixed on a semi-circular adjustable track above the part to be tested and the image is displayed on a computer monitor in real time. So the position of the light source and the angle of reflected light can be used to adjust the position of the camera in time and obtain the optimal matching parameters of the optical system.

In this study, the surface to be detected is an irregularly curved surface. In order to detect the entire surface, the component needs to be rotated. Therefore, a rotatable fixing device is used to photograph the surface at different angles. The rotation angle is 0-360 [deg.]. Adjust the angle between the light source and the detection part before shooting, and fully adjust the relative positions of the light source, the part, and the camera, so that the CCD camera can accurately capture the reflected light on the surface of the part, and obtain the best images.

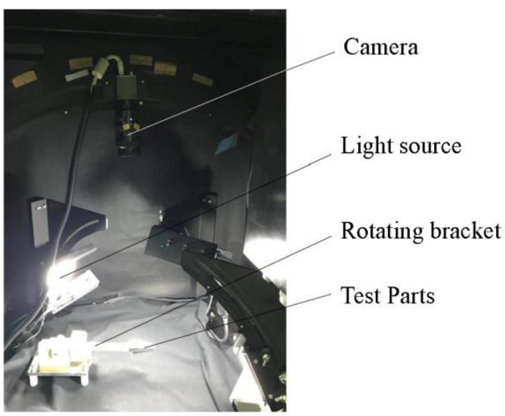

Figure 4. Image acquisition device

3.2.2 Image Pre-preprocessing

Before image recognition and identification, image preprocessing is required for the collected original image. The main purpose of image preprocessing is to eliminate irrelevant information and interference in the image, restore useful real information, simplify the data to the utmost, and improve the reliability of image detection.

(1) Image segmentation. Image segmentation refers to extracting meaningful parts of an image and analyzing them. The purpose of doing so was to remove the interference of other parts. The speed at which the computer performs convolution calculations is positively related to the size of the image. Compared to the original image, the processed image is smaller size, leading to a faster computer operation.

(2) Noise reduction of images. The shooting environment in this study was a closed black box and there is no other light irradiation light source causing interference. However, under long-term sensation, the current of the photosensitive element (CCD) will be locally overheated, it is able to result in excessively bright spots or noise, and the noise may interfere with the judgment of the defect in the machine vision system. Therefore, before the image recognition, the original image needs to be de-noised. Under the same optical conditions, noise appears to be random, and it can be offset to eliminate the effect of noise. In this study, the parts at the same angle were photographed five times in succession, and the image average algorithm [11] was used to reduce the noise. Figure 3 shows de-noise results, where the picture on the left shows the picture before the de-noising while the picture on the right shows the picture after the de-noising.

3.2.3 Defects Identification using Ensemble CNN

In this study, the Ensemble CNN that shown in Section 3.1 was used to identify after pre-processed images, where the parameters were continuously adjusted according to machine learning algorithm to achieve optimal results. Details of the machine learning will be introduced in the next session.

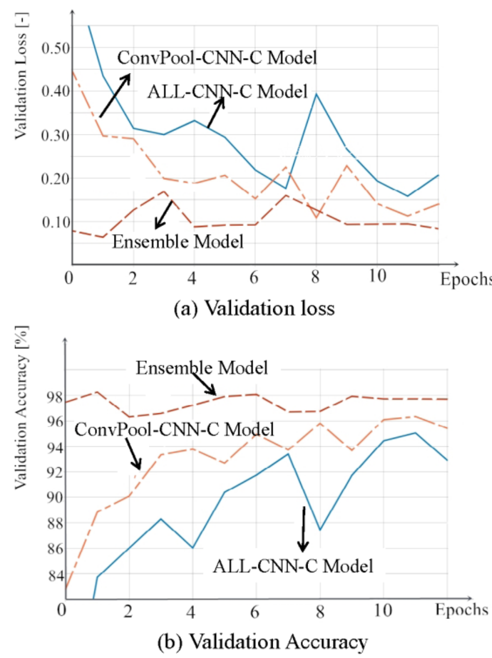

Figure 5. Validation Loss and validation accuracy of 3 models

4. EXPERMENTAL RESULTS AND DISCUSSION

For neural network weight setting and parameter setting, the following works have been done. (1) According to learning curve feedback, set epoch to 10, train 10 rounds to prevent over-fitting. (2) Specify the number of samples each batch contains when performing gradient descent 50 times. (3) The optimizer uses the Adam optimizer [7] and the learning rate is set to 0.0005, then start training.

In this study, the input is a single-channel grayscale image of 64×64 pixels. After the pre-processed picture is produced, the training set is input into the convolutional neural network for learning. The data set consists of 30,000 pieces of 64×64 RGB images. Of these, 24,000 were used for training/verification and the other 6000 pieces were used for testing.

After initialization, two separated CNN models are trained. At the very beginning of the learning process, the exact derivative of the loss function is not calculated. Instead, it is an estimate of a small batch of data, which means that the learning process does not always move in the best direction. Therefore, during the initial learning process, the validation and loss fluctuate. When the training process is over, save the weight of the model. The two models are then combined into one ensemble model, and both models are re-instantiated and loaded with the best saved weights. Because the weights of the previous models have been saved, the model already contains much useful information before the first training, so the learning accuracy of the ensemble model at the beginning is relatively high.

Table 1. Test results of ensemble model

Detect surface properties	Detected as flawless	Detected as defective	Test Accuracy
Defect-free surface (1364 sheets)	1235	129	90.54%
Defective surface (1634 sheets)	20	1616	98.77%

The performances of the two independent models and the ensemble model on the test set are shown as Figure 5 (The fluctuations in the initial learning process are omitted), where figure (a) shows validation loss of 3 models and (b) shows validation accuracy of 3 models. It can be seen from Figure 5 (b) that after 10 epochs, the verification accuracy of the ensemble model for the test set is about 97%, which is about 2% better than a single neural network model. Additionally, Figure 5 (a) shows the validation loss also performs better, and the ensemble model has a lower error rate than a single model.

It can be seen from that for the same test set, the ensemble model has higher recognition accuracy by comparing to a single CNN model, because each single CNN model has its own drawbacks. Therefore, by stacking different models that characterize different hypotheses about the data, a better hypothesis can be found out that is not in a hypothetical space from which to build an integrated model, and an ensemble model can be achieved a lower error rate than a single model. As expected, this ensemble model achieves better performance on the test set than using either model alone.

Table 1 shows test results of ensemble model, where performance evaluation of ensemble network modules with 3000 different tests images. It can be seen that the integrated deep learning model has a good recognition rate for the surface of the crucible, but it is often easy to mistake the flawless surface as the surface of the crucible, thereby increasing the production cost. Therefore, it is necessary to perform a secondary detection on the surface considered to be flawed by a subsequent image detecting step.

Image registration analysis was performed on the detected 1745 defective surfaces, and it was found that the false detection was 129 pictures with defects, and the Hamming distance range from the standard image template was 0-25. For the 1616 pictures with defects, the Hamming distance is 17-50. If the threshold is set to 17, only 46 pictures are falsely detected, which greatly improves the detection success rate.

But this method also has limitations. Because it can only be compared under the same illumination, its versatility is small. Moreover, this method has an overlapping area for the output of the Hamming distance of the defective image and the image without the defect. For an image in this interval, it is difficult to determine whether it is caused by a slight illumination change under photographing conditions or a misjudgment. Moreover, in the actual production process, compared with the convolutional neural network which is insensitive to illumination and shooting conditions, the detection versatility is also not good because the template image of the alignment is set in an fixed optical scene. Therefore, in this paper, it is only a supplement to the judgment of the ensemble model.

5. CONCLUSION AND REMARKS

In this study, we proposed a new integrated network called the ensemble model by combining two kinds of different neural networks and applied it to analyze a single surface image, whereby a good accuracy rate was achieved. That is, the verification accuracy of the ensemble model for the test set is about 97%, which is 2% better than a single neural network model.

However, there are still problems in the detection of surface defects. Firstly, the CNN has obtained a high recognition rate for the analysis of a single picture. If multiple images are taken on the surface of a component, as long as a flawless image on the surface is mistakenly detected as defective, the entire parts need to be viewed. Secondly, because the surface defects of the parts are tiny, and before the input of the CNN learning data set image, it is possible that small defects are also lost. So the next step of this study is to try to strengthen the defect features in the picture and reduce image loss caused by misdetection.

6. REFERENCES

[1] Golnabi, H., and Asadpour, A. 2007. *Design and application of industrial machine vision systems*. Robotics and Computer-Integrated Manufacturing, 23, 6, 630-637.

[2] Beyerer, Jürgen, Fernando Puente León and Christian Frese, Machine vision. 2015. *Automated visual inspection: Theory, practice and applications*. Springer. 015-017.

[3] Steger, C., Ulrich, M. and Wiedemann, C. 2018. *Machine vision algorithms and applications*. Will-VCH. 97-369.

[4] Bruck, H. A., McNeill, S. R., Sutton M. A. and Peters III W. H. 1989. *Digital image correlation using Newton-Raphson method of partial differential correction*. Experimental mechanics. 29, 3, 261-267.

[5] Lecun, Y., Bottou, L., Bengio Y., and Haffner, P. 1998. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*. 86, 11, 2278-2324.

[6] Krizhevsky, A., Sutskever, I. and Hinton, G. E. 2012. *ImageNet classification with deep convolutional neural networks*. Advances in neural information processing systems 25.

[7] Bengio, Yoshua, Goodfellow, Ian J. and Aaron Courville. 2015. *Deep learning*. Nature. 521, 7553, 436-444.

[8] Jan Erik Solem. 2012. *Programming Computer Vision with Python: Tools and algorithms for analyzing images*. O'Reilly.

[9] Thomas G. Dietterich. 2002. *Ensemble learning*. The handbook of brain theory and neural networks 2. 110-125.

[10] Springenberg, J. T., Dosovitskiy, A., Brox, T. and Riedmiller, M. 2014. *Striving for simplicity: The all convolutional net*. arXiv preprint arXiv:1412.6806 .

[11] Abdulkadir, S., Kadir, A. P., Ahmet, G. Y. and Emre, D. 2016. Fabric defect detection using deep learning. In *Proc. of 24th Signal Processing and Communication Application Conference*, 1437-1440.