International Journal of Precision Engineering and Manufacturing (2019) 20:363–374
https://doi.org/10.1007/s12541-019-00074-4

Online ISSN 2005-4602
Print ISSN 2234-7593

REGULAR PAPER

Convolutional Neural Network Based Surface Inspection System for Non-patterned Welding Defects

Je-Kang Park[1] · Woo-Hyun An[2] · Dong-Joong Kang[1]

Received: 29 March 2018 / Revised: 12 November 2018 / Accepted: 2 December 2018 / Published online: 22 February 2019

Abstract

In this paper, we propose a convolutional neural network (CNN) based method that inspects non-patterned welding defects (craters, pores, foreign substances and fissures) on the surface of the engine transmission using a single RGB camera. The proposed method consists of two steps: first, extracting the welding area to be inspected from the captured image, and then determining whether the extracted area includes defects. In the first step, to extract the welding area from the captured image, a CNN based approach is proposed to detect a center of the engine transmission in the image. In the second stage, the extracted area is identified by another CNN as defective or non-defective. To train the second stage CNN stably, we propose a class-specific batch sampling method. With our sampling method, biased learning caused by data imbalance (number of collected defective images is much less than that of non-defective images) is effectively prevented. We evaluated our system with a large amount of samples (about 32,000 images) collected manually from the production line, and our system shows a remarkable performance in all experiments.

Keywords Defect detection · Automatic inspection · Convolutional neural network · Machine vision · Image processing · Deep learning

List of symbols

i	Input nodes of layer
o	Output nodes of layer
w	Learnable weights of layer
b	Bias of layer
$\mathcal{F}$	Activation function
c	Estimated center of the engine transmission
$\hat{c}$	Ground truth center of the engine transmission
N	Number of training samples in a batch
r_1	Lower radius of the welding area
r_2	Upper radius of the welding area
θ_1	Lower angle of the welding area
θ_2	Upper angle of the welding area
p	Probability distribution of the estimated class
$\hat{p}$	Probability distribution of the ground truth class

✉ Dong-Joong Kang
djkang@pusan.ac.kr

[1] School of Mechanical Engineering, Pusan National University, 2, Busandaehak-ro, 63beon-gil, Geumjeong-gu, Busan 46241, South Korea

[2] Transmission Production Engineering Team 1 - Production and Development Division, Hyundai Motor Company, 700 Yeompo-ro, Buk-Gu, Ulsan 44259, South Korea

1 Introduction

Welding is a process of joining two metals by applying high temperature heat to the metal. This is a high-risk process for the worker because of the high temperature and high voltage working environment. Since the welding process is frequently required in the production of metal products, automation has been employed for a long time. However, various defects are frequently generated due to the characteristics of welding. Certain defects, where the width or height of the welding bead deviates from a prescribed range, can be measured using a non-contact distance sensor such as a laser sensor. However, other defects such as craters, pores, foreign substances inflow, and fissures on the welding surface are cannot be inspected accurately by the distance sensor alone. These defects show various patterns of a visual texture on the welding surface. Therefore, it should be recognized (not measurement) through machine vision.

The machine vision process for detecting defects on the surface of material generally proceeds in two steps: feature extraction and defect detection. The feature extraction usually extracts features using general image processing techniques such as smoothing, morphology, and edge detection [1–7]. Several methods [8–11] have been proposed a new

filter to extract a visual feature, or suggested a histogram-based feature extraction method. Traditionally, defect detection methods [4–6, 9–11] have been used to analyze a defect pattern and set thresholds on defective images. Since 2010, a data-driven approach [2, 3, 8, 12, 13] has been rapidly developed to learn a visual feature extracted from defective and non-defective images. Since 2012, convolutional neural networks [14] (CNN) based methods have been widely researched [15–19]. Since CNN can perform end-to-end learning to directly learn the input image, it is capable of learning the decision boundary and the convolution filter that optimized for defect detection.

The goal of this paper is to detect non-patterned defects on the welding surface of the engine transmission using CNN. Figure 1 shows the welded engine transmission. To achieve this goal, we first collected approximately 32,000 sample images from the production line, and manually labeled all samples for defective or non-defective. This large amount of samples allows CNN to learn various type of welding defects and ensure the reliability of the experiments results. In order to train CNN stably with such a large number of samples, the data imbalance problem should be solved. Generally, non-defective images are collected more than defective images. Since CNN depends heavily on training data, this imbalance of training samples leads to biased learning. Existing methods [15–19] use mini batch sampling which assumes that the number of samples per class is uniformly distributed. Therefore, it is not suitable for the actual production line such as our problem. Another problem in our case (inspecting the surface of engine transmissions) is the position of the welding area in the image is not fixed due to the vibration of the image acquisition system. This problem occurs frequently, but most existing methods [15–19] do not consider this problem and commonly assume the entire region of the image as an inspection region. On the other hand, some methods [3, 12] deal with this problem. They detect a circular-shaped product in the image by using the edge-based image processing. However, in our case, the edges are blurred due to the rapid rotation of the engine transmission, and the edges are often occluded by foreign substances. Therefore, it is difficult to extract the welding area accurately by existing method.

Fig. 1 Top view of the welded engine transmission. The shape of the welding area is circular

To overcome these problems, we propose a CNN based method that detecting welding area. Unlike existing methods [15–19] that only use CNN for defect identification, we also use CNN to train the regression function to detect the center of the engine transmission. Since CNN extracts high-level features, the arc-shaped welding area can be detected well even if the edges are blurred or damaged or if the illumination changes are severe. Furthermore, to solve the data imbalance problem, we propose a class-specific batch sampling method. Proposed sampling method keeps a proportion of defective and non-defective samples used in one step of learning constantly. Our sampling method is efficient and less complex than the hard negative mining technique of AdaBoost [13], which computes the weights of all training samples in every steps of learning.

2 Related Works

Early vision-based welding defect inspection methods [1, 4] typically use the traditional image processing algorithms. These methods first blur the image using a Gaussian filter to remove a high frequency noise. Then, the image segmentation and morphological operations are used to enhance the features of the inspection target area and to suppress the features of the background area. The region of interest (RoI), which is the region to be inspected, is extracted from the preprocessed image. For defect detection, a threshold based heuristic method is used. The threshold is selected from the extracted RoI by analyzing a difference of the pixel intensity or a histogram. The selected threshold is used as a criterion for detecting a defect in the test image. Since these threshold-based methods are sensitive to changes in the ambient environments such as illumination changes, the performance is frequently degraded. On the other hand, these threshold-based methods can easily detect the characteristic defects even with a small number of samples.

Recently, the data-driven approach that collecting and learning a large number of samples has been widely researched. Especially, CNN based method [20] can learn a visual feature of the fine defects. This method models the defect inspection problem as an image classification problem. First, the sample images for training network are collected and the collected images are classified according to the type of the defects. A CNN composed of multiple layers learns the weights in a gradient direction that minimizes

the loss of the cross-entropy objective function. This CNN-based method is robust to illumination changes and shows a high performance even in complicated inspection problems. However, CNN-based method requires large number of training samples and takes a long time to tune a number of hyperparameters.

3 Convolutional Neural Networks

In this section, we first briefly review CNN. CNN consists of one or more layers, and each layer has input nodes i and output nodes o. Depending on the type of layer, the number of output nodes and the types of operations performed between nodes are different. There are layers that learn weights (convolution layer, fully-connected layer) and layers that only perform computation (max-pooling layer). Training is proceeded through a feed-forward and backward process. In forward process, overall loss of CNN is computed. In backward process, loss is propagated backward by chain rules from the last layer to the front most layer. The weights of each layer are updated in the direction of the gradient.

3.1 Convolution Layer

Convolution means a filtering operation between the input nodes and the mask. With a specific mask, convolution operations can blur an image, remove noises, and extract edges. Unlike normal filtering operations where the mask is predefined, the mask of convolution layer is learnable weights.

$$o = \mathcal{F}(i * w + b) \tag{1}$$

In Eq. (1), * denotes a convolution operation, $\mathcal{F}$ denotes an activation function, w is a learnable mask, and b is a bias. Convolution layer extracts visual features from the input nodes and abstracts high-dimensional features by consecutively placing the convolution layer.

3.2 Max-Pooling Layer

Pooling refers to a subsampling operation which reduces the input nodes. Max-pooling layer repeats the operation of computing the maximum value at some adjacent input nodes for the all input nodes.

$$o = max(i_0, i_1, \ldots, i_{k-1}, i_k) \tag{2}$$

In Eq. (2), k is the size of the pooling window. Max-pooling layer preserves meaningful features with high activation on the input nodes while eliminating unnecessary features. Therefore, max-pooling layer allows CNN to learn features that are invariant to positional changes and significantly reduces CNN's learning complexity. In addition, since the layers placed after the max-pooling layer receive the reduced nodes, the computation of CNN is considerably reduced. In general, max-pooling layer is located after the convolution layer to sample features that extracted by the convolution layer.

3.3 Fully-Connected Layer

Fully-connected layer completely connects the input nodes and the output nodes and performs the same operation as the multi-layer perceptron of the conventional neural network. The output node is represented by a weighted sum of all input nodes.

$$o = \mathcal{F}(i \cdot w + b) \tag{3}$$

In Eq. (3), $\cdot$ denotes a vector–matrix multiplication, $\mathcal{F}$ denotes an activation function, w is a learnable weight, and b is a bias. Fully-connected layer can replace a classifier such as SVM [13] or Random Forest [8] and can be learned as a regression function. Typically, fully-connected layer is placed at the end of CNN after all convolution layers have extracted the features.

4 Defect Inspection System for Welding Surface

In order to inspect the welding area of the engine transmission, we first construct an image acquisition system that captures the surface of the engine transmission. Figure 2 shows the configuration of our image acquisition system. In Fig. 3, the vision sensor continuously captures a partial portion of the rotating engine transmission to observe the entire welding area. Note that since a high-resolution image is required to detect a minute welding defect, it is impossible to capture the entire welding area at once.

The proposed defect inspection system consists of two steps: extracting the welding area from the collected image

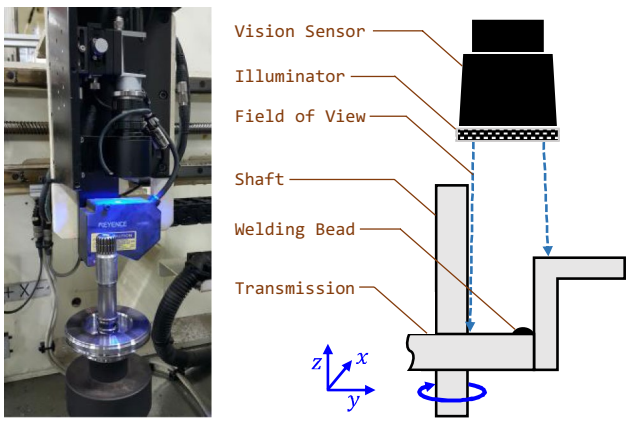


Fig. 2 Configuration of our image acquisition system

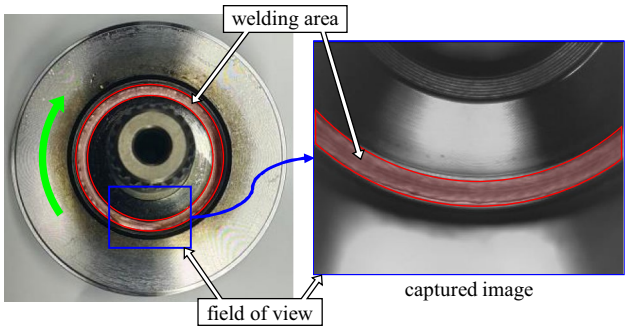


Fig. 3 Field of view of the vision sensor and the captured image. Since the shaft and the vision sensor are fixed, the welding area in the captured image always appears in the horizontal direction and shows only a portion of the entire welding area. To capture the entire welding area, we continuously rotate the transmission and capture the transmission several times

and detecting the defect in the extracted welding area. Figure 4 shows the flow diagram of the overall system. In the first stage, the center of the engine transmission is estimated from the collected sample image. In the second stage, the Polar to Cartesian (P2C) transformation is performed with the estimated center position and the radius of the engine transmission to extract the welding area. Then, the rectified welding area is identified as defective or non-defective by CNN. The rest of this chapter describes each stage in more detail.

4.1 Center Estimation

In the sample images collected from the production line, a positional error occurs due to the various types of mechanical vibration. As shown in Fig. 5, the arc-shaped welding area, which is the region to be inspected, vibrates in x and y directions. To extract the welding area accurately from the collected images it is necessary to estimate the center of the engine transmission in the collected images. This section describes how to select the region of interest (RoI) from the input image and the structure of CNN for estimating the center of engine transmission.

4.1.1 RoI Selection

To estimate the center of the circular transmission in the image, a circle or an arc must be included in the image. An arc that is long and not damaged by foreign substances or soot is a good clue for estimating the center of the transmission. In Fig. 6a, all four arcs with the same center are clearly visible. Thus, these arcs are good clues for estimating the center of the transmission. On the other hand, in Fig. 6b, l_1 and l_4 are missing due to the change of transmission model. In Fig. 6c, l_2, l_3 and l_4 are damaged due to foreign substances inflow, and in Fig. 6d, l_1, l_2 and l_4 are occluded due to the soot. Based on these observations, l_4 is often missed or damaged. Also, if l_4 is included in RoI, the size of the RoI region increases. Therefore, the region containing l_1, l_2 and l_3 is selected as RoI for estimating the center, as shown in Fig. 6a.

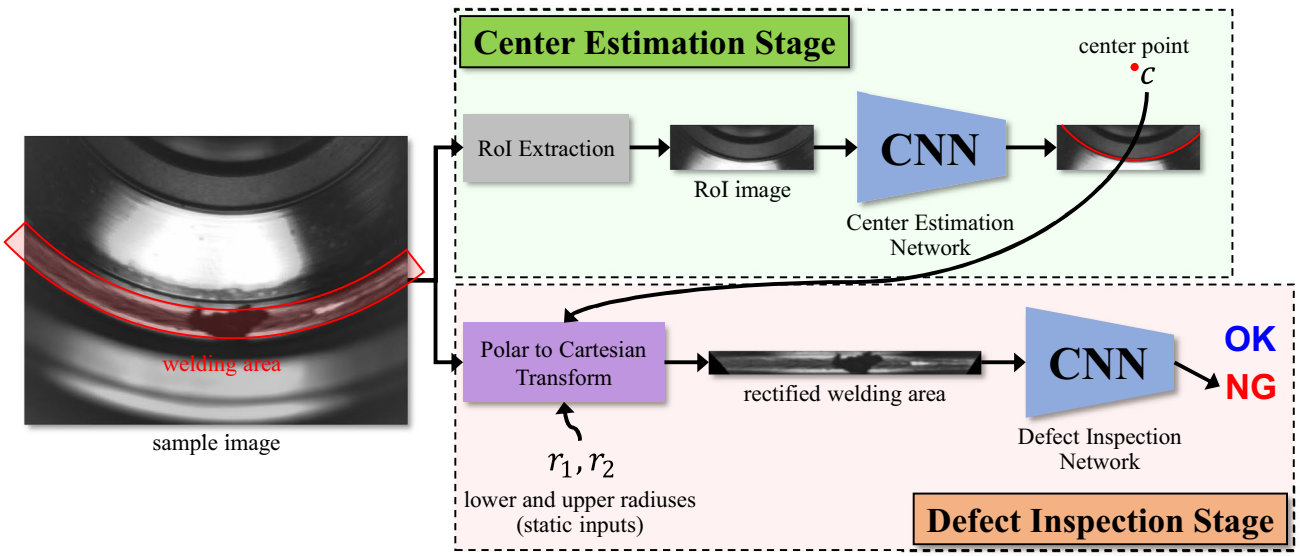


Fig. 4 Overview of the proposed system. First, the sample image is input to the center estimation stage to find the center of the transmission. In the next defect inspection stage, the welding area is extracted from the sample image and classified whether the extracted welding area is defective or non-defective

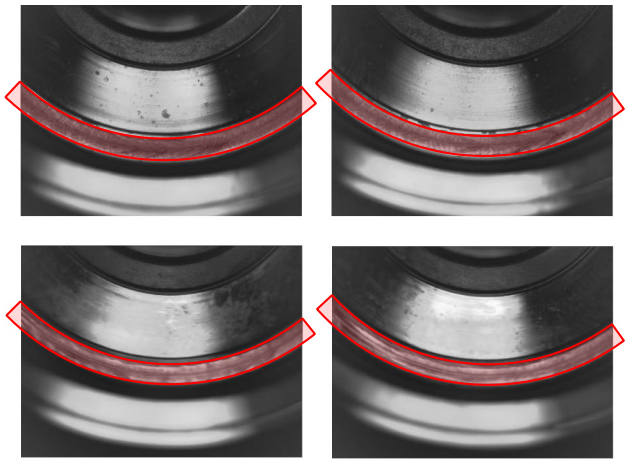


Fig. 5 Collected sample images. The arc-shaped red area represents the welding area to be inspected. Note that the position of the welding area in each sample is not fixed due to the vibration of the image acquisition system

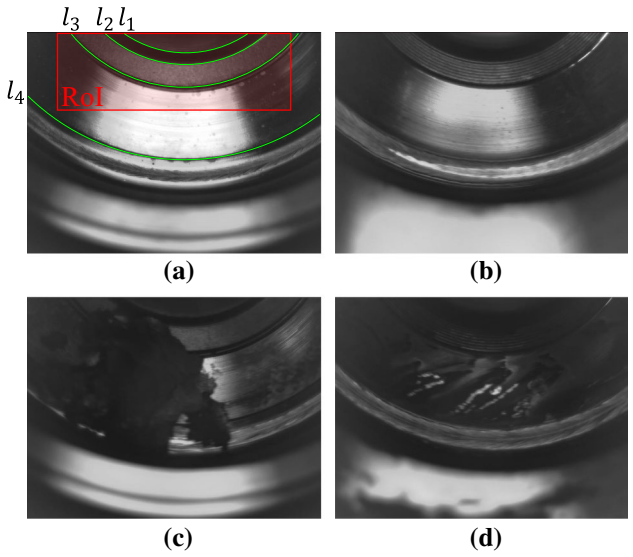


Fig. 6 **a** A sample whose center of the engine transmission can be easily detected. All arcs (l_1, l_2, l_3 and l_4) are clearly visible. **b–d** Some arcs are missing, invisible or corrupted

4.1.2 CNN for Center Estimation

A CNN for estimating the center of the engine transmission (CE-CNN) consists of a convolution layer for feature extraction and a fully-connected layer for learning the regression function. Figure 7 shows the structure of the CE-CNN that receives RoI image and outputs c (x and y coordinates) the center of the transmission. Max-pooling layers are placed after the convolution layers. To reduce the computational complexity and the number of

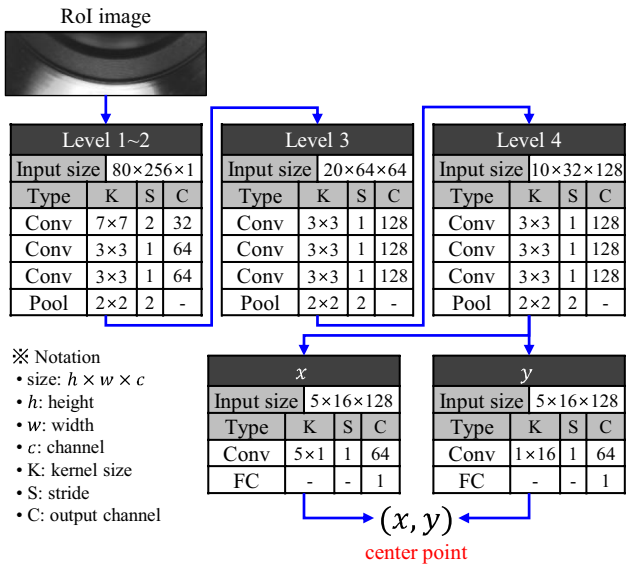


Fig. 7 Detailed structure of the center estimation network (CE-CNN)

learnable weights, the fully-connected layers that output x and y coordinates are constructed separately as in Park et al. [21]. In our case, the separation of x and y coordinates increases the accuracy by 0.11% and reduces the number of trainable weights by about 35%. The activation function of all convolution and fully-connected layers is ReLU [22] except for the last fully-connected layer. Since CE-CNN should be learned to output the center c closest to the ground truth center $\hat{c}$ of the transmission, the loss function is defined as Euclidean loss.

$$L = \frac{1}{N} \sum_{j=1}^{N} \left\| \hat{c}_j - c_j \right\|_2^2 \tag{4}$$

In Eq. (4), N is the number of training samples, and $\hat{c}$ is the manually labeled ground truth center of the transmission. All the learnable weights in the network compute the gradient for losses and update the weights in the direction of loss reduction.

4.2 Defect Inspection

Prior to inspecting the defect, the region other than the welding area (background) should be removed because it is not related to the defect detection and it interrupts the stable learning. In our system, the defect is detected through CNN, and CNN requires a rectified input image. In this section, to extract the arc-shaped welding area from the input image, we describe the process of P2C transformation and the structure of CNN for detecting the defects.

4.2.1 P2C Transform

Polar coordinate system is a two-dimensional coordinate system that represents a point on a plane in terms of a radius and an angle axis. Since the engine transmission does not move in the z direction during the image acquisition (see Fig. 2), r_1 and r_2 in Fig. 8 are constants. θ_1 is a angle between the x-axis and the point where the circle of radius r_1 intersects the right boundary of the image, and θ_2 is the angle between the x-axis and the point where the circle of radius r_1 intersects the left boundary of the image. P2C transform function that using the estimated center in the previous stage is as follows.

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \quad (r_1 < r \le r_2, \theta_1 < \theta \le \theta_2) \tag{5}$$

In Eq. (5), (r, θ) is the point on the Polar coordinate system and (x, y) is the point on the Cartesian coordinate system. After P2C transform, width and height of the transformed image are θ_2–θ_1 and r_2–r_1 respectively.

4.2.2 CNN for Defect Inspection

A CNN for inspecting the defect (DI-CNN) consists of convolution layers for feature extraction and fully-connected layers for learning a classification function. Figure 9 shows the structure of the DI-CNN that receives the P2C transformed image and outputs probability distribution p for the presence of the defect. Max-pooling layers are placed after the convolution layers. Activation function of all convolution and fully-connected layers is ReLU except for the last fully-connected layer. The output node o of the last fully-connected layer forms a probability distribution p through the softmax function of Eq. (7). Since it is a probability distribution of two classes of defective or non-defective, p is a vector of size 2 ($p \in R^2$). The objective value $\hat{p}$ is the probability distribution in the form of a one-hot vector, with a value of 1 for the element that corresponding to the ground truth class of the input image, and 0 for the other. The loss function is defined as a cross-entropy function that represents the distance of two probability distributions.

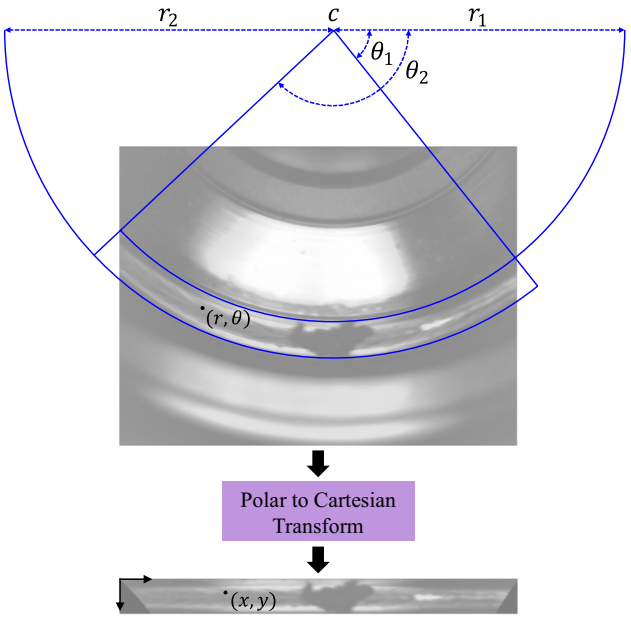


Fig. 8 The arc-shaped welding area is rectified through Polar to Cartesian (P2C) transform. Note that the center c comes from the center estimation stage, and both the inner radius r_1 and the outer radius r_2 are constant

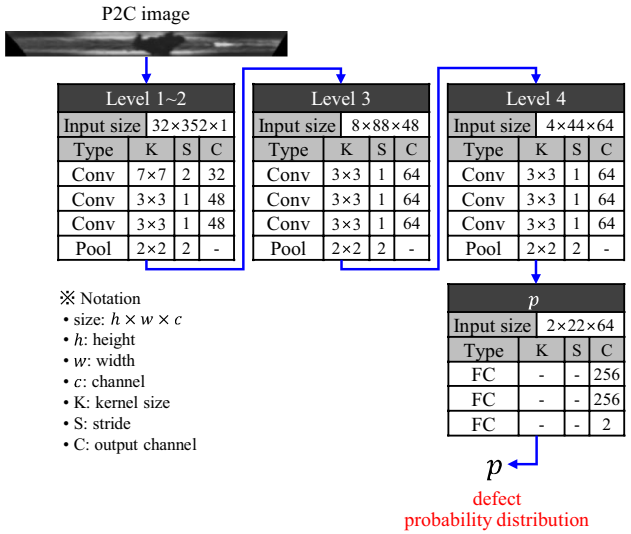


Fig. 9 Detailed structure of the defect inspection network (DI-CNN)

$$L = -\frac{1}{N}\sum_{j=1}^{N}\left[\hat{p}_j \log p_j + (1-\hat{p}_j)\log(1-p_j)\right] \tag{6}$$

$$p = \frac{e^{o}}{\sum_k e^{o_k}} \tag{7}$$

In Eq. (6), N is the number of training samples. All the learnable weights in the network compute the gradient for losses and update the weights in the direction of loss reduction.

4.3 Implementation Details

4.3.1 Class-Specific Batch Sampling

CNN finds weights that minimize the loss of the training sample through the gradient descent method. The process of computing the loss of training samples is called

feed-forward, while the process of updating weights in the direction of decreasing loss is called feed-backward, and the gradient descent method repeats this forward–backward process until CNN is converged. In this process, the increase in the number of training samples also increases the amount of computation (see Eqs. 4, 6). Thus, the mini-batch gradient descent method is generally utilized. Mini-batch gradient descent method uses a batch of the randomly selected training sample, rather than using all the training samples in a single forward–backward procedure. Mini-batch gradient descent method is established when the distribution of the number of training samples per class is similar to the uniform distribution. If the number of training samples per class is unbalanced, the learning may be biased or fail, as shown in Fig. 10b.

In the samples collected from the production line, the number of defective samples is very small compared to the number of non-defective samples, thus conventional mini-batch sampling results in a biased learning. In order to solve this problem, we sample batches from each class individually as shown in Fig. 10c. With this class-specific sampling method, the number of samples per class included in the batch is always maintained uniformly, so that the biased learning can be effectively prevented.

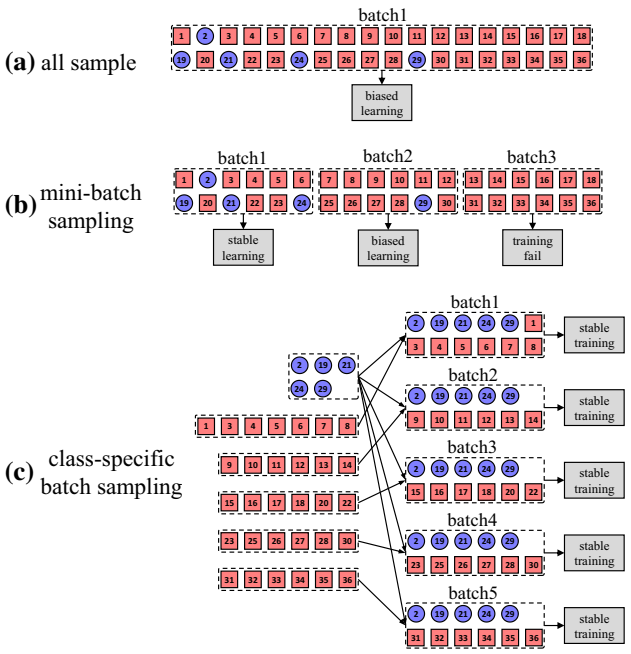


Fig. 10 Comparison of three types of sampling methods. **a** Sampling all samples always results in a biased learning. **b** Mini-batch sampling sometimes causes a biased learning or training fail. **c** Proposed class-specific batch sampling keeps the number of samples per class contained in the batch uniformly to prevent the biased learning

4.3.2 Data Augmentation

In the training of CNN, a large amount of training samples and various patterns of training samples are improve CNN's generalization performance. However, the number of samples collected in the production line is limited (especially in the case of defective images). To overcome the lack of training samples, we used a data augmentation technique. Data augmentation generates additional fictitious training samples by randomly transforming and distorting a raw image sample. Random image transformation consists of rotation, translation, and scaling of an image. Random image distortion consists of adjusting brightness and contrast of an image.

In training of CNN, we first sample a batch from the training samples, and then the image augmentation methods are applied to the sampled batch. Therefore, newly generated samples are always used for each step of training. Through this data augmentation strategy, we supplemented insufficient defective samples and achieved high generalization performance.

4.3.3 Parameters for Training

We used the gradient descent method to optimize our network. Initial learning rate was set to 0.01 and continuously decreased by multiplying it by 0.1 when the current iteration reached 50% and 83% of the total 6000 iterations. Class-specific batch sampling was used to form a batch. One batch has 128 samples consisting of 32 defective samples and 96 non-defective samples. Momentum is set to 0.9 and a weight decay is set to 0.0005, which were the same as AlexNet [14].

5 Experiments

5.1 Dataset

In the experiment, 32,014 transmission images that collected directly from the production line were used. The collected images were manually labeled as OK (indicating non-defective) or NG (indicating defective). The number of labeled OK samples is 30,937 and the number of NG samples is 1077. Due to the nature of the high temperature welding process for metals, some of the collected images contain a severe noise, such as soot and light reflection. Figure 11 (a) shows a normal OK, (b) shows a sooty OK, and (c) shows an OK where light is reflected in the welding area. Figure 11 (d) is an NG with a missing welding, (e) is an NG where a foreign substance is located on the welding area, and (f) is NG where a crater was generated.

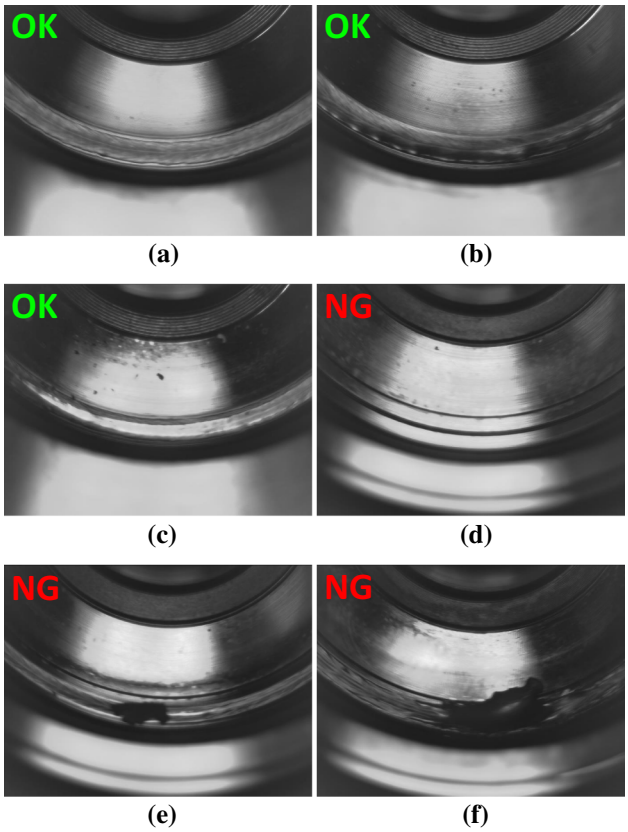


Fig. 11 Various patterns of OK (non-defective) and NG (defective) sample images. **a** normal OK, **b** sooty OK, **c** bright OK, **d** NG with a missing welding, **e** NG where a foreign substance is located on the welding area, **f** NG where a crater was generated

5.2 Evaluation of Center Estimation

We conducted a comparative experiment with an existing method to evaluate the center estimation performance of our CE-CNN. First, in order to train the CE-CNN, we selected 1000 images from the 32,014 transmission images and manually labeled the center position. The selected 1000 images have various illumination conditions and consist of 341 NG images and 659 OK images. The labeling criteria for the center position are l_1, l_2 and l_3 arcs in Fig. 6. A point closest to the center of the three arcs was labeled as the center of the engine transmission. We randomly selected 900 images for the training and 100 images for the validation for tenfold cross validation. For all training images, data augmentation techniques that randomly translating an image and distorting brightness and contrast were applied. The evaluation metric was the root mean squared error of Eq. (8).

$$error = \left\| \hat{c} - c_2 \right\| \tag{8}$$

In Eq. (8), $\hat{c}$ is the ground truth center and c is the estimated center of the engine transmission. Note that since the observation distance between the vision sensor and the surface of the engine transmission is always fixed in the image acquisition system, the scale of the engine transmission in the all images is also fixed. Thus, we do not use the normalization term in Eq. (8).

Table 1 Comparative evaluation results of center estimation

	Error (pixel)	Time (s/image)
Hough circle [23]—CPU	26.02	0.3509
Ours—CPU	5.41	0.0214
Ours—GPU	5.41	0.0073

Resolution of the test image is 2584 × 1944. Note that reported results are average performance of tenfold cross validation

The comparative experiment was performed with Hough circle detection method [23]. Hough circle is a method of accumulating the edges that extracted from the image in the Hough space and detecting N circles with the most votes. For fair comparison with the proposed method using training data, among the circles detected by the Hough circle, a circle nearest to the mean center of the training data was finally selected.

The mean error and the processing speed of each method are shown in Table 1, and the center estimation results are shown in Figs. 12 and 13. In Fig. 12a, Hough circle does not detect the exact center position because the edge was not extracted correctly due to the dark illumination. On the other hand, since the CE-CNN have learned with the distorted samples (brightness and contrast), CE-CNN estimated the center accurately, even in the blurry image. In Fig. 12b, there are dense circular patterns on the surface of the transmission between the arc l_2 and the arc l_3. Since the space between the dense circular patterns are too narrow, shifted center is estimated as a result. In the training of the CE-CNN, there were many samples like the one shown in Fig. 12b, and consecutive circular patterns can be used as good texture features. Thus, CE-CNN estimated the accurate center position.

We have further tested CE-CNN with several sample images of various illumination. Figure 13 shows the center estimation results. CE-CNN shows good estimation results in all images. Since randomly augmented samples of different brightness are used in the training of CE-CNN, CE-CNN shows good performance even in the samples of dynamic illumination. Note that the brightness change is not severe in the test images in Fig. 13 because the illuminator has installed on our image acquisition system.

Regarding the processing time, CE-CNN was very fast due to GPU parallelism. Hough circle was relatively slow because all operations should be processed sequentially within the CPU due to the voting operation, which is difficult to parallelize.

Hough transform has one important advantage. Hough transform is easy to apply because it does not require a

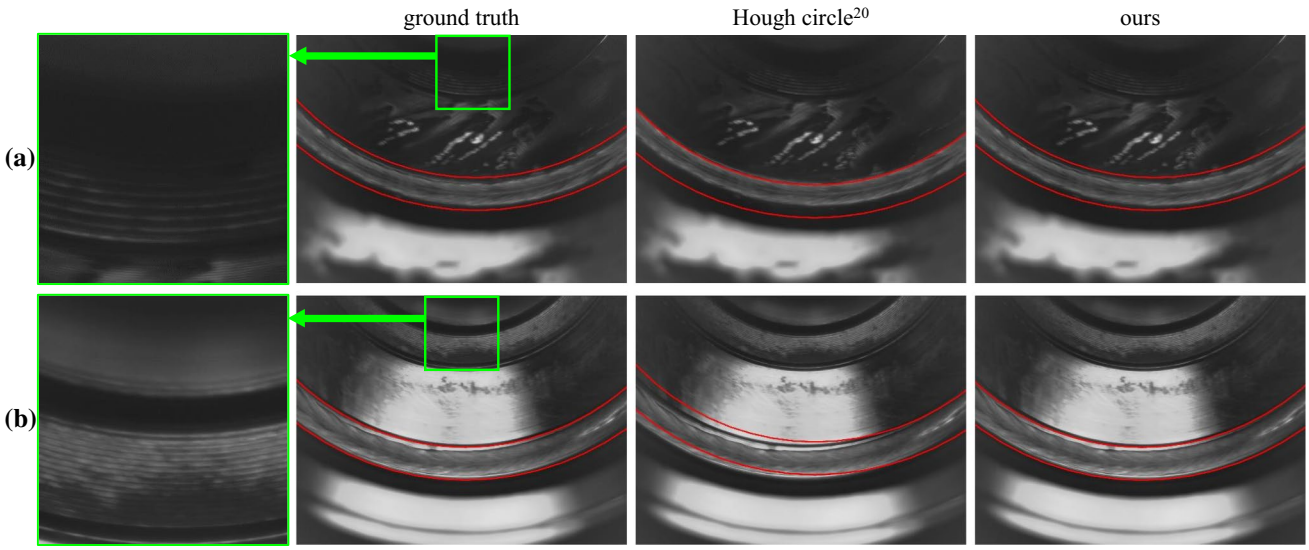


Fig. 12 Center estimation results of Hough circle and CE-CNN. The two red lines represent the inner arc and the outer arc of the welding area. If the center is correctly estimated, the welding area is located between the two arcs, like the ground truth image (2nd column). Conventional Hough circle (3th column) does not detect the center correctly for cases of **a** if the RoI (1st column) is too dark or **b** if dense circular patterns are observed in the RoI. On the other hand, our CE-CNN (4th column) estimates the center accurately in all cases

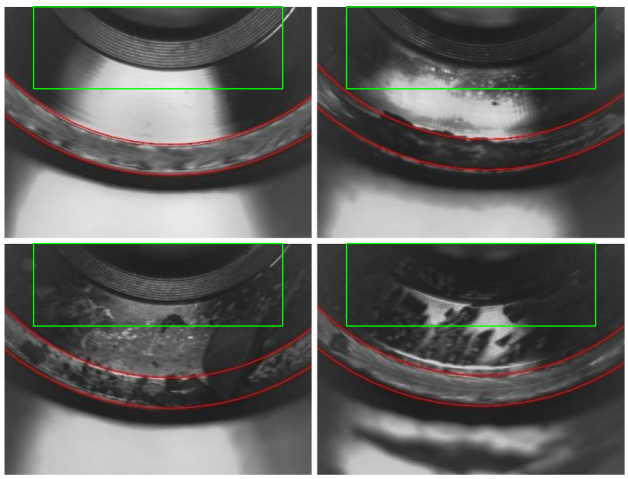


Fig. 13 Some center estimation results of CE-CNN. Note that the brightness of RoI (green box) varies in the sample images. CE-CNN estimates the center well even for the sample images under various and dynamic illumination

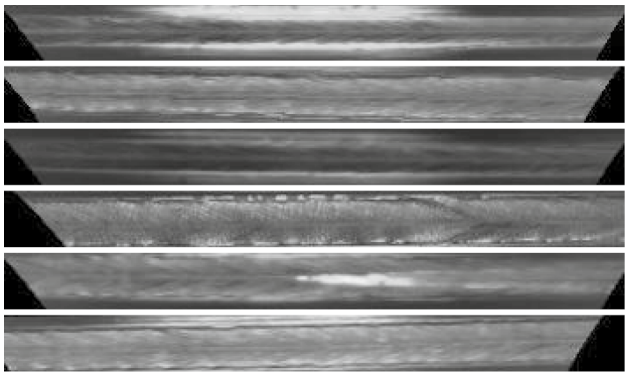
large number of training samples. On the other hand, proposed CE-CNN has higher accuracy and faster processing time. Thus, which method to use is a trade-off problem. In the inspection problem of the defective automobile engine transmission, accuracy is very important because the risk for safety accident is high when the defective product is predicted as a non-defective product. Therefore, even if it takes some time to collect the training samples, it is right to use the proposed CE-CNN because the CE-CNN shows better detection performance than the Hough transform. CE-CNN requires only 1000 sample images for training and this amount of sample images can be collected in half a day.

5.3 Evaluation of Defect Inspection

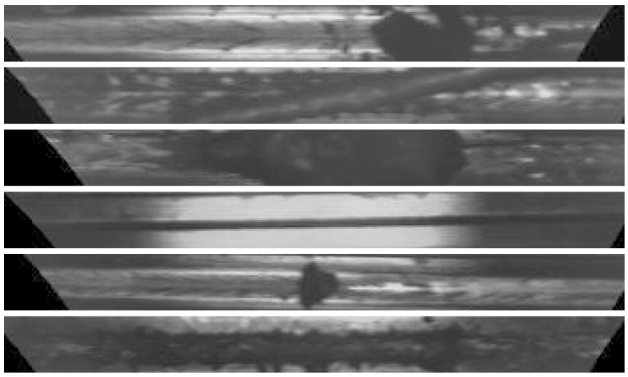
We conducted a comparative experiment with existing methods to verify the defect inspection performance of DI-CNN. All 32,014 images were used for the training and validation of DI-CNN. For tenfold cross validation, we randomly selected 28,809 images for the training and 3201 images for the validation. In order to extract the welding area from the image, the center estimation process of CE-CNN and P2C transformation were performed on all images. All data augmentation techniques were applied to the training images. Some of the training images are shown in Fig. 14. The evaluation metrics we used were true positive rate (TPR) and true negative rate (TNR). TPR is a ratio of the case where the image predicted as OK is actually OK, and TNR is a ratio of the case where the image predicted as NG is actually NG.

$$\begin{aligned} TPR &= \frac{n_{TP}}{n_{TP}+n_{FN}} \\ TNR &= \frac{n_{TN}}{n_{TN}+n_{FP}} \end{aligned} \tag{9}$$

In Eq. (9), n_{TP} is the number of correctly classified OK samples, and n_{FN} is the number of misclassified OK samples. n_{TN} is the number of correctly classified NG samples, and n_{FP} is the number of misclassified NG samples. Since TPR and TNR are mutually complementary, adjustment of the threshold that classifies OK and NG sample changes TPR



(a) OK images



(b) NG images

Fig. 14 Sample images used to train DI-CNN. All images are generated through P2C transformation based on the estimated center. Note that the samples images contain only the welding area

and TNR. In a general classification problem, the threshold is selected to make TPR and TNR are balanced. However, in our case (production line), the threshold for maximizing TNR should be selected since n_{FP} should be minimized.

In the comparative experiment, CNN based methods [15–17] and the classic feature based methods [13] were used. The classic feature based method [13] extracts the HOG feature from the image and learns a SVM classifier. We also added an experiment using the LBP feature that describes the texture patterns well. Differences between the CNN based methods are their network structure and the types of defects. The process for preparing the training samples and the optimization process of the networks are the same with the experiments of CE-CNN. For fair comparison with the proposed method, the P2C transformed training images were used and the same data augmentation techniques were used for all methods. We also conducted a comparative experiment between the mini-batch sampling method and our class-specific batch sampling method.

The results of the comparative experiments are shown in Table 2. The HOG feature based method [13] showed the lowest performance. HOG feature is not suitable for the defect inspection because it is mainly used for detecting the upright objects such as humans. LBP feature is suitable for pattern inspection, so it provides higher performance than the HOG feature, but it showed a not sufficient performance for using production line. On the other hand, all CNN based methods [15–17] showed an outstanding performance. Since the binary classification problem does not require a large network, all CNN based methods showed little variation in performance. In the comparison of the sampling methods, TPR was increased in all cases of using the class-specific batch sampling method. Our class-specific batch sampling method improved the performance in all CNN-based methods by preventing the network from being biased toward the large amount of OK samples.

6 Conclusions

In this paper, we proposed the CNN based method for the automatic inspection of the welding defects. To remove the background region from the collected images and to rectify the welding area, the center of the engine transmission was estimated by CE-CNN, and P2C transformation was performed. Then, the CNN based classifier is learned to inspect the welding defects in the extracted area. The imbalance

Table 2 Results of comparative evaluation of defect inspection

	Number of weights	Mini-batch sampling		Class-specific batch sampling	
		TPR (%)	TNR (%)	TPR (%)	TNR (%)
HOG + SVM [13]	–	73.52	96.23	–	–
LBP + SVM	–	78.23	97.84	–	–
CNN [15]	546,578	96.87	99.73	98.21	100.00
CNN [16]	711,652	97.55	99.64	98.64	100.00
CNN [17]	653,522	97.79	99.73	98.77	100.00
Ours	1,036,066	**98.83**	**99.73**	**99.34**	**100.00**

Note that reported results are the average performance of tenfold cross validation

The results shown in bold indicate the highest performance in that column

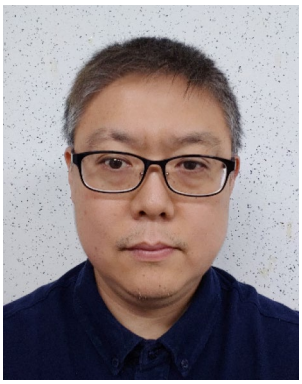
problem in the collection of the defective and the non-defective samples was successfully solved using a class-specific sampling method. The proposed system shows a remarkable performance in all experiments.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (No. 2016R1A2B4007608) and National IT Industry Promotion Agency (NIPA) Grant funded by the Korea government (MSIT) (No. S0602-17-1001) and partly supported by Cooperative R&D fund of Korea Ministry of SMEs and Startups (S2605414).

References

1. Chu, H. H., & Wang, Z. Y. (2017). A study on welding quality inspection system for shell-Tube heat exchanger based on machine vision. *International Journal of Precision Engineering and Manufacturing, 18*(6), 825–834.
2. Jia, H., Murphey, Y. L., Shi, J., & Chang, T. S. (2004). An intelligent real-time vision system for surface defect detection. In *Proceedings of the 17th international conference on pattern recognition* (Vol. 3, pp. 239–242).
3. Shen, H., Li, S., Gu, D., & Chang, H. (2012). Bearing defect inspection based on machine vision. *Measurement, 45*(4), 719–733.
4. Chu, H. H., & Wang, Z. Y. (2016). A vision-Based system for post-Welding quality measurement and defect detection. *The International Journal of Advanced Manufacturing Technology, 86*(9–12), 3007–3014.
5. Funck, J. W., Zhong, Y., Butler, D. A., Brunner, C. C., & Forrer, J. B. (2003). Image segmentation algorithms applied to wood defect detection. *Computers and Electronics in Agriculture, 41*(1–3), 157–179.
6. Yang, W., Li, D., Zhu, L., Kang, Y., & Li, F. (2009). A new approach for image processing in foreign fiber detection. *Computers and Electronics in Agriculture, 68*(1), 68–77.
7. Min, H. G., Kang, D. J., Kim, K. J., & Park, J. H. (2017). New non-contact measurement method of deformation at tensile test of thin film via digital image correlation technique. *International Journal of Precision Engineering and Manufacturing, 18*(11), 1509–1517.
8. Kwon, B. K., Won, J. S., & Kang, D. J. (2015). Fast defect detection for various types of surfaces using random forest with VOV features. *International Journal of Precision Engineering and Manufacturing, 16*(5), 965–970.
9. Ngan, H. Y., Pang, G. K., Yung, S. P., & Ng, M. K. (2005). Wavelet based methods on patterned fabric defect detection. *Pattern Recognition, 38*(4), 559–576.
10. Kumar, A., & Pang, G. K. (2002). Defect detection in textured materials using Gabor filters. *IEEE Transactions on Industry Applications, 38*(2), 425–440.
11. Tsai, D. M., & Lai, S. C. (2008). Defect detection in periodically patterned surfaces using independent component analysis. *Pattern Recognition, 41*(9), 2812–2832.
12. Zhou, W., Fei, M., Zhou, H., & Li, K. (2014). A sparse representation based fast detection method for surface defect detection of bottle caps. *Neurocomputing, 123,* 406–414.
13. Shumin, D., Zhoufeng, L., & Chunlei, L. (2011). AdaBoost learning for fabric defect detection based on HOG and SVM. In *2011 International conference on multimedia technology* (pp. 2903–2906).
14. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
15. Park, J. K., Kwon, B. K., Park, J. H., & Kang, D. J. (2016). Machine learning-based imaging system for surface defect inspection. *International Journal of Precision Engineering and Manufacturing-Green Technology, 3*(3), 303–310.
16. Masci, J., Meier, U., Ciresan, D., Schmidhuber, J., & Fricout, G. (2012). Steel defect classification with max-pooling convolutional neural networks. In *2012 International joint conference on neural networks (IJCNN)* (pp. 1–6).
17. Faghih-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R., & De Schutter, B. (2016). Deep convolutional neural networks for detection of rail surface defects. In 2016 *International joint conference on neural networks (IJCNN)* (pp. 2584–2589).
18. Soukup, D., & Huber-Mörk, R. (2014). Convolutional neural networks for steel surface defect detection from photometric stereo images. In *Advances in visual computing lecture notes in computer science* (pp. 668–677).
19. Wu, X., Cao, K., & Gu, X. (2017). A surface defect detection based on convolutional neural network. In *Lecture notes in computer science computer vision systems* (pp. 185–194).
20. Khumaidi, A., Yuniarno, E. M., & Purnomo, M. H. (2017). Welding defect classification based on convolution neural network (CNN) and Gaussian kernel. In *2017 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya* (pp. 261–265).
21. Park, J. K., & Kang, D. J. (2018). Unified convolutional neural network for direct facial keypoints detection. *The Visual Computer*. https://doi.org/10.1007/s00371-018-1561-3.
22. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning* (ICML-10) (pp. 807–814).
23. Kimme, C., Ballard, D., & Sklansky, J. (1975). Finding circles by an array of accumulators. *Communications of the ACM, 18*(2), 120–122.

Je-Kang Park received his BS degree in the school of mechanical engineering from Pusan National University, Busan, South Korea. He is currently a Ph.D. student in the same graduate school. His current research interests are computer vision, machine learning, deep neural network, and object detection.



Woo-Hyun An received his BS degree in the school of mechanical engineering from Yeungnam University. He is currently the manager of the Transmission Manufacturing Engineering Team 1 at the Hyundai Motor Company. His current research interests are Machine Learning and Smart Factory.

Dong-Joong Kang received a BS in precision engineering from Pusan National University in 1988 and a PhD in automation and design engineering at KAIST (Korea Advanced Institute of Science and Technology) in 1998. From 2004 to 2005, he was postdoc research at Cornell university and from 1997 to 1999, he was a research engineer at Samsung Advanced Institute of Technology (SAIT). He has been a professor at the School of Mechanical Engineering at Pusan National University since 2006 and associate editor of the International Journal of Control, Automation, and Systems since 2007. His current research interests include visual surveillance, intelligent vehicles/robotics, machine vision, and deep learning.