IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 June 2023, accepted 26 June 2023, date of publication 28 June 2023, date of current version 6 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3290914

RESEARCH ARTICLE

Robust Inspection of Integrated Circuit Substrates Based on Twin Network With Image Transform and Suppression Modules

EUNJEONG CHOI[1] AND JEONGTAE KIM[1,2]
[1]Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul 03760, South Korea
[2]Graduate Program in Smart Factory, Ewha Womans University, Seoul 03760, South Korea

Corresponding author: Jeongtae Kim (jtkim@ewha.ac.kr)

This work was supported by the Advanced Technology Incorporated (ATI), Incheon, Republic of Korea; and in part by the National Research Foundation (NRF), South Korea, through the Project Brain Korea 21 Fostering Outstanding Universities for Research (BK21 FOUR).

ABSTRACT Because existing IC substrate inspection methods do not utilize information in the design file, those are prone to failing detection of critical defects such as missing patterns. To remedy the problem, we propose a novel twin network-based inspection system for integrated circuit (IC) substrates that compares the design file (*i.e.,* a Gerber image) with a test image to be inspected. The proposed method is composed of an image transform module and an image comparison block. The image transform module transforms a Gerber image into an image that has similar characteristics to the test image. Without the transform module, many false positives may occur because the characteristics of the Gerber and test images such as noise, color, and pattern thickness are different. To compare the transformed Gerber image with the test image, we propose a twin network-based image comparison block with a feature suppression module that suppresses features from regions where defects do not exist while emphasizing features from defective regions. We confirmed the performance of the proposed method in comparison with existing methods using a real-world IC substrate dataset. Within the experiments, the proposed method achieved significantly improved performance from the existing inspection methods.

INDEX TERMS Deep learning, defect detection, packaging, integrated circuit substrate, transform module, attention module, printed circuit board, reference comparison, siamese network, twin network.

I. INTRODUCTION

Integrated circuit (IC) substrates are a type of printed circuit board (PCB) that are used extensively in packaging bare IC chips [1]. The IC substrate connects the bare chip to the main board, in addition to protecting the chip from physical damage [2]. As the quality of IC substrates can directly affect the performance of final products, inspecting them is an essential step in the semiconductor manufacturing process.

Inspecting the quality of IC substrates can be conducted by applying defect detection methods that have been studied for wafer or PCB inspection [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. These methods include non-referential approaches [21], [22] and referential approaches [22], [23], [24], [25], [26], [27]. The non-referential approaches determine whether a test image is defective without using any reference image [22]. Although these approaches benefit from not requiring cumbersome engineer interventions and image registration processes for using a reference image, they might not work

FIGURE 1. Illustrations of Gerber and test images with differences in characteristics: (a) and (c) Gerber images, (b) and (d) test images.

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Huang.

well when detecting defects that are not pre-defined [22]. In comparison, the referential approaches detect defects by comparing a test image with a defect-free reference image [22]. Although these approaches require image calibration processes, they can detect any type of defect by using the difference between two images [22]. We think that the goal of quality inspection is to detect all defects even if they are not pre-defined. Accordingly, we focus more on the reference comparison approaches in this paper.

Since the wafer and PCB typically contain multiple copies of the patterned images, many referential approaches generate a reference image that has similar characteristics to a test image by using images adjacent to a test image [28]. However, the reference image obtained using the aforementioned method could include some defects, which may cause degradation of inspection performance [20]. Additionally, if the same pattern missing which is a critical defect occurs in both the reference and test images, it may be difficult to detect such errors using only existing referential methods. Moreover, there are situations in which it is sometimes difficult to acquire a reference image.

To address the aforementioned problems, some studies generated a defect-free reference image using the design file for the wafer or PCB fabrication and compared the generated image with the test image to detect defects by using image processing techniques [29], [30], [31]. By using the design file as a reference image, it is able to detect any type of defect which differs from the intended design. However, there are differences between the characteristics of the test and generated images (*e.g.,* pattern thickness, noise, and color), which cause many false positives. Fig. 1 illustrates generated reference and test images with the differences in characteristics. We think that a referential inspection method robust to these differences is desirable for inspecting IC substrates because most design files for the substrates are available. As a matter of fact, investigating a novel robust referential inspection method using the design file is the primary motivation of this study. In this paper, we refer to the generated reference image using the design file for the IC substrate as the Gerber image.

Herein, we propose a novel deep learning-based referential inspection method that can address the performance degradation caused by differences in characteristics between the Gerber and test images. To minimize the effect of these differences, we present an image transform module that transforms the Gerber image into an image that has similar characteristics to the test image. The concept behind the image transform module is that if various filters for blur, morphological operations, or colorization are applied to the Gerber image, the module can reduce errors due to thickness and color differences between the designed and fabricated IC substrates. Note that the optimal filters for the image transform module are determined during training. In addition to the transform module, we propose an image comparison block based on an encoder-decoder twin network (also known as a Siamese network) to effectively compare the test and transformed Gerber images. The twin network consists of two encoders that share weights and considers the differences between the feature maps extracted from the encoders to conduct the image comparison [32]. Moreover, to further improve the performance of the image comparison block, we add a feature suppression module that can emphasize defective regions and suppress regions unnecessarily highlighted by differences in characteristics between the transformed Gerber and test images. We expect that the suppression module can reduce false positives from the differences in characteristics without compromising the performance when detecting defects.

The main contributions of this study can be summarized as follows:

- To the best of our knowledge, the proposed method is the first attempt to utilize the design file for inspecting IC substrates. By doing that, it is possible to detect any type of defect which differs from the design. As a result, the proposed method is effective in detecting fatal defects such as missing patterns which are extremely difficult to detect.
- We proposed an autoencoder-based transform module to reduce characteristic differences between the design file and fabricated substrates. Thanks to the transform module, it is possible to reduce false positives caused by differences in characteristics such as pattern thickness, color, and noise.
- We confirmed the performance of the proposed method using a real-world IC substrate dataset. In the experiments, the proposed method outperforms the existing inspection methods.

The remainder of this paper is organized as follows. In Section II, we review related studies. We also explain the proposed method in detail in Section III. The experimental results, conclusions, and discussion are presented in Sections IV, V, and VI, respectively.

II. RELATED WORK

As deep learning has achieved promising performance in computer vision tasks [33], [34], [35], [36], deep learning-based PCB/wafer inspection methods have been actively studied [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. These inspection methods can be divided into two categories of approaches: non-referential and referential. Deep learning-based non-referential inspection methods are based on object detectors or segmentation networks and use only a test image to identify defects. For example, Adibhatla et al. applied You Only Loock Once (YOLO) [34], which is a one-stage object detector, to classify defect types and localize defects [12], [13]. Similarly, Hu et al. [37] and Fan et al. [10] applied faster regions with convolutional neural networks (Faster R-CNN) [35] that is a two-stage object detector. Khalilian et al. [14] and Kim et al. [15] trained a convolutional autoencoder using defective images and forced it to repair defective parts. Subsquently, they detected defects by

subtracting the repaired output from the input. Ranjan et al. proposed a deep convolutional autoencoder-based model for defect segmentation [38]. However, although these deep learning-based non-referential inspection methods exhibited impressive performance, we think that reference comparison methods can be more effective in improving inspection performance because defects are defined as different parts from defect-free regions.

Recently, several investigations have been conducted on deep learning-based referential inspection methods [16], [17], [18], [19], most of which are based on a twin network. For example, Miao et al. [16] and Haddad et al. [18] classified images inspected from image processing-based defect detection methods into either defects or pseudo-defects (*i.e.,* false positives) using a twin network. Ling et al. proposed an encoder-decoder-based twin network with skip connections that transfer correlation maps between feature maps extracted from encoders [17]. In addition, Choi et al. applied Bayesian learning to a twin network to improve the inspection performance of wafers [19]. However, although these inspection methods exhibited impressive results, none of them utilized the design file that is effective in detecting critical defects such as pattern missing and in generating a defect-free reference image. Only a few studies attempted to investigate deep learning-based reference comparison methods using the design file for wafer inspection [20]. Kim et al. proposed a convolutional autoencoder-based wafer inspection method that uses a layout image and a wafer image combined in the channel direction as the input [20]. However, this method did not consider differences in characteristics between the generated reference image and the test image, which could cause many false positives.

The proposed method differs from existing methods in the following aspects. Firstly, the proposed method utilizes information from the design file (*i.e.,* a Gerber image) which has never been used in the existing methods for inspecting IC substrates. By doing that, the proposed method is able to detect critical defects such as missing patterns. Secondly, the proposed method prevents many false positives from happening by introducing an image transform module that reduces characteristic differences between a Gerber image and a test image such as pattern thickness, color, and noise.

III. PROPOSED METHOD

We propose a deep learning-based reference comparison method for IC substrate inspection using a Gerber image as a reference image. Fig. 2 illustrates the structure of the proposed network. As depicted in the figure, for IC substrate inspection, the network first applies the image transform module to a Gerber image and obtains the transformed Gerber image that has similar characteristics to a test image. Subsequently, the image comparison block based on a twin network compares the test and transformed images. Note that a feature suppression module is used when comparing the two images in the image comparison block.

A. IMAGE TRANSFORM MODULE

Fig. 3 provides an overview of the transform module, which learns the optimal filters for mapping a Gerber image to a defect-free test image. As shown in Fig. 3, the proposed module includes atrous convolution blocks to control the field of view of filters without compromising the spatial resolution [39]. Atrous convolution is well-known as a powerful tool that can capture contextual information at multiple scales without compromising spatial resolution [39].

Considering two-dimensional signals, the output $y[i, j]$ of the atrous convolution of input signal $x[i, j]$ with filter w of size $M \times N$ is defined as follows [39]:

$$y[i,j] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x[i + r \cdot m, j + r \cdot n] w[m, n], \qquad (1)$$

where rate r determines the stride with which we sample the input signal. Note that standard convolution is a special case in which $r = 1$. We applied several parallel atrous convolutions with different rates to extract multi-scale features that reflect local and global contextual information. Then, we combined the extracted features using a 1×1 convolution and used the combined features as the decoder input to obtain the transformed Gerber image.

To encourage a transformed image to be more similar to a defect-free test image, we added the following mean square error (MSE) loss function:

$$L_{trans} = \frac{1}{H} \frac{1}{W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} (1 - t_{ij})(x_{ij} - \hat{x}_{ij})^2, \qquad (2)$$

where H and W respectively denote the height and width of the input, and x_{ij}, $\hat{x}_{ij}$, and t_{ij} represent the target test image, transformed Gerber image, and label of the target test image at position (i, j), respectively. To prevent the transformed Gerber image from becoming defective, we utilized the labels of the target test image. Moreover, we set the label to 1 in the defective region and the label to 0 in the defect-free region. Therefore, if the target test image x_{ij} is defective, the loss term $(1 - t_{ij})(x_{ij} - \hat{x}_{ij})^2$ is zero, which implies only defect-free regions of the target test image affect the loss term.

B. IMAGE COMPARISON BLOCK

The twin network-based image comparison block comprises two sub-encoders that share weights, a decoder, and feature suppression modules. As shown in Fig. 2, the transformed Gerber and test images are the inputs to the image comparison block. Then, the proposed block extracts the feature maps from the two images in parallel using the same convolutional layers and then combines their feature maps to use the combined feature maps as the input of the decoder. In addition, the block computes the absolute difference of the feature maps between the encoder-convolutional layers and then transfers the difference map to the corresponding layer of the decoder in the form of skip connections [40]. We used these skip connections to recover spatial information that was

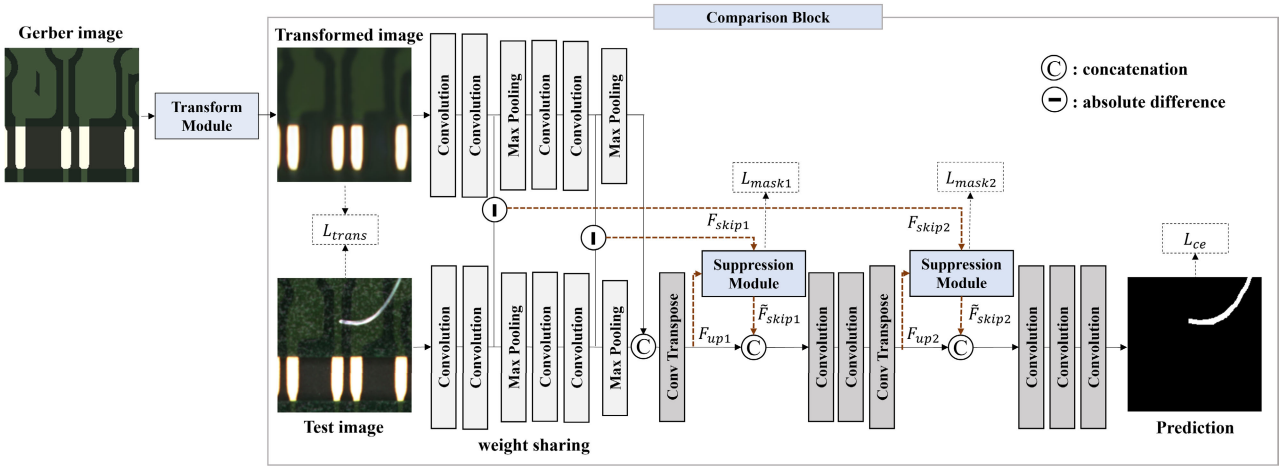

FIGURE 2. Overview of the proposed network.

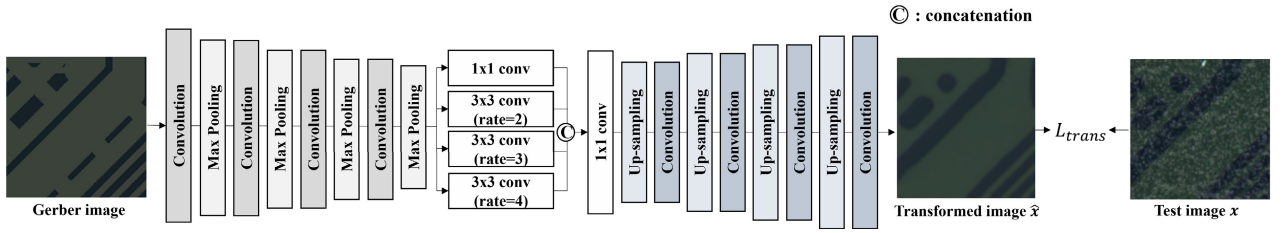

FIGURE 3. Overview of the proposed transform module.

lost during down-sampling and to make the network more focused on image comparison. However, although absolute difference skip connections can improve defect detection performance, they may cause false positives by transferring the errors from differences in characteristics to the decoder. Therefore, to maintain the advantages of the absolute difference skip connections and to address this problem, we applied a feature suppression module to the difference feature map of the skip connection to emphasize informative features and suppress any that are unnecessary, which will be explained in more detail later in the paper. Finally, the decoder determines whether each pixel belongs to the defective region using the combined feature maps from the sub-encoders and the output feature maps from the suppression modules.

We also used a weighted cross-entropy loss function to put more weight on the loss term for the misclassification of a defective pixel [41], [42]. The weighted cross entropy is defined as follows:

$$L_{\text{wce}} = -\frac{1}{D}\sum_{i}\sum_{j} w_1 t_{ij}\log p_{ij} + w_2(1-t_{ij})\log(1-p_{ij}), \tag{3}$$

where p_{ij} represents the prediction probability of defect for a pixel at position (i, j), D denotes the product of the height and width of the output, and w_1 and w_2 are the weights for defect and defect-free classes, respectively. In this study, we set w_1 to 0.6 and w_2 to 0.4.

C. FEATURE SUPPRESSION MODULE

Inspired by convolutional block attention module (CBAM) [43], we present a feature suppression module to further improve the performance of the image comparison block. Fig. 4 provides an overview of the suppression module, which consists of the channel and spatial attention modules. Given a feature map $F \in \mathbb{R}^{H\times W\times C}$, the channel attention module produces a channel attention mask $M_{\text{c}} \in \mathbb{R}^{1\times 1\times C}$ to focus on meaningful channels and suppress any that are unnecessary. Initially, the module gathers spatial information from the feature map using average- and max-pooling operations. Then, the average-pooled features $F^c_{\text{avg}} \in \mathbb{R}^{1\times 1\times C}$ and max-pooled features $F^c_{\text{max}} \in \mathbb{R}^{1\times 1\times C}$ are sequentially forwarded to a shared multi-layer perceptron (MLP) with one hidden layer. Next, the module merges the output feature vectors using element-wise summation. Finally, the channel attention mask $M_{\text{c}} \in \mathbb{R}^{1\times 1\times C}$ is computed as follows:

$$M_{\text{c}}(F) = \sigma(W_1(W_0(F^c_{\text{avg}})) + W_1(W_0(F^c_{\text{max}}))), \tag{4}$$

where σ is the sigmoid function and terms $W_0 \in \mathbb{R}^{L\times C}$ and $W_1 \in \mathbb{R}^{C\times L}$ denote the weights of the MLP. The weights of the MLP are shared for both inputs (*i.e.*, F^c_{avg} and F^c_{max}) and the rectified linear unit activation function is followed by W_0.

The spatial attention module operates in a similar way, by generating a spatial attention mask to emphasize informative regions and to suppress irrelevant ones. First, the module applies average- and max-pooling operations in the channel axis and then concatenates average-pooled features

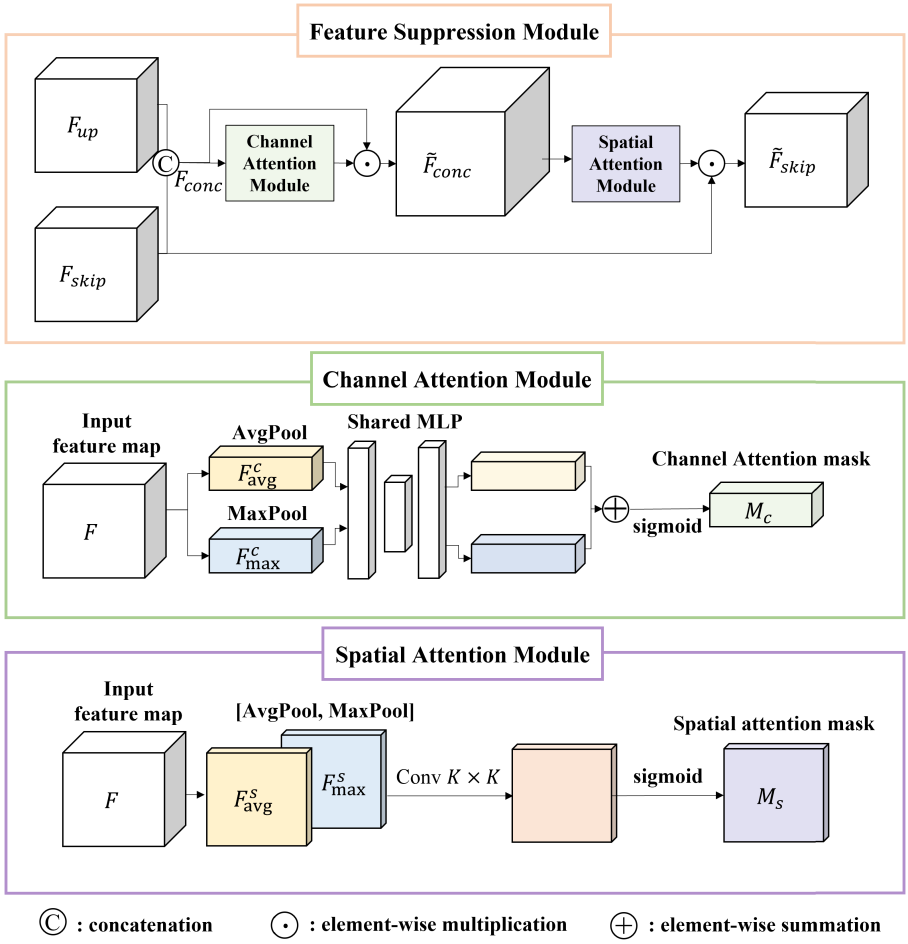

FIGURE 4. Overview of the proposed suppression module.

$F^s_{\text{avg}} \in \mathbb{R}^{H\times W\times 1}$ and max-pooled features $F^s_{\text{max}} \in \mathbb{R}^{H\times W\times 1}$ to compute the spatial mask $M_s \in \mathbb{R}^{H\times W\times 1}$, as follows:

$$M_s(F) = \sigma(f^{K\times K}([F^s_{\text{avg}}; F^s_{\text{max}}]), \tag{5}$$

where [;] denotes the concatenation operation and $f^{K\times K}$ represents a convolution operation with a filter size $K \times K$.

The main goal of a feature suppression module is to emphasize defective regions and to suppress differences in characteristics highlighted by the absolute difference skip connection using the spatial mask. In detail, the suppression module uses a feature map from the up-sampling layer of the decoder (*i.e.,* F_{up}) and an absolute difference feature map from the skip connection (*i.e.,* F_{skip}) as the input. Then, the module obtains a channel attention mask by applying a channel attention module to the input and generates the refined feature map $\tilde{F}_{conc}$ through element-wise multiplication of the input and channel attention mask. Subsequently, the suppression module applies the spatial attention module to the refined feature map and produces the spatial attention mask. Finally, by multiplying the spatial attention mask by the absolute difference feature map, the suppression module can reduce the influence of any differences in characteristics. The overall process can be summarized as follows:

$$\begin{aligned}\tilde{F}_{conc} &= M_c(F_{conc}) \odot F_{conc} \in \mathbb{R}^{H\times W\times 2C},\\ \tilde{F}_{skip} &= M_s(\tilde{F}_{conc}) \odot F_{skip} \in \mathbb{R}^{H\times W\times C},\end{aligned} \tag{6}$$

where F_{conc} is the concatenated feature map of F_{up} and F_{skip}, $\odot$ denotes element-wise multiplication, and $\tilde{F}_{skip}$ denotes the suppressed absolute difference feature map using the spatial attention mask.

We also added the MSE loss function to encourage the spatial mask to be more similar to the ground truth, as follows:

$$L_{mask} = \frac{1}{H}\frac{1}{W}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}(t_{ij} - s_{ij})^2, \tag{7}$$

where s_{ij} is the spatial mask at position (i, j). We believe that if the spatial mask is similar to the labels of the test image, it can be more effective in emphasizing the defective regions and suppressing regions that are irrelevant to the defects.

D. TOTAL LOSS FUNCTION

We applied feature suppression modules to the first and second convolutional blocks of the decoder, as shown in Fig. 2.

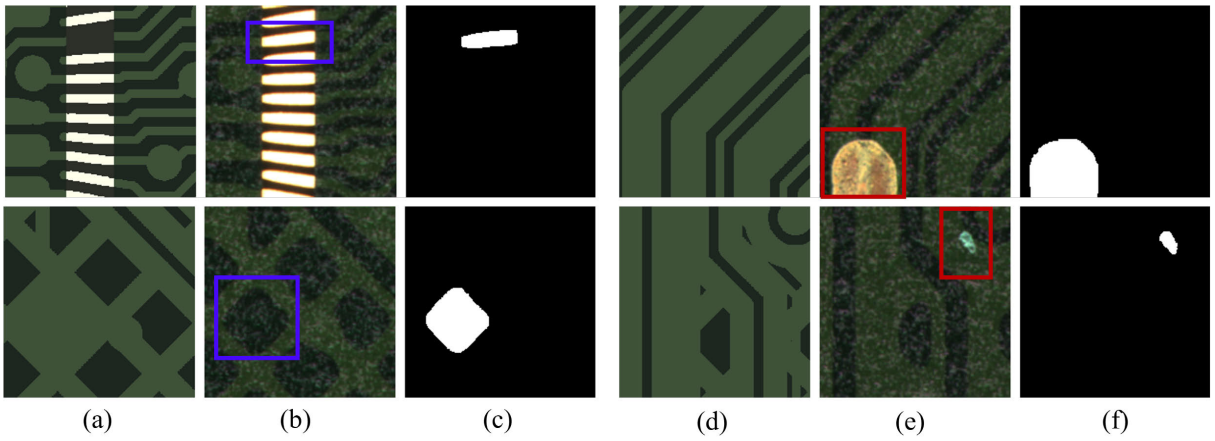

FIGURE 5. Pair-set examples: (a) and (d) Gerber images, (b) and (e) test images, and (c) and (f) ground truth. The blue boxes represent synthetic defects with missing patterns and the red boxes indicate real defects.

Therefore, L_{mask1} and L_{mask2} were added to the total loss function, which is defined as follows:

$$L_{total} = \lambda_1 L_{trans} + \lambda_2 L_{mask1} + \lambda_3 L_{mask2} + \lambda_4 L_{wce}, \quad (8)$$

where λ_1, λ_2, λ_3, and λ_4 are the weights for balancing the four losses. In this paper, we set λ_1, λ_2, λ_3, and λ_4 to 0.4, 0.15, 0.15, and 0.3, respectively.

IV. EXPERIMENT

The proposed method was implemented as follows: the image transform module and the image comparison block were trained simultaneously using the weights of the pre-trained transform module as the initial weights. In this paper, we call the proposed method Gerber-Twin. To verify the effectiveness of the Gerber-Twin method, we compared its performance with that of existing methods. These existing methods included non-referential inspection method (*i.e.,* a convolutional autoencoder using a single input [38]) and referential inspection methods (*i.e.,* a convolutional autoencoder using a combined input in the channel direction [20] and a twin network [17]). We refer to these methods as Single-Net (a convolutional autoencoder using a single input), Concat-Net (a convolutional autoencoder using the combined input), and Base-Twin (a twin network).

A. DATASET

We tested the performance of the proposed method by comparing it with other methods using a real-world IC substrate dataset that contains repetitive substrate images of which sizes are $2,879 \times 3,238$. The dataset was acquired using a 12k line scan camera with RGB LED lighting, 48mm field of view, and 4.0μm per pixel resolution in a vacuumed stage.

For the experiments, we used 190 substrate images and generated a Gerber image using the GerbView program that converts the design file into an image. Although the original Gerber image was grayscale, we manually changed the color of the Gerber image to be somewhat similar to that of the substrate images for a fairer comparison. Fig. 6 illustrates the original Gerber, colored Gerber, and substrate images.

FIGURE 6. Grayscale and color Gerber images: (a) grayscale Gerber image (b) colored Gerber image, and (c) real substrate image.

To generate the training and test sets, we first cropped a full-size IC substrate image into 200×200 image patches and then found image patches in a full-size Gerber image that were similar to the cropped patches using a template-matching technique. We then used image pairs composed of the cropped patches and the matched Gerber patches for training and testing. In addition, we applied rotation (90°, 180°, and 270°) and flip augmentation for defective image pairs in the training set to alleviate the problem of imbalanced dataset.

Moreover, we added synthesized defective images with missing patterns to verify the usefulness of the reference comparison methods in terms of detecting defects. Although defects with missing patterns are not frequent in the manufacturing process, they must be detected because they are fatal defects. To generate synthetic images with missing patterns, we deleted patterns in the Gerber image because removing some parts in the substrate images could have resulted in artifacts. Although one may think that they are defects with patterns added, we call them defects with missing patterns in this paper.

Finally, we collected 17,640 pair-set images (4,488 defective image pairs and 13,152 defect-free image pairs) for training and 10,671 pair-set images (150 defective image pairs and 10,521 defect-free image pairs) for testing. Fig. 5 displays pair-set examples used for experiments, where the red and blue boxes indicate real and synthetic defects, respectively.

TABLE 1. Training options for each method.

	batch	lr	ep	d_rt	d_ep	pt
Single-Net	16	1×10^{-4}	200	1	0	15
Concat-Net	16	1×10^{-4}	100	0.1	30	15
Base-Twin	16	1×10^{-4}	100	1	0	15
Gerber-Twin	16	1×10^{-4}	200	0.1	30	15

TABLE 2. Experimental results for IC substrate dataset.

Method	R (%)	P (%)	F (%)
Single-Net	83.96	61.57	71.04
Concat-Net	89.30	65.75	75.74
Base-Twin	89.30	72.29	79.90
Gerber-Twin	90.91	93.41	92.14

B. IMPLEMENTATION DETAILS

The training options for each method are listed in Table 1; batch, lr, ep, d_rt, d_ep, and pt represent the mini-batch size, initial learning rate, maximum number of epoch, decay rate, decay epoch number, and patience for early stopping, respectively. In Table 1, we reported d_ep as 0 when a fixed learning rate was used for training (*i*.e., d_rt = 1). In addition, we trained all methods using the Adam optimizer on an NVIDIA TITAN Xp graphics card (Nvidia Corporation, USA) and used the tensorflow2 library [44].

C. EXPERIMENTAL RESULTS

We trained and tested all methods three times, and report the average recall, precision, and f1-score of each method in Table 2. As presented in Table 2, the proposed method demonstrates the best performance compared with other inspection methods in the sense that it exhibits the highest recall, precision, and f1-score. The recall, precision, and f1-score of the proposed Gerber-Twin method are approximately 1.61%, 21.12%, and 12.24%, respectively higher than those of the Base-Twin method, which exhibits the second-best performance. We believe that one of the main reasons for this significant improvement is that the image transform module in the proposed method effectively reduced the characteristic differences between the Gerber and test images. In addition, the feature suppression module in the proposed method also contributes to reducing false positives by suppressing regions of feature maps where no defect exists.

Figs. 7 (a), (b), (c), and (d) show the Gerber image, transformed Gerber image, test image, and ground truth of five different regions of an IC substrate, respectively, and Figs. 7 (e), (f), (g), and (h) present corresponding inspection result images of Single-Net, Concat-Net, Base-Twin, and Gerber-Twin, respectively. Note that the Single-Net method implements the inspection using only test images while the Concat-Net, Base-Twin, and the proposed Gerber-Twin methods use both Gerber and test images to detect defective regions.

As shown in the second and third rows of Fig. 7, the proposed Gerber-Twin method was able to detect fatal defects of missing patterns without generating many false positives. On the contrary, the Single-Net method was not able to detect the critical defects because it is difficult to know that a pattern is missing without the aid of the design file. Although the Concat-Net and Base-Twin methods detected the missing pattern defects since the methods also used the Gerber image, those methods generated many false positives. This can be further illustrated in the fourth and fifth rows of Fig. 7. As shown in the figures, both Concat-Net and Base-Twin methods produced many false positives because the characteristics of the Gerber images and test images were different. In contrast, the proposed Gerber-Twin method showed significantly improved results thanks to the image transform and feature suppression modules.

To further verify the usefulness of the image transform module, we compared the absolute difference of the input images (*i.e.,* the Gerber and test images) with the absolute difference between the transformed Gerber and test images. If the image transform module effectively managed to reduce errors resulting from characteristic differences between input images, the absolute differences between the transformed Gerber and test images would have smaller values than the absolute differences of the input images. Fig. 8 presents the absolute difference maps averaged in the channel axis to aid visualization. Figs. 8 (a), (b), and (c) display the Gerber images, test images, and ground truth, respectively, and Figs. 8 (d), (e), and (f) present the transformed Gerber images, the absolute difference between input images, and the absolute difference between the transformed Gerber and test images, respectively. As shown in Figs. 8 (e) and (f), the absolute differences between the transformed Gerber and test images had smaller values than the absolute differences between the input images, which was consistent with our expectations. From these results, we confirmed that the image transform module was useful for improving precision.

We also verified whether the spatial attention mask obtained from the feature suppression module was similar to the ground truth. If true, this module could effectively emphasize the defective regions and suppress any unnecessary regions highlighted by differences in characteristics. Figs. 9 (a), (b), and (c) illustrate Gerber, test, and ground truth images, respectively, and Figs. 9 (d), (e), and (f) present the spatial attention masks, absolute difference feature maps from the skip connection, and refined absolute difference feature maps using spatial attention masks, respectively. Figs. 9 (c) and (d) indicate that the spatial masks were similar to the ground truth. In addition, from Figs. 9 (e) and (f), we observed that the suppression module effectively suppresses regions irrelevant to the defects and emphasizes the defective regions, which could improve precision and recall.

The inference time per 200 $\times$ 200 image pair for each method was reported in Table 3. We measured the inference time using a PC with NVIDIA GeForce GTX 1070 except for data pre-processing steps such as template matching. As shown in Table 3, the inference time of the Gerber-Twin method was longer than that of other methods because more

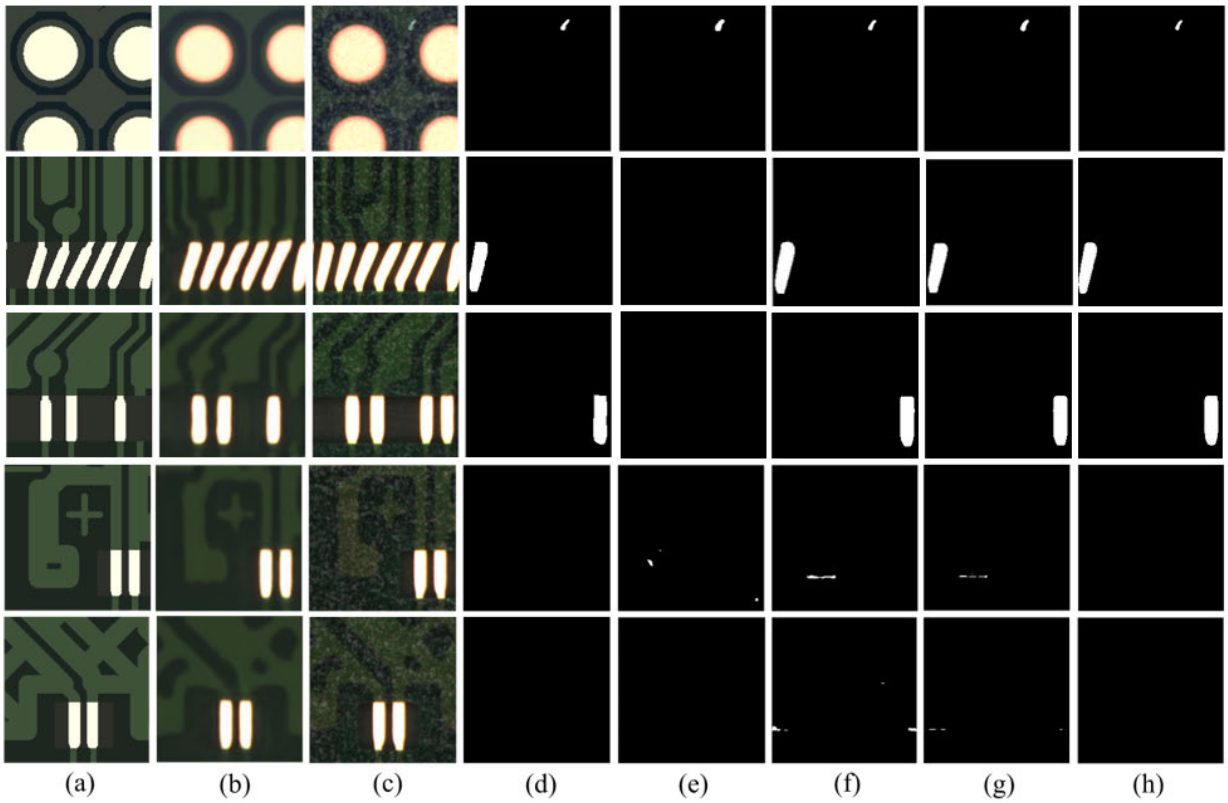

FIGURE 7. Gerber, test images and corresponding prediction result of each method: (a) Gerber images, (b) transformed Gerber images, (c) test images, (d) ground truth, (e) predictions using Single-Net, (f) predictions using Concat-Net, (g) predictions using Base-Twin, and (h) predictions using Gerber-Twin.

TABLE 3. Inference time for one image pair.

Method	Inference time (ms)
Single-Net	30.01
Concat-Net	30.07
Base-Twin	30.90
Gerber-Twin	44.93

TABLE 4. Experimental results for training dataset size.

Method	R (%)	P (%)	F (%)
50% of training data	89.30	82.27	85.64
70% of training data	90.37	84.50	87.34
100% of training data	90.91	93.41	92.14

computations were required for the image transform and feature suppression modules.

We conducted additional experiments using 50% and 70% of the training dataset to investigate how training dataset size affects the performance of the proposed method and report the experimental results in Table 4. As expected, we confirmed that increasing the size of the training dataset improved performance.

In this paper, we applied rotation and flip augmentation for defective images to alleviate the problem of imbalanced dataset. To investigate how imbalanced data affects

TABLE 5. Experimental results for imbalanced dataset.

Method	R (%)	P (%)	F (%)
W/O augmentation	89.30	67.61	76.96
W/ augmentation	90.91	93.41	92.14

TABLE 6. Ablation study of the proposed modules on IC substrate dataset.

Transform module	Suppression module	R (%)	P (%)	F (%)
		89.30	72.29	79.90
✓		90.91	86.73	88.77
	✓	90.91	82.93	86.74
✓	✓	90.91	93.41	92.14

inspection performance, we conducted additional experiments with and without augmentation and report the experimental results in Table 5. From the study, we confirmed that alleviating the problem of imbalanced dataset improved the inspection performance. One may improve the performance further using more balanced dataset (if more real defective data is available).

We conducted ablation study to verify the effectiveness of each module by comparing the Gerber-Twin method with three variants of the proposed network: no modules, only a transform module, and only a feature suppression module. The experimental results are reported in Table 6. As shown

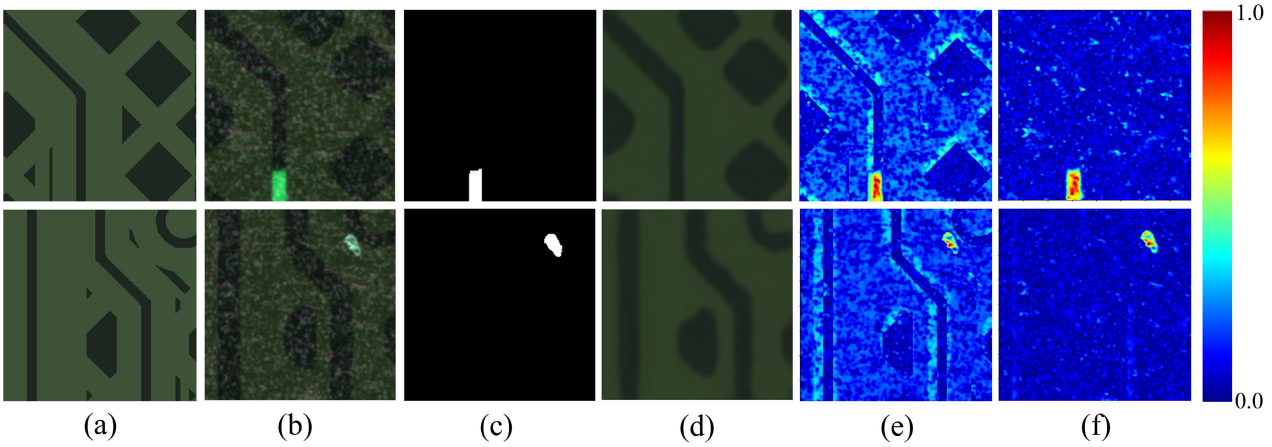

FIGURE 8. Illustrations of the results of the image transform module: (a) Gerber images, (b) test images, (c) ground truth, (d) transformed Gerber images, (e) absolute difference maps between Gerber and test images, and (f) absolute difference maps between transformed Gerber and test images.

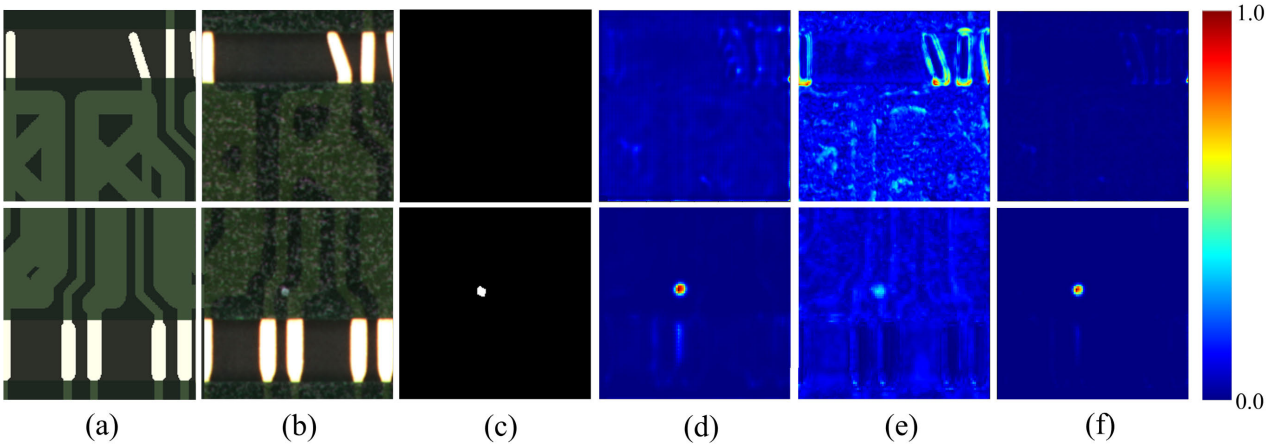

FIGURE 9. Illustrations of the results of the feature suppression module: (a) Gerber images, (b) test images, (c) ground truth, (d) spatial attention masks, (e) absolute difference feature maps from skip connection, and (e) refined absolute difference feature maps.

in Table 6, the proposed method using both modules achieved a significant performance improvement compared with other methods by exhibiting the highest recall, precision, and f1-score. Table 6 also indicates that both proposed modules contributed to reducing false positives effectively. These experimental results demonstrated the proposed network achieved improved robustness to differences in characteristics between Gerber and test images.

V. DISCUSSION

We investigated a novel twin network-based referential inspection system for IC substrate that compares a test image with the design file (Gerber image). Through intensive experiments using a real-world IC substrate dataset, we have confirmed that the proposed method exhibits significantly improved performance from existing methods. In addition, we verified the effectiveness of each module in the proposed method such as the image transform module and feature suppression module by ablation studies. We firmly believe that the proposed method can be effectively used for quality assurance during the manufacturing of IC substrates. Furthermore, the proposed method has potential applicability for many other inspection tasks. For example, it can be efficiently used for the inspection of products of which design files are available such as semiconductor wafers [20].

The inference time of the proposed method is longer than conventional methods. Since the acceptable inference time during a manufacturing process depends on the specification of a product line, it is difficult to mention whether the current inference time of the proposed method is fast enough for real-time processing. Nevertheless, it is for sure that faster computation is more desirable for real-time processing. We believe that the inference time of the proposed method can be reduced by storing pre-computed transformed images from a Gerber image. In addition, the scalability of the proposed method can be improved further with the advance of computing hardware as well as algorithms such as model compression [45].

We confirmed that the proposed method was able to detect defects of 2×2 pixel size. Because the IC substrate dataset

used in our experiments was acquired using a camera with 4 μm per pixel resolution, the smallest size of defect that was detected using the proposed method was approximately 64 μm^2. Since this limitation is not due to the proposed method but to the resolution of imaging devices, it is possible to detect smaller-size defects using the proposed method with finer-resolution images.

It is desirable to confirm the performance of the proposed method using a standard dataset. However, since a standard open dataset for IC substrate inspection using Gerber images does not exist, we were not able to do that. Instead, to confirm the performance of the proposed method thoroughly, we conducted various experiments such as using different size training datasets and using different amounts of imbalanced dataset, etc. We think that the effectiveness of the proposed method is verified through the experiments even though those were conducted without using *unavailable* standard dataset.

One may be concerned about the possibility of overfitting in training the proposed method if the amount of labeled data is not sufficient, which is frequently happening situation in practice. To remedy the concern, one may attempt to apply various augmentation strategies [46]. In addition, one may expand dataset without labeling using semi-supervised learning methods that utilize unlabeled data [47].

VI. CONCLUSION

In this paper, we propose a twin network-based referential inspection system for integrated circuit (IC) substrates that uses the design file (*i.e.,* a Gerber image) as a reference image. Compared with existing inspection methods, the proposed method is more robust to differences between the characteristics of the Gerber and test images, such as noise, color, and pattern thickness. One of the key components of the proposed method for reducing characteristic differences is an image transform module that transforms the Gerber image into an image that has similar characteristics to the test image. Another key element is a feature suppression module that emphasizes meaningful features and suppresses any unnecessary features using spatial attention masks. We conducted various experiments using a real-world IC substrate dataset and demonstrated that the proposed method is more robust to differences in characteristics between the Gerber and test images compared to the existing methods.

ACKNOWLEDGMENT

The authors are very grateful to Advanced Technology Incorporated (ATI) in Incheon, Republic of Korea for providing us with IC substrate data. This work was supported by grants from ATI and the National Research Foundation (NRF), Korea, under project Brain Korea 21 Fostering Outstanding Universities for Research (BK21 FOUR).

REFERENCES

[1] L. Wang, L. Zheng, C. Y. Wang, S. Li, Y. Song, L. Zhang, and P. Sun, "Experimental study on micro-drills wear during high speed of drilling IC substrate," *Circuit World*, vol. 40, no. 2, pp. 61–70, Apr. 2014.

[2] H. Kang and A. H. I. Lee, "A new supplier performance evaluation model: A case study of integrated circuit (IC) packaging companies," *Kybernetes*, vol. 39, no. 1, pp. 37–54, Mar. 2010.

[3] K. Imoto, T. Nakai, T. Ike, K. Haruki, and Y. Sato, "A CNN-based transfer learning method for defect classification in semiconductor manufacturing," in *Proc. Int. Symp. Semiconductor Manuf. (ISSM)*, Dec. 2018, pp. 1–3.

[4] M. Saqlain, Q. Abbas, and J. Y. Lee, "A deep convolutional neural network for wafer defect identification on an imbalanced dataset in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 436–444, Aug. 2020.

[5] M. Saqlain, B. Jargalsaikhan, and J. Y. Lee, "A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 171–182, May 2019.

[6] S. Cheon, H. Lee, C. O. Kim, and S. H. Lee, "Convolutional neural network for wafer surface defect classification and the detection of unknown defect class," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 163–170, May 2019.

[7] L. H. D. S. Silva, G. O. D. A. Azevedo, B. J. T. Fernandes, B. L. D. Bezerra, E. B. Lima, and S. C. Oliveira, "Automatic optical inspection for defective PCB detection using transfer learning," in *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, Nov. 2019, pp. 1–6.

[8] K. Shioiri and S. Hamao, "Application of machine learning to printed circuit board external inspection," Anritsu, Atsugi, Japan, Tech. Rep. 28, 2020, pp. 45–51.

[9] J. S. S. V. Mamidi, S. Sameer, and J. Bayana, "A light weight version of PCB defect detection system using YOLO V4 tiny," in *Proc. Int. Mobile Embedded Technol. Conf. (MECON)*, Mar. 2022, pp. 441–445.

[10] F. Fan, B. Wang, G. Zhu, and J. Wu, "Efficient faster R-CNN: Used in PCB solder joint defects and components detection," in *Proc. IEEE 4th Int. Conf. Comput. Commun. Eng. Technol. (CCET)*, Aug. 2021, pp. 1–5.

[11] Q. Zhang and H. Liu, "Multi-scale defect detection of printed circuit board based on feature pyramid network," in *Proc. IEEE Int. Conf. Artif. Intell. Comput. Appl. (ICAICA)*, Jun. 2021, pp. 911–914.

[12] V. A. Adibhatla, H.-C. Chih, C.-C. Hsu, J. Cheng, M. F. Abbod, and J.-S. Shieh, "Defect detection in printed circuit boards using you-only-look-Once convolutional neural networks," *Electronics*, vol. 9, no. 9, p. 1547, Sep. 2020.

[13] V. A. Adibhatla, H.-C. Chih, C.-C. Hsu, J. Cheng, M. F. Abbod, and J.-S. Shieh, "Applying deep learning to defect detection in printed circuit boards via a newest model of you-only-look-once," *Math. Biosci. Eng.*, vol. 18, no. 4, pp. 4411–4428, 2021.

[14] S. Khalilian, Y. Hallaj, A. Balouchestani, H. Karshenas, and A. Mohammadi, "PCB defect detection using denoising convolutional autoencoders," in *Proc. Int. Conf. Mach. Vis. Image Process. (MVIP)*, Feb. 2020, pp. 1–5.

[15] J. Kim, J. Ko, H. Choi, and H. Kim, "Printed circuit board defect detection using deep learning via a skip-connected convolutional autoencoder," *Sensors*, vol. 21, no. 15, p. 4968, Jul. 2021.

[16] Y. Miao, Z. Liu, X. Wu, and J. Gao, "Cost-sensitive Siamese network for PCB defect classification," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–13, Oct. 2021.

[17] Z. Ling, A. Zhang, D. Ma, Y. Shi, and H. Wen, "Deep Siamese semantic segmentation network for PCB welding defect detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[18] B. M. Haddad, S. F. Dodge, L. J. Karam, N. S. Patel, and M. W. Braun, "Locally adaptive statistical background modeling with deep learning-based false positive rejection for defect detection in semiconductor units," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 357–372, Aug. 2020.

[19] E. Choi and J. Kim, "Die-to-die inspection of semiconductor wafer using Bayesian twin network," *IEIE Trans. Smart Process. Comput.*, vol. 10, no. 5, pp. 382–389, Oct. 2021.

[20] J. Kim, Y. Nam, M. Kang, K. Kim, J. Hong, S. Lee, and D. Kim, "Adversarial defect detection in semiconductor manufacturing process," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 3, pp. 365–371, Aug. 2021.

[21] A. M. Darwish and A. K. Jain, "A rule based approach for visual pattern inspection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 1, pp. 56–68, Jan. 1988.

[22] M. Moganti, F. Ercal, C. H. Dagli, and S. Tsunekawa, "Automatic PCB inspection algorithms: A survey," *Comput. Vis. Image Understand.*, vol. 63, no. 2, pp. 287–313, Mar. 1996.

[23] I. Ibrahim, Z. Ibrahim, K. Khalil, M. M. Mokji, and S. Abu-Bakar, "An algorithm for classification of five types of defects on bare printed circuit board," *Int. J. Comput. Sci. Eng. Syst.*, vol. 5, no. 3, pp. 201–208, 2011.
[24] J. Ma, "Defect detection and recognition of bare PCB based on computer vision," in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 11023–11028.
[25] B.-J. Park and K.-S. Hahn, "A research of PCB pattern visual inspection system using cad data," in *Proc. Korean Inf. Sci. Soc. Conf.*, 2007, pp. 446–449.
[26] H. Liu, W. Zhou, Q. Kuang, L. Cao, and B. Gao, "Defect detection of IC wafer based on spectral subtraction," *IEEE Trans. Semiconductor Manuf.*, vol. 23, no. 1, pp. 141–147, Feb. 2010.
[27] M. H. Annaby, Y. M. Fouda, and M. A. Rushdi, "Improved normalized cross-correlation for defect detection in printed-circuit boards," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 199–211, May 2019.
[28] M. Zontak and I. Cohen, "Kernel-based detection of defects on semiconductor wafers," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Sep. 2009, pp. 1–6.
[29] A. Doudkin and A. Inyutin, "Computer-aided technique for defect and project rules inspection on PCB layout image," *Int. J. Comput.*, vol. 5, pp. 107–111, 2014.
[30] B.-J. Park and K.-S. Hahn, "Automated visual inspection system of pcb using cad information," *J. Korea Multimedia Soc.*, vol. 12, no. 3, pp. 397–408, 2009.
[31] S.-G. Youn, Y.-G. Kim, and T.-H. Park, "Wavelet transform based defect detection for PCB inspection machines," *Trans. Korean Inst. Electr. Eng.*, vol. 66, no. 10, pp. 1508–1515, 2017.
[32] E. Choi and J. Kim, "Robust change detection using channel-wise co-attention-based Siamese network with contrastive loss function," *IEEE Access*, vol. 10, pp. 45365–45374, 2022.
[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 91–99.
[36] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[37] B. Hu and J. Wang, "Detection of PCB surface defects with improved faster-RCNN and feature pyramid network," *IEEE Access*, vol. 8, pp. 108335–108345, 2020.
[38] N. Ranjan, S. Bhandari, Y. Kim, and H. Kim, "Polycrystalline silicon wafer scratch segmentation based on deep convolutional autoencoder," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Feb. 2022, pp. 1–4.
[39] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
[40] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 4063–4067.
[41] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.
[42] S. Panchapagesan, M. Sun, A. Khare, S. Matsoukas, A. Mandal, B. Hoffmeister, and S. Vitaladevuni, "Multi-task learning and weighted cross-entropy for DNN-based keyword spotting," in *Proc. Interspeech*, Sep. 2016, pp. 760–764.
[43] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
[44] M. Abadi, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.
[45] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 6, pp. 1789–1819, Jun. 2021.
[46] C. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-supervised learning for anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9659–9669.
[47] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1195–1204.

EUNJEONG CHOI received the B.S., M.S., and Ph.D. degrees in electronic and electrical engineering from Ewha Womans University, Seoul, South Korea, in 2016, 2018, and 2023, respectively. She is currently with Mechatronics Research of Samsung Electronics, South Korea. Her research interests include deep learning for machine vision and digital signal processing.

JEONGTAE KIM received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Seoul, South Korea, in 1989 and 1991, respectively, and the Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor, in 2004. From 1991 to 1998, he was with Samsung Electronics, South Korea, where he had been engaged in the development of digital camcorders and digital TVs. Since 2004, he has been with the Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul, where he is currently a Professor. His research interests include machine learning, computer vision, and radar signal processing.