이화여자대학교 대학원
2023학년도
석사학위 청구논문

A Combined Twin and Single Network for Fast and Robust Inspection of IC Substrates

전 자 전 기 공 학 과
윤 희 준
2024

A Combined Twin and Single Network for Fast and Robust Inspection of IC Substrates

이 논문을 석사학위 논문으로 제출함

2024 년 6 월

이화여자대학교 대학원

전 자 전 기 공 학 과 윤 희 준

윤 희 준 의 석사학위 논문을 인준함

지도교수 김 정 태 ____

심사위원 김 정 태 ____

강 제 원 ____

박 수 현 ____

이화여자대학교 대학원

Table of contents

List of Tables

List of Figures

Summary

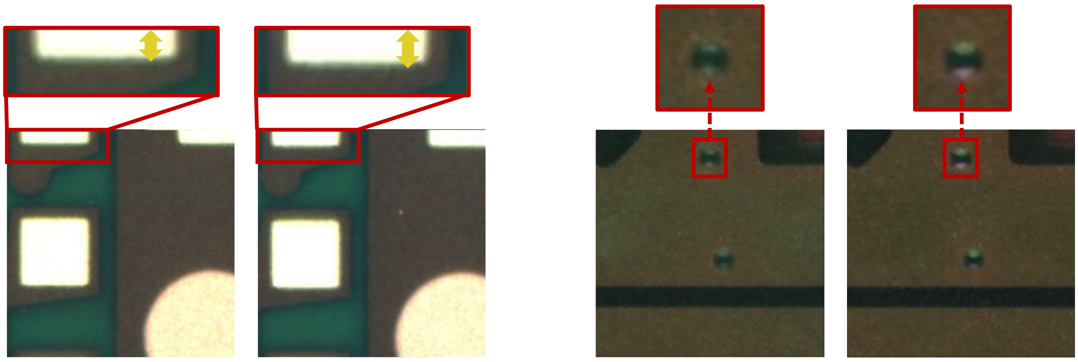
This thesis explores a novel approach to inspecting IC substrates by integrating twin and single network architectures for fast and accurate defect detection. For inspecting IC substrates, the repetitive nature of the substrates allows comparison-based inspection approaches that detect changed regions between test and reference images. One deep learning-based approach for change detection is the twin network structure, which processes two inputs using two identical encoders to extract features consistently, and then compares these features to identify changes. However, the twin network is vulnerable to errors caused by pseudo changes and mis-registration, which complicates the detection of micro defects crucial for accurate inspections. To address these challenges, this research proposes a hybrid approach of Combined Twin and Single Network (abbreviated by C-TSNet). We incorporate a single network that is not affected by misalignments and color variations with a twin network, the proposed network achieves robustness to mis-registration while maintaining low computational latency. Additionally, we introduce CC-TSNet, which is extended version of C-TSNet by adding a channel co-attention module to further improve robustness against characteristic differences. The proposed methods were validated on a real-world IC substrate dataset, showing significantly better performance compared to traditional deep-learning approaches such as U-Net and Base-Twin network. The integration of the channel co-attention module improved the system's ability to handle pseudo-changes due to characteristic differences, as resulting in higher F1-scores and reduced false positives. Moreover, the proposed methods achieved accuracy comparable to the module-based method but with only one-third of computational latency, making them highly suitable for real-time manufacturing processes. Overall, this thesis presents a comprehensive solution for IC substrate inspection, combining the strengths of both twin and single network architecture to deliver fast, accurate, and robust defect detection.

I. Introduction

Deep-learning based inspection methods have gained significant prominence in recent years for quality control in manufacturing process. During the early days of manufacturing, product inspection in manufacturing was primarily a manual task performed by human inspectors [1]. However, manual inspection has limitations in terms of precision, speed and consistency among different inspectors [2]. Consequently, the adoption of automated defect detection using deep learning networks in manufacturing has become essential to meet the increasing demands for high product quality [3].

Particularly in the semiconductor industry, inspecting products such as integrated circuit (IC) chips or wafers is crucial for ensuring product quality and detecting micro defects that could impact performance of the final products [4]. Defect inspection techniques for semiconductor products can be broadly classified into non-referential and referential methods. Non-referential deep-learning approaches, such as segmentation [5-8] or object detection networks [9-15], learn the characteristics of defects during the training process and inference test images without reference. However, these single-image methods face limitations in identifying defects that are not present in the training datasets, leading to potential errors [3]. Conversely, referential inspection methods detect changes by comparing test images with reference images. This approach is more adaptable for identifying defects in complex patterns, even with relatively few samples [16]. Leveraging these advantages, change detection (CD) techniques can be effectively applied in semiconductor manufacturing by utilizing the repetitive substrates of products [17].

Change detection is defined as the task recognizing differences between two images captured at different times or under varying conditions [18]. This technique acts a crucial role in a variety of applications, including remote sensing [19-26], medical imaging [27-29], and industrial inspection [17, 30, 31]. For CD tasks, twin network (also called as a Siamese network [32]), consists of two weighted-shared subnetworks that process two input images and compare their outputs, is widely used for various applications [17, 19, 20, 22, 24, 29, 31, 33-35]. However, twin networks may face significant challenges such mis-registration and pseudo-changes due to different lightning conditions and unwilling changes while processing images [17, 34]. These examples of challenged paired images are illustrated in Figure 1, which are examples of the IC substrate dataset with misalignments and color changes due to variation in illumination.



(a) (b)

Figure 1. Example of IC substrate dataset with pseudo-changes: (a) paired images with misregistered and (b) paired images with characteristic differences

To address these problems, several CD methods incorporated attention modules [17, 19, 20, 22, 34], primarily focusing on remote sensing applications. In the context of semiconductor inspection, a few studies have been conducted. Unlike remote sensing applications, which are less vulnerable to certain issues, semiconductor inspections are crucial for detecting micro defects and therefore demand higher accuracy and precision. Choi *et al*. [17] tackled issues arising from misalignment and characteristic differences by utilizing a co-attention module that considers both spatial and channel dimensions. Despite its effectiveness, this module requires extra computation time for calculating correlations between feature vectors, which could be a notable drawback for real-world applications [36].

We focused on the fact that although attention mechanisms can address the issues raised in twin networks, they significantly increase computation time, especially when correcting registration errors [17]. This prompted us to consider alternative strategies to resolve the issues by leveraging the benefits of both referential and non-referential methods. Considering the challenges, we propose the Combined Twin and Single Network (C-TSNet), which offers fast and robust inspection of IC substrates. The single network can avoid the drawbacks of the twin network such as pseudo-changes and alignment errors, thus this integrating can be an alternative to spatial attention modules. By combining these two architectures, C-TSNet aims to overcome the limitations of each individual approach, providing a more robust solution without extra module. Additionally, we introduce Channel co-attention attached C-TSNet (abbreviated by CC-TSNet), an extension of C-TSNet that utilizes a channel co-attention module [34] to enhance robustness to characteristic differences. We validate the proposed methods on a real-world IC substrate dataset. In the experiments, we present both qualitative and visualization results showing that our proposed methods achieved competitive performance compared to existing methods while ensuring computational efficiency.

II. Related works

In this section, we provide a comprehensive review of the previous literatures relevant to our study. We begin with a discussion on inspection methods using deep learning. And we explain about twin networks, followed by an exploration of change detection techniques utilizing twin networks, and examination of twin network-based inspection methods.

A. Deep learning-based inspection

With the advancement of deep learning-based computer vision technologies, various techniques such as classification [37-39], image segmentation. [5-8, 25, 31], and object detection [9-15], have been adopted for inspection in manufacturing fields. Inspection techniques can be broadly categorized into non-referential and referential methods. The primary difference between these two approaches is whether the input consists of a single image or a pair of images (test image and reference image). Referential inspection methods require reference images that are assumed to be flawless [40]. In situations where a reference image is not available, non-referential inspection methods need to be used. For example, Yang *et al*. [41] modified the U-Net architecture [42], which is widely used for image segmentation tasks, to inspect concrete surfaces. Similarly, many researchers have employed U-Net for various inspection tasks [8, 43, 44]. On the other hand, You Only Look Once (YOLO) [45], a network designed for object detection tasks, has also been extensively utilized for localizing defects and identifying their types. Zou *et al*. [15]

adapted the YOLO architecture for metal surface inspection, while Lin *et al*. [14] used YOLO to detect PCB defect types to verify assembly processes. Although non-referential inspection methods showed impressive quality of inspection, we believe that referential methods are more effective when available. Given that IC substrates which we want to inspect have repetitive characteristics, we use referential methods as our baseline. Detailed explanations review of previous studies on referential methods is provided in the following sections.

B. Twin network architecture

The concept of twin network architecture (also known as Siamese network) was initially introduced by Bromley *et al*. [46] in 1993 to address the problem of signature verification by treating it as an image matching problem. This architecture consisted of two identical sub-networks that process two input images and then compare their outputs to determine their similarity. These sub-networks extract features from the two signatures, and the angle between two feature vectors is computed to determine a distance value, indicating the degree of similarity between the signatures. This weight-shared encoder structure of twin networks ensures that both sub-networks process the input data in the same way, extracting features using the identical weights. This ability of learning similarity provides a powerful tool for various similarity-based tasks across different domains. Koch *et al*. [32] adapted Siamese neural networks with a convolutional architecture to rank the similarity between inputs for one-shot learning. This approach involves training with just a single example per class and then identifying the class that most closely matches the input test image. Moreover, this network structure has found extensive application in object tracking, where it measures the similarity between a template image, which is the object to be tracked, and a larger search image to locate the object's position [47-50].

C. Twin network-based change detection methods

Change detection (CD), the task of recognizing changed areas between two images, is widely utilized in remote sensing, medical imaging, and quality inspection. Traditional CD methods typically rely on statistical and mathematical techniques to identify changes [18, 51, 52] but are less robust compared to deep-learning approaches which enables automatic feature learning without complex feature design [53]. In this point of view, twin networks have found significant use in change detection tasks as a deep-learning approach. Daudt *et al*. [35] were among the first to apply twin networks to CD, using a weight-shared encoder to extract features from paired input images, which were then passed to a decoder using fusion strategies like concatenation or difference. However, this method was vulnerable to pseudo-changes, such as color variations, and registration errors [53].

Recent advancements in twin networks include the integration of attention mechanisms to improve robustness [19, 20, 22, 34, 53, 54]. These mechanisms enhance the network's ability to concentrate on relevant parts of the input images and reduce the impact of less important features. Chen *et al*. [20] introduced a dual attention mechanism to recognize features of objects that belong to the same category but have different appearances. However, this mechanism was applied within a single image, thus it is crucial to account for feature relationships between paired images for improved robustness. The

Spatial-Temporal Attention neural Network (STANet) [19] was developed to account for feature dependencies along the spatial dimension between image pairs. Although STANet is resilient to registration errors, it remains sensitive to characteristic differences such as environmental changes. Choi *et al*. [34] introduced the channel-wise co-attention module, making the twin network more robust to pseudo-changes. This attention module computes correlations between each channel to find similar feature maps in the other image, reducing false positives by comparing one set of feature maps with the similar feature maps identified by the co-attention module. Recently, with the rise of transformer architectures in computer vision tasks [55], several research has focused on replacing convolutional layers in twin networks with self-attention and cross-attention mechanisms to enhance performance [53, 56, 57]. However, this shift has led to a significant increase in computational complexity [53].

D. Twin network-based inspection methods

Detecting defects in industrial products is a crucial component of quality control in manufacturing processes. To enhance the accuracy and efficiency of these inspections, twin networks can be employed. These networks are particularly useful in various domains where referential images are available for comparison, ensuring that even subtle defects are identified and addressed. Wu *et al*. [58] utilized twin networks for button surface defect detection, while Deshpande *et al.* [59] applied them to steel surfaces. Feng *et al*. [60] applied twin network-based spatial correlation attention mechanism in CNN classification network for armature defect detection.

In the context of semiconductor inspection, twin networks can be effectively used to inspect products like integrated circuit (IC) chips and printed circuit boards (PCBs) due to their repetitive characteristics. For instance, Miao *et al*. [61] applied this architecture to classify PCB's real defects from pseudo-defects. One characteristic of semiconductor inspection is that it is essential to inspect products to ensure product quality and identify micro defects that could affect the performance of the final products. To meet these demands for detailed inspection without false positives, research has concentrated on addressing issues caused by unintended alterations during the inspection process, such as slight shifts between test and reference images or changes in lighting conditions when capturing product images. Ding *et al*. [62] integrated the spatial pyramid pooling network to enhance feature representation, improving robustness to variations in input images. Ling *et al*. [31] proposed a system for detecting welding defects on PCBs using a twin network with skip connections, introducing a correlation module that computes similarity between the input image and the template image. To reduce errors from mis-registration, they applied a global pooling operation to features before entering the correlation module. Despite its benefits, this pooling operation may potentially compromise spatial information. Choi *et al*. [17] proposed a twin-network based system incorporating a co-attention module that simultaneously accounts for spatial and channel-wise dependencies between image pairs for IC substrate inspection. This attention mechanism improved robustness to registration errors and characteristic differences. Nevertheless, the increased computational demand of this approach may limit its practicality in real-time manufacturing environments [36]. Considering these observations, we propose CD method that reduces errors arising from registration errors and color variations while remaining low computational complexity for inspecting semiconductor products.

III. Proposed methods

In this section, we begin by describing the architecture of the proposed network, followed by an introduction to the method with the channel co-attention module and the loss function. For clarity, we refer to the auto-encoder structure that processes a single image as a single network, in contrast to the twin network which takes two images as input. The proposed methods are named as C-TSNet (Combined-Twin and Single Network) and CC-TSNet (C-TSNet with channel co-attention module).

A. C-TSNet

We introduce a Combined structure of a Twin network and a Single network (abbreviated by C-TSNet) for IC substrate inspection system that is robust to mis-registration. The proposed network's overall architecture is depicted in Figure 2. In the illustration, the "Conv Block" consists of a convolution, batch normalization, and ReLU layers, and "T_Conv Block" comprises a transposed convolution, batch normalization, and ReLU layers. For the final layers of each decoder, only a convolution layer is employed to classify each pixel into two classes, that are defect or defect-free.

In this network, pairs of images (*i.e.*, X_a and X_b) are processed through a weight-shared encoder to extract feature maps in the same manner. After processing inputs, the resulting feature maps are concatenated and passed into a twin decoder for comparing differences, while only the feature maps of X_a are used as the input of the single decoder. The final prediction maps are obtained by averaging the outputs of the twin and single decoders. These maps are then passed through a softmax function, and thresholds are adjusted to generate binary masks, optimizing recall, precision, and F1 score.

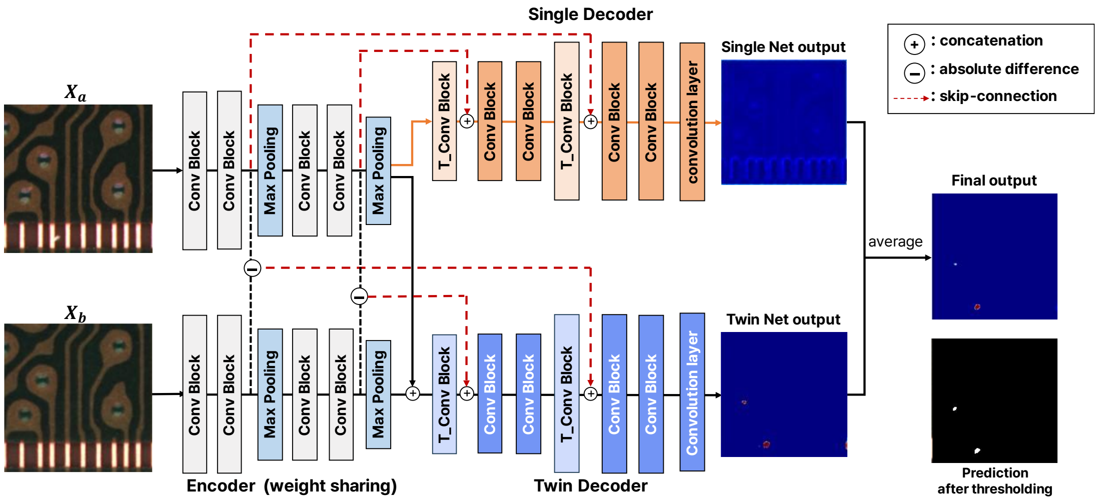


Figure 2. Overall network architecture of the proposed C-TSNet

For the twin decoder, the outputs of the 2^{nd} and 4^{th} convolutional blocks of the encoder are passed as a form of skip connection, involving absolute difference operations between the feature maps of X_a and X_b. This setup allows the decoder to directly receive information about the differences between the two input images. However, if registration errors and pseudo-changes exit in the image pairs, the absolute difference operation may cause errors because the intensity values for these areas can be significantly different.

To address this issue, we merge a single decoder with the twin network, using the output feature maps of X_a as input of single decoder. The intuition beneath this is that single auto-encoder structure only takes one image as an input, making it free from pseudo-change issues or mis-registration. By combining the single auto-encoder structure with the twin network, we aim to improve the ability to reduce false positives and enhance its overall robustness without attention modules.

B. CC-TSNet

Inspired by previous work [34], we incorporate a channel co-attention module, which is introduced by Choi *et al.* [34], into the C-TSNet to address errors arising from characteristic differences. The aim is to enhance the robustness of the network in detecting changes accurately in the presence of pseudo-changes and registration errors. As depicted in Figure 3 the overall architecture of the CC-TSNet integrates the channel co-attention module into the feature maps from the last convolutional block. “Conv Block” and “T_Conv Block” shown in Figure 3 are made up of same layers that are described in the previous section.

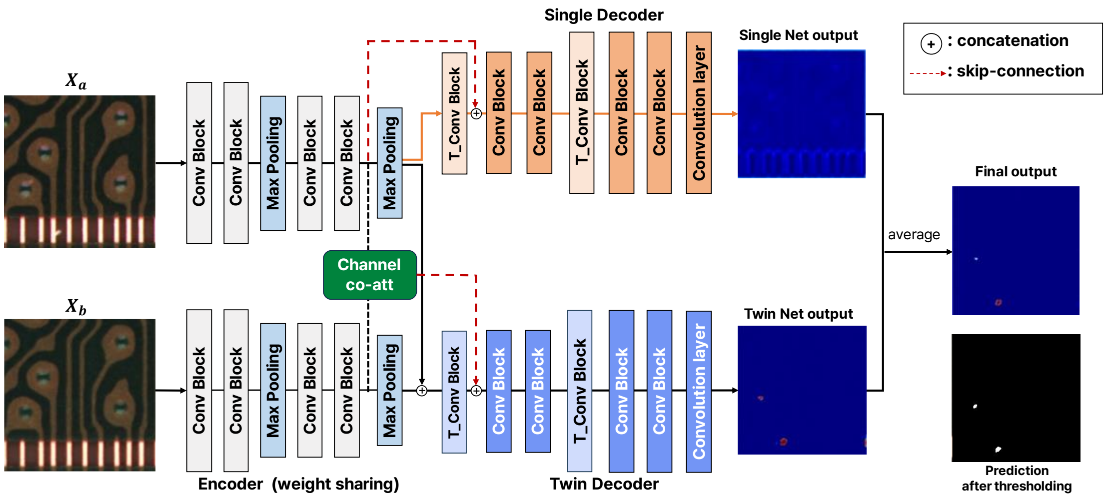


Figure 3. Overall network architecture of the proposed CC-TSNet

The structure of the channel co-attention module that we utilized is illustrated in Figure 4 [34]. Here, h, w and c denotes the shape of feature maps comes from the module. The module computes the cosine similarity between F_a and F_b, the feature maps of images X_a and X_b, respectively. This process generates an affinity matrix S^c that identifies similar feature maps between the two images. For instance, if the j^{th} feature map of F_a is similar to the i^{th} feature map of F_b, the value S_{ij}^c, indicating the degree of similarity, will be high.

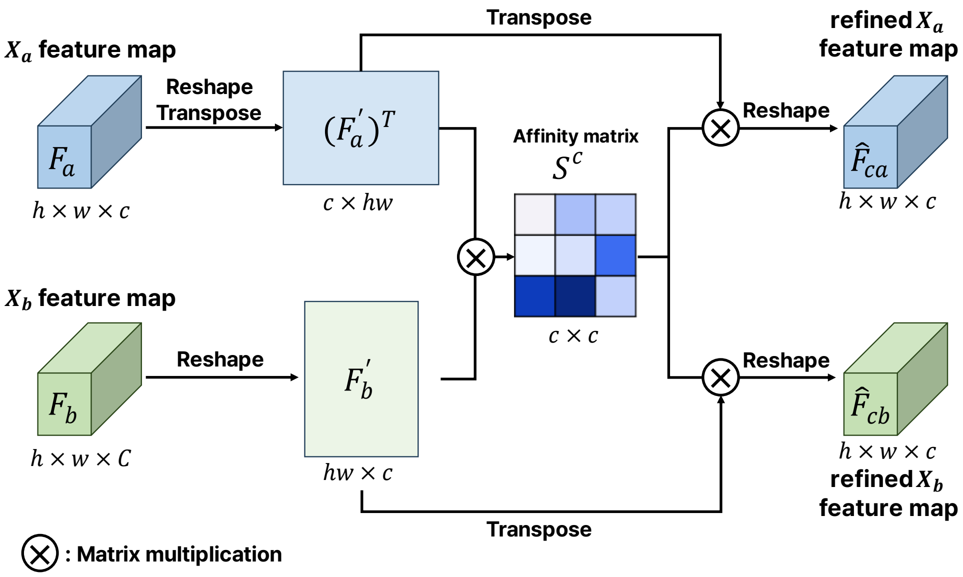


Figure 4. Overview of a channel co-attention module

The refined feature maps $\hat{F}_{ca}$ and $\hat{F}_{cb}$ are produced by multiplying the reshaped feature maps with the channel-wise correlations calculated as S^c. These refined feature maps are then concatenated with original feature maps and pass through a 1×1 convolutional layer. The differences between the feature maps are obtained as follows:

$$D_{att} = abs\left(f_g([F_a; \hat{F}_{ca}]) - f_g([F_b; \hat{F}_{cb}])\right), \quad (1)$$

where [;] represents the concatenate operation, and f_g denotes a 1×1 convolutional layer.

To ensure defects are not missed, the absolute difference of F_a and F_b is concatenated with D_{att}, generating the combined difference map D_{out} as follows:

$$D_{out} = f_h([D_{att}; abs(F_a - F_b)]), \tag{2}$$

where f_h indicates a 1×1 convolutional layer. Finally, D_{out} is passed to the twin decoder in the form of a skip connection.

The integration of C-TSNet and the channel co-attention moduel ensures the network effectively manages both registration errors and pseudo-changes, resulting in more accurate change detection in IC substrate inspections. This hybrid approach leverages the strengths of both twin and single networks, providing a comprehensive and robust solution for the defect detection.

C. Loss function

The number of pixels in defect class is much smaller than normal classes for this pixel-wise classification. Therefore, we utilized a weighted cross-entropy loss function to address the class imbalance by assigning different weights w_1 and w_2 to each class [17]. The weighted cross-entropy loss that we employ is defined as follows:

$$L_{wce} = -\frac{1}{D}\sum_{i=1}^{H}\sum_{j=1}^{W} w_1 y_{ij} log p_{ij} + w_2(1 - y_{ij})\log(1 - p_{ij}) \tag{3}$$

Here, y_{ij} and p_{ij} denotes the ground truth where 1 indicates a defect and 0 indicate a defect-free pixel, and the predicted probability of being a defect for a pixel at position (i, j), respectively. H is height of the output images and W is the width of the output image. And D is the total number of pixels in the images (*i.e.*, $H \times W$). In this study, we set w_1, weights for the defect class to 0.6 and w_2, weights for the defect-free class to 0.4. By assigning higher weights to the defective pixels, the networks are better equipped to learn and detect defects.

IV. Experiment

In this section, we begin by describing the dataset used to verify the performance of our proposed methods and detail the experimental setup. Finally, we outline the evaluation metrics and present the experimental results. For comparison, we employed U-Net as a non-referential method, and Base-Twin (an encoder-decoder based twin network with skip-connections) and Twin-Comb-Co-Att [17] as referential methods.

A. Dataset

To validate the performance of our proposed methods, we conducted experiments on a real-world IC substrate dataset. The dataset is consisted of 88 full-size image pairs sized 4,040×5,036 pixels and 16 full-size image pairs sized 3,280×1,900 pixels. We cropped the same positions from the full-size image pairs into patches sized 200×200 pixels. This process yielded a dataset of 22,432 image pairs for training, including 7,264 defective image pairs and 15,168 defect-free image pairs with horizontal and vertical flip augmentation. For testing, we created 6,352 image pairs consist of 823 defective and 5,529 defect-free pairs The test set contains 838 defect segments. Figure 5 shows examples of the cropped dataset with ground truth annotations. We randomly selected 20% of the training dataset for validation to monitor over-fitting during training.

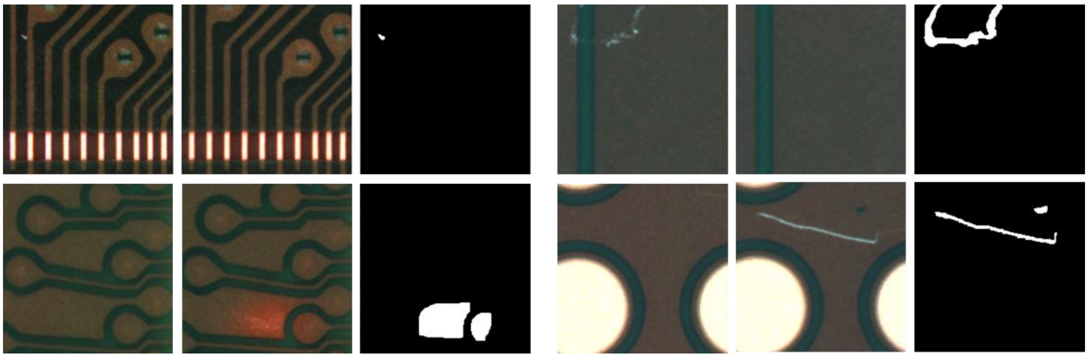


Figure 5. Examples of IC substrate dataset with ground truth

B. Implementation details

We conducted experiments using the TensorFlow 2 framework on an NVIDIA GeForce RTX 3090 graphics card (Nvidia corporation, USA). The networks were trained using the Adam optimizer with step learning rate scheduling. We set initial learning rate to 1×10^{-4} with a decay rate of 0.1. To prevent overfitting, we employed early stopping technique, executing training when the validation loss did not decrease for a specified number of epochs. Table **1** outlines the training options for each method, where B, Ep, D-ep, and P indicate batch size, maximum epoch, decay epoch and patience for early stopping, respectively.

Table 1. Training options for each method

Method	B	Ep	D-ep	P
U-Net	32	300	20	20
Base-Twin	16	200	30	20
Twin-Comb-Co-Att	32	200	20	15
C-TSNet	16	300	20	20
CC-TSNet	8	200	30	20

C. Performance metrics

To evaluate each method qualitatively, we utilized precision, recall, and F1-score. The metrics are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

where TP, FP, and FN represents the number of true positives, false positives, and false negatives, respectively. TP indicates that defect segments are correctly detected, while FP indicates that defect-free areas are incorrectly predicted as defect. FN represents defect segments that are not detected. Precision is the fraction of correctly detected defects among all segments predicted as defects, and recall is the ratio of correctly detected defects out of the actual defect segments.

D. Experimental results

To consider both computational latency and quantitative aspects, we report number of parameters and inference time and then experimental results with performance metrics. First, we listed the number of parameters and the inference for one 200×200 image in Table 2 that was measured using GeForce RTX 3090 graphics card.

Table 2. The number of parameters and inference time for one 200×200 image

Method	Params	Inference time (ms)	Ratio
U-Net	91,010	0.716	0.157
Base-Twin	116,610	1.253	0.275
Twin-Comb-Co-Att	118,722	4.557	1
C-TSNet	190,580	1.459	0.320
CC-TSNet	**190,338**	**1.542**	**0.338**

As shown in Table 2, the Twin-Co-Attention network required the longest processing time due to the extensive computations involved in spatial-channel-wise correlations. The proposed methods were approximately three times faster than the Twin-Comb-Co-Att network but were 1.16 times slower than the Base-Twin network. The CC-TSNet was 1.06 times slower than the C-TSNet.

Table 3 presents the experimental results for the proposed and comparison methods on the IC substrate dataset. To ensure a fair comparison, we fixed the recall values for all methods. The performance of each method was averaged over three repetitions using the same hyper-parameters.

Table 3. Experimental results for IC substrate dataset

Method	Recall (%)	Precision (%)	F1-score (%)
U-Net	93.68	84.59	88.90
Base-Twin	93.68	88.21	90.85
Twin-Comb-Co-Att	93.68	94.54	**94.11**
C-TSNet	93.68	93.68	93.68
CC-TSNet	**93.68**	**94.09**	**93.88**

As indicated in Table 3, the quantitative results of the U-Net, which is a non-referential method, performed worse than the referential methods. The proposed methods achieved an F1-score 2.83 to 3.03 percentage points higher than the Base-Twin network. By utilizing the channel co-attention module into the C-TSNet, the performance slightly increased, resulting in an F1-score only 0.23 percentage points lower than the Twin-Comb-Co-Att network. We believe that slight performance degradation of the proposed networks is acceptable given that the inference time is one-third that of the Twin-Comb-Co-Att, as shown in Table 2.

For qualitative performance evaluation, Figure 6 demonstrates that U-Net struggles to identify defect-free areas that look similar to defects. In contrast, referential methods, including the proposed ones, better distinguish pattern from defects. Figure 7 provides examples of image pairs with registration errors, showing the Base-Twin network's vulnerability to mis-registration. Conversely, the proposed methods exhibit robustness to registration errors, performing comparably to the Twin-Comb-Co-Att method. Figure 8 shows the prediction results for each method when inspecting image pairs with characteristic differences. Similar to Figure 7, it indicates that the Base-Twin network is also susceptible to color variations, whereas the proposed methods demonstrate greater robustness to characteristic differences. This may suggest that the single decoder part of the C-TSNet effectively corrects some false positives caused by registration errors and color changes. However, despite some corrections, the C-TSNet still has errors due to characteristic differences as shown in Figure 9. These false caused by color variations are corrected with the channel-co-att module (*i.e.*, CC-TSNet), similar to the Twin-Comb-Att network.

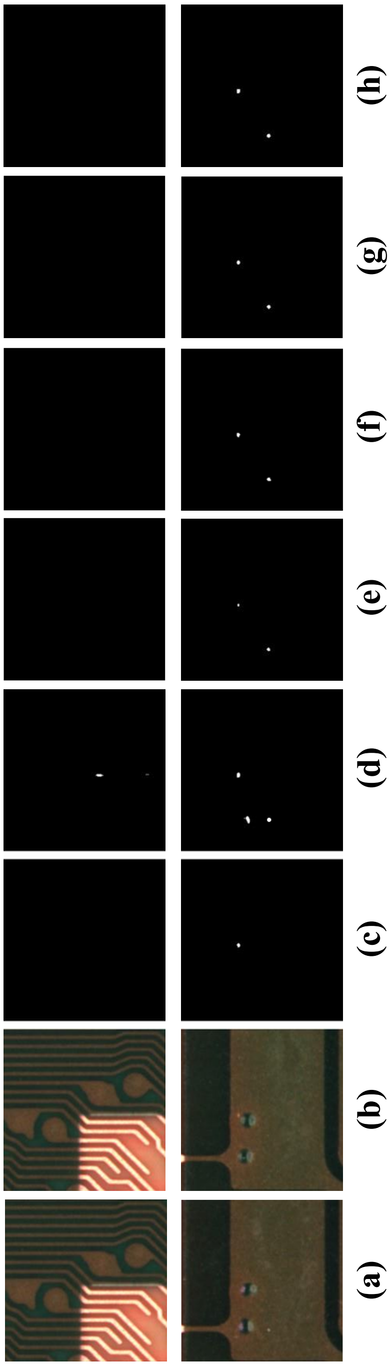


Figure 6. Visualization results of each method on IC substrate dataset: (a) $\boldsymbol{X_a}$ image, (b) $\boldsymbol{X_b}$ image, (c) ground truth, (d) U-Net, (e) Base-Twin, (f) Twin-Comb-Co-Att, (g) C-TSNet, and (h) CC-TSNet

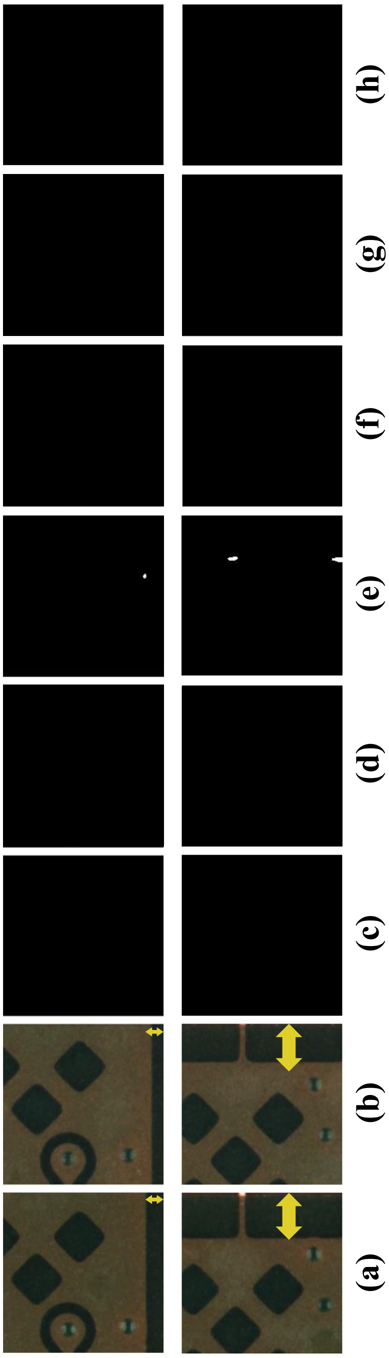


Figure 7. Visualization results of each method on IC substrate dataset with registration errors: (a) $\boldsymbol{X_a}$ image, (b) $\boldsymbol{X_b}$ image, (c) ground truth, (d) U-Net, (e) Base-Twin, (f) Twin-Comb-Co-Att, (g) C-TSNet, and (h) CC-TSNet. The yellow arrows indicate the area where mis-registration occurred.

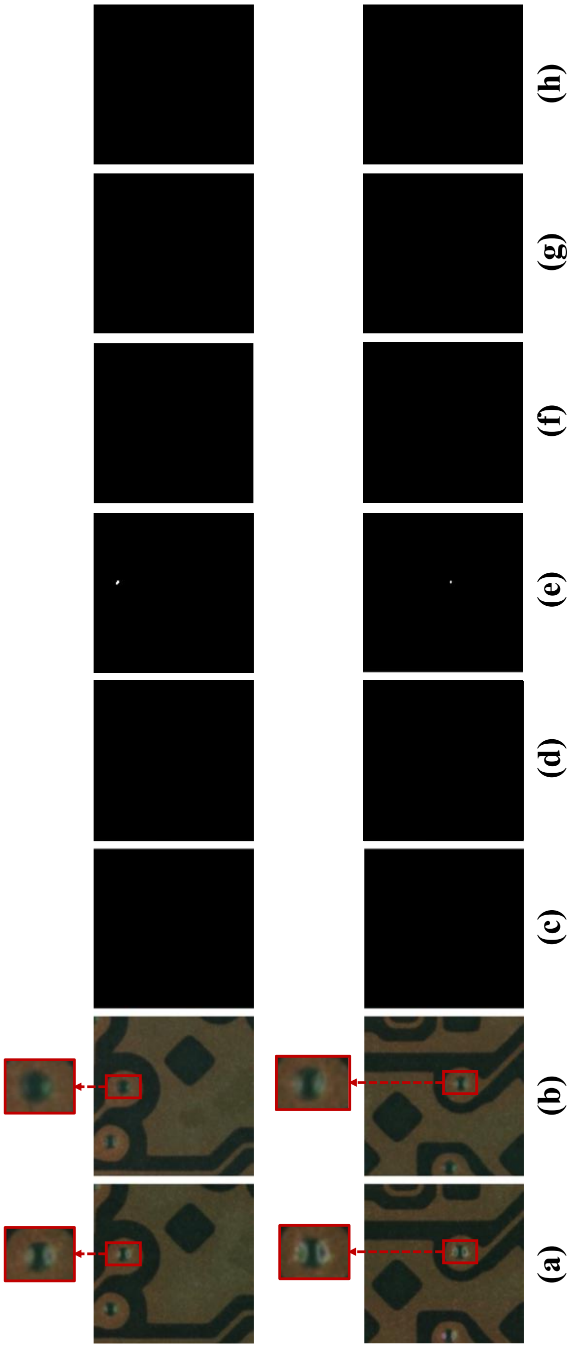


Figure 8. Visualization results of each method on IC substrate dataset with characteristic differences: (a) $\boldsymbol{X_a}$ image, (b) $\boldsymbol{X_b}$ image, (c) ground truth, (d) U-Net, (e) Base-Twin, (f) Twin-Comb-Co-Att, (g) C-TSNet, and (h) CC-TSNet

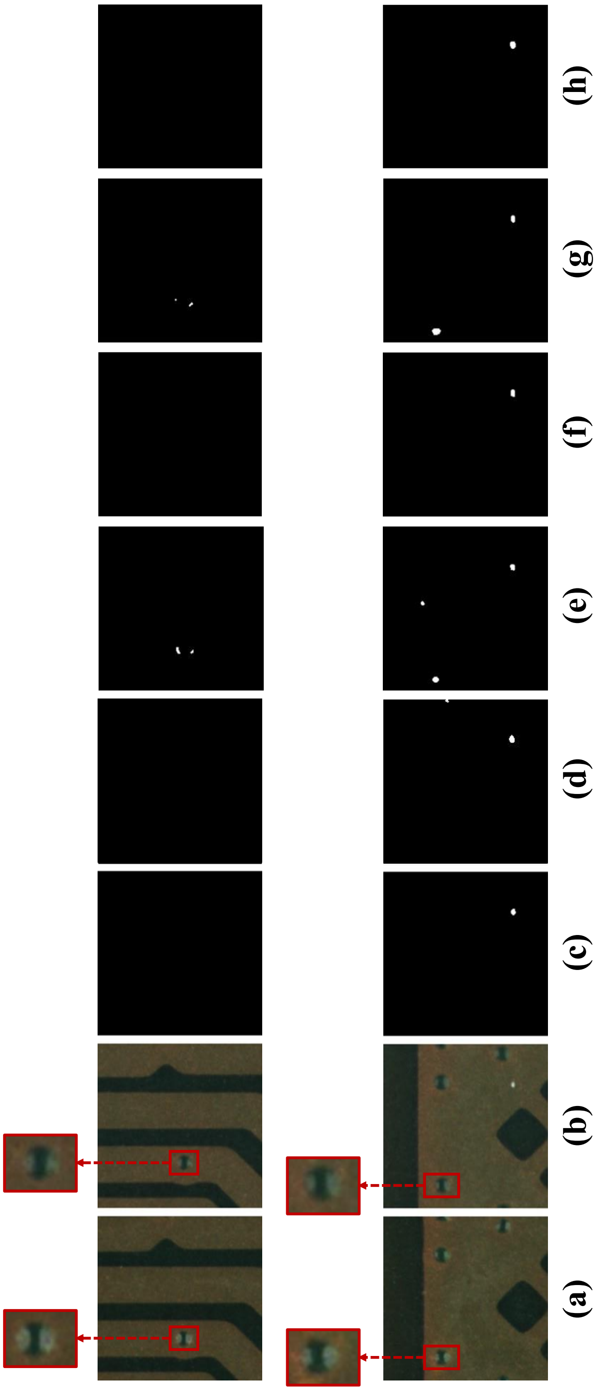


Figure 9. Visualization results of each method on IC substrate dataset with characteristic differences: (a) $\boldsymbol{X_a}$ image, (b) $\boldsymbol{X_b}$ image, (c) ground truth, (d) U-Net, (e) Base-Twin, (f) Twin-Comb-Co-Att, (g) C-TSNet, and (h) CC-TSNet

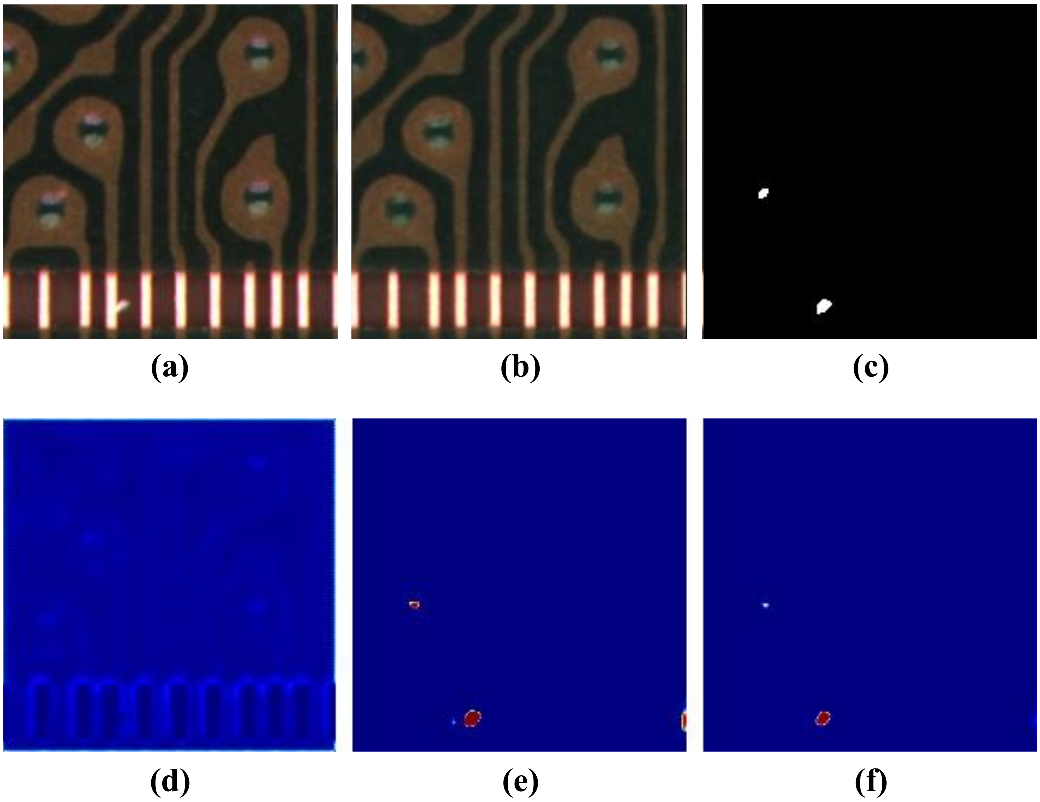


Figure 10. Illustrations of logits from C-TSNet: (a) $\boldsymbol{X_a}$ image, (b) $\boldsymbol{X_b}$ image, (c) ground truth, (d) single decoder logit, (e) twin decoder logit, and (f) final logit after averaging outputs from single and twin decoder.

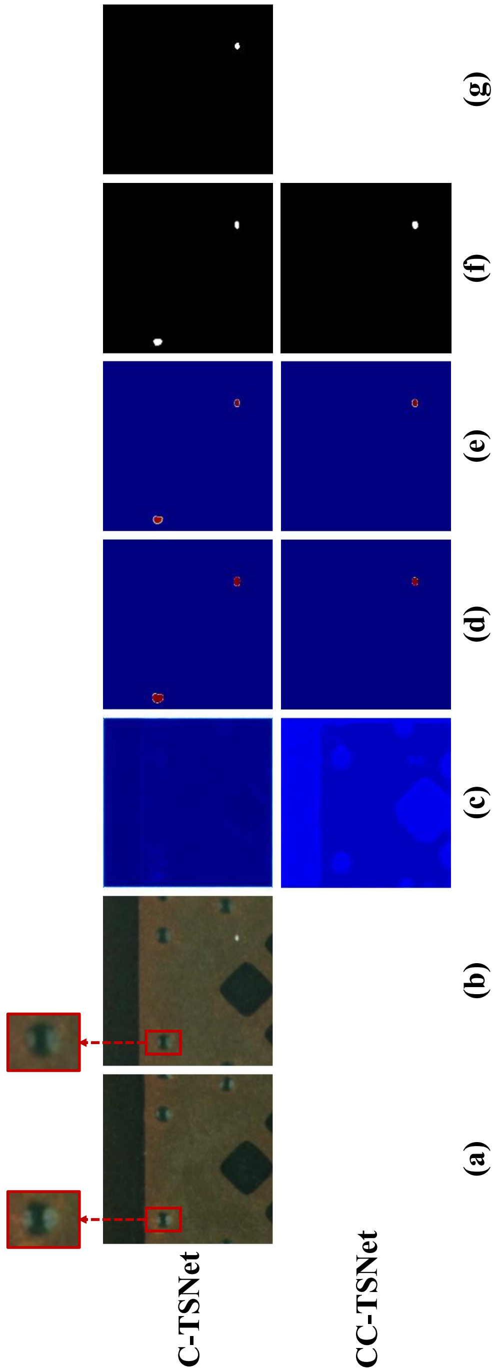


Figure 11. Illustrations of logits from C-TSNet and CC-TSNet: (a) $\boldsymbol{X_a}$ image, (b) $\boldsymbol{X_b}$ image, (c) single decoder logit, (d) twin decoder logit, (e) final logit after averaging outputs from single and twin decoder, and (g) ground truth

To further verify the effectiveness of the proposed methods, we visualize the logits from both the single decoder and twin decoder parts. As shown in Figure 10, the output from the single decoder is highly influenced by both patterns and defects. In contrast, the twin decoder output does not assign high scores to patterns but does show errors in the lower-right corner due to registration issues. Figure 10 (f) demonstrates that averaging the logits from the single and twin decoders reduce the false positive by mis-registration. Additionally, Figure 11 illustrates the effectiveness of the channel co-attention module by comparing C-TSNet and CC-TSNet. The image pair used in Figure 11 contains illumination changes. While the C-TSNet is influenced by characteristic differences that are not corrected by single decoder, some of these errors are corrected in CC-TSNet. Combining all quantitative and qualitative results, we can conclude that the proposed methods effectively reduce errors caused by mis-registration and characteristic differences, offering a cost-effective solution.

V. Conclusion

We propose twin network-based change detection systems that are both fast and robust against mis-registration and pseudo changes for integrated circuit (IC) substrate inspection. To mitigate the negative effects caused by referencing paired images in a twin network, we combine a twin network with a single decoder, which remains unaffected by registration errors and pseudo changes. Additionally, we utilize a channel co-attention module to enhance the network's robustness to characteristic differences. The experimental results of the proposed methods demonstrated a higher F1-score compared to the Base-Twin and U-Net, and nearly matched the performance of the Twin-Comb-Co-Att. Our proposed methods have shown to be effective in not only identifying defects but also minimizing errors caused by misalignment. Moreover, the inference time of the proposed network has been decreased to one-third of that of the Twin-Comb-Co-Att, making it more practical for using in the manufacturing process.

Bibliography

[1] G. Wen, Z. Gao, Q. Cai, Y. Wang, and S. Mei, "A novel method based on deep convolutional neural networks for wafer semiconductor surface defect inspection," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 12, pp. 9668-9680, 2020.

[2] S. Mei, H. Yang, and Z. Yin, "An unsupervised-learning-based approach for automated defect inspection on textured surfaces," *IEEE transactions on instrumentation and measurement,* vol. 67, no. 6, pp. 1266-1277, 2018.

[3] X. Zheng, S. Zheng, Y. Kong, and J. Chen, "Recent advances in surface defect inspection of industrial products using deep learning techniques," *The International Journal of Advanced Manufacturing Technology,* vol. 113, pp. 35-58, 2021.

[4] C. Bu, R. Li, T. Liu, R. Shen, J. Wang, and Q. Tang, "Micro-crack defects detection of semiconductor Si-wafers based on Barker code laser infrared thermography," *Infrared Physics & Technology,* vol. 123, p. 104160, 2022.

[5] J. Han and S. Hong, "A New Backbone Network for Instance Segmentation: Application on a Semiconductor Process Inspection," *IEEE Access,* vol. 8, pp. 218110-218121, 2020.

[6] T. Schlosser, M. Friedrich, F. Beuth, and D. Kowerko, "Improving automated visual fault inspection for semiconductor manufacturing using a hybrid multistage system of deep neural networks," *Journal of Intelligent Manufacturing,* vol. 33, no. 4, pp. 1099-1123, 2022.

[7] N. Ranjan, S. Bhandari, Y.-C. Kim, and H. Kim, "Polycrystalline silicon wafer scratch segmentation based on deep convolutional autoencoder," in *2022 International Conference on Electronics, Information, and Communication (ICEIC)*, 2022: IEEE, pp. 1-4.

[8] X. Fan, P. Cao, P. Shi, J. Wang, Y. Xin, and W. Huang, "A nested unet with attention mechanism for road crack image segmentation," in *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*, 2021: IEEE, pp. 189-193.

[9] Q. Zhang and H. Liu, "Multi-scale defect detection of printed circuit board based on feature pyramid network," in *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2021: IEEE, pp. 911-914.

[10] P. V. Pereira *et al.*, "Automatic Fault Detection in Soldering Process During Semiconductor Encapsulation," in *International Conference on Informatics in Control, Automation and Robotics*, 2022: Springer, pp. 130-145.

[11] P. P. Shinde, P. P. Pai, and S. P. Adiga, "Wafer defect localization and classification using deep learning techniques," *IEEE Access,* vol. 10, pp. 39969-39974, 2022.

[12] H. Wu, R. Lei, and Y. Peng, "Pcbnet: A lightweight convolutional neural network for defect inspection in surface mount technology," *IEEE Transactions on Instrumentation and Measurement,* vol. 71, pp. 1-14, 2022.

[13] H. Chen, Z. He, B. Shi, and T. Zhong, "Research on recognition method of electrical components based on YOLO V3," *IEEE Access,* vol. 7, pp. 157818-157829, 2019.

[14] Y.-L. Lin, Y.-M. Chiang, and H.-C. Hsu, "Capacitor detection in PCB using YOLO algorithm," in *2018 International Conference on System Science and Engineering (ICSSE)*, 2018: IEEE, pp. 1-4.

[15] Y. Zou and Y. Fan, "An Infrared Image Defect Detection Method for Steel Based on Regularized YOLO," *Sensors,* vol. 24, no. 5, p. 1674, 2024.

[16] K. Wu, J. Tan, and C. Liu, "Cross-domain few-shot learning approach for lithium-ion battery surface defects classification using an improved Siamese network," *IEEE Sensors Journal,* vol. 22, no. 12, pp. 11847-11856, 2022.

[17] E. Choi and J. Kim, "Spatial and Channel-wise Co-Attention-based Twin Network System for Inspecting Integrated Circuit Substrate," *IEEE Transactions on Semiconductor Manufacturing,* 2023.

[18] P. C. Smits and A. Annoni, "Toward specification-driven change detection," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 38, no. 3, pp. 1484-1488, 2000.

[19] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sensing,* vol. 12, no. 10, p. 1662, 2020.

[20] J. Chen *et al.*, "DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 14, pp. 1194-1206, 2020.

[21] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Transactions on geoscience and remote sensing,* vol. 57, no. 1, pp. 574-586, 2018.

[22] H. Jiang, X. Hu, K. Li, J. Zhang, J. Gong, and M. Zhang, "PGA-SiamNet: Pyramid feature-based attention-guided Siamese network for remote sensing orthoimagery building change detection," *Remote Sensing,* vol. 12, no. 3, p. 484, 2020.

[23] L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis," *Ieee Access,* vol. 8, pp. 126385-126400, 2020.

[24] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 42, pp. 565-571, 2018.

[25] L. Ma *et al.*, "Object-based change detection in urban areas: The effects of segmentation strategy, scale, and feature space on unsupervised methods," *Remote Sensing,* vol. 8, no. 9, p. 761, 2016.

[26] J. Sublime and E. Kalinicheva, "Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami," *Remote Sensing,* vol. 11, no. 9, p. 1123, 2019.

[27] Y. Cai, Y. Li, C. Qiu, J. Ma, and X. Gao, "Medical image retrieval based on convolutional neural network and supervised hashing," *IEEE access,* vol. 7, pp. 51877-51885, 2019.

[28] K. Wang, M. Zheng, H. Wei, G. Qi, and Y. Li, "Multi-modality medical image fusion using convolutional neural network and contrast pyramid," *Sensors,* vol. 20, no. 8, p. 2169, 2020.

[29] B. Zhang *et al.*, "Short-term lesion change detection for melanoma screening with novel siamese neural network," *IEEE transactions on medical imaging,* vol. 40, no. 3, pp. 840-851, 2020.

[30] Y. Fridman, M. Rusanovsky, and G. Oren, "ChangeChip: A reference-based unsupervised change detection for PCB defect detection," in *2021 IEEE Physical Assurance and Inspection of Electronics (PAINE)*, 2021: IEEE, pp. 1-8.

[31] Z. Ling, A. Zhang, D. Ma, Y. Shi, and H. Wen, "Deep Siamese semantic segmentation network for PCB welding defect detection," *IEEE Transactions on Instrumentation and Measurement,* vol. 71, pp. 1-11, 2022.

[32] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, 2015, vol. 2, no. 1.

[33] H. Chen *et al.*, "Anatomy-aware siamese network: Exploiting semantic asymmetry for accurate pelvic fracture detection in x-ray images," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, 2020: Springer, pp. 239-255.

[34] 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.

[35] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE international conference on image processing (ICIP)*, 2018: IEEE, pp. 4063-4067.

[36] J. Yang, G. Fu, W. Zhu, Y. Cao, Y. Cao, and M. Y. Yang, "A deep learning-based surface defect inspection system using multiscale and channel-compressed features," *IEEE transactions on instrumentation and measurement,* vol. 69, no. 10, pp. 8032-8042, 2020.

[37] C.-F. J. Kuo, C.-Y. Lai, C.-H. Kao, and C.-H. Chiu, "Integrating image processing and classification technology into automated polarizing film defect inspection," *Optics and Lasers in Engineering,* vol. 104, pp. 204-219, 2018.

[38] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE transactions on instrumentation and measurement,* vol. 63, no. 4, pp. 877-888, 2013.

[39] L. Zhang, Y. Bian, P. Jiang, and F. Zhang, "A transfer residual neural network based on ResNet-50 for detection of steel surface defects," *Applied Sciences,* vol. 13, no. 9, p. 5260, 2023.

[40] E. Choi and J. Kim, "Robust Inspection of Integrated Circuit Substrates Based on Twin Network With Image Transform and Suppression Modules," *IEEE Access,* 2023.

[41] L. Yang, B. Li, W. Li, B. Jiang, and J. Xiao, "Semantic metric 3d reconstruction for concrete inspection," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 1543-1551.

[42] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 2015: Springer, pp. 234-241.

[43] L. Cheng, J. Yi, A. Chen, and Y. Zhang, "Fabric defect detection based on separate convolutional UNet," *Multimedia Tools and Applications,* vol. 82, no. 2, pp. 3101-3122, 2023.

[44] F. Liu and L. Wang, "UNet-based model for crack detection integrating visual explanations," *Construction and Building Materials,* vol. 322, p. 126265, 2022.

[45] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[46] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems,* vol. 6, 1993.

[47] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14*, 2016: Springer, pp. 850-865.

[48] A. He, C. Luo, X. Tian, and W. Zeng, "A twofold siamese network for real-time object tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4834-4843.

[49] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8971-8980.

[50] Z. Zhang and H. Peng, "Deeper and wider siamese networks for real-time visual tracking," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4591-4600.

[51] I. V. Nikiforov, "A generalized change detection problem," *IEEE Transactions on Information theory,* vol. 41, no. 1, pp. 171-187, 1995.

[52] S.-C. Liu, C.-W. Fu, and S. Chang, "Statistical change detection with moments under time-varying illumination," *IEEE Transactions on Image Processing,* vol. 7, no. 9, pp. 1258-1268, 1998.

[53] X. Zhang, S. Cheng, L. Wang, and H. Li, "Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 61, pp. 1-15, 2023.

[54] J. Huang, Q. Shen, M. Wang, and M. Yang, "Multiple attention Siamese network for high-resolution image change detection," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 60, pp. 1-16, 2021.

[55] A. a. B. Dosovitskiy, Lucas and Kolesnikov, Alexander and Weissenborn, Dirk and Zhai, Xiaohua and Unterthiner, Thomas and Dehghani, Mostafa and Minderer, Matthias and Heigold, Georg and Gelly, Sylvain and Uszkoreit, Jakob and Houlsby, Neil, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations*, 2021.

[56] W. G. C. Bandara and V. M. Patel, "A transformer-based siamese network for change detection," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022: IEEE, pp. 207-210.

[57] P. Yuan, Q. Zhao, X. Zhao, X. Wang, X. Long, and Y. Zheng, "A transformer-based Siamese network and an open optical dataset for semantic change detection of remote sensing images," *International Journal of Digital Earth,* vol. 15, no. 1, pp. 1506-1525, 2022.

[58] S. Wu, Y. Wu, D. Cao, and C. Zheng, "A fast button surface defect detection method based on Siamese network with imbalanced samples," *Multimedia Tools and Applications,* vol. 78, pp. 34627-34648, 2019.

[59] A. M. Deshpande, A. A. Minai, and M. Kumar, "One-shot recognition of manufacturing defects in steel surfaces," *Procedia Manufacturing,* vol. 48, pp. 1064-1071, 2020.

[60] T. Feng, S. Huang, J. Liu, J. Wang, and X. Fang, "Welding surface inspection of armatures via CNN and image comparison," *IEEE Sensors Journal,* vol. 21, no. 19, pp. 21696-21704, 2021.

[61] Y. Miao, Z. Liu, X. Wu, and J. Gao, "Cost-Sensitive Siamese Network for PCB Defect Classification," *Computational Intelligence and Neuroscience,* vol. 2021, no. 1, p. 7550670, 2021.

[62] R. Ding, C. Zhang, Q. Zhu, and H. Liu, "Unknown defect detection for printed circuit board based on multi-scale deep similarity measure method," *The Journal of Engineering,* vol. 2020, no. 13, pp. 388-393, 2020.

국문초록

윤희준
전자전기공학과
이화여자대학교 대학원

최근 반도체는 인공 지능과 차량용 반도체 등 다양한 산업 분야에서 폭넓게 적용되며, 수요 증가와 성능 고도화를 위한 고정밀 패키징 기술이 요구되고 있다. 반도체 제품은 매우 작은 불량도 성능을 크게 저하시킬 수 있으므로 품질 신뢰성 보장을 위해 고도화된 품질 검사 네트워크가 필수적이다. 반도체 기판 (IC substrate) 불량 검사는 입력 영상 안에 동일한 패턴의 칩이 반복되어 있는 특징을 활용하여 기준 영상과 검사 영상을 비교하는 방법을 사용할 수 있다. 그러나 이 방법은 조명 차이와 같은 의도치 않은 변화(pseudo change)나 및 정합 오차에 취약하여 과검출이 발생할 확률이 높다는 단점이 있다. 이를 해결하기 위해 어텐션 모듈을 활용한 방법들이 제시되었으나, 이러한 모듈은 많은 연산량과 메모리가 요구되어 실제 양산 과정에 적용하기 어려움이 있었다. 따라서 본 논문에서는 IC substrate 검사를 위해 정합 오차 및 pseudo change에 강인하면서도 계산 시간을 최소화할 수 있는 딥러닝 기반 네트워크 C-TSNet (Combined-Twin and Single Network) 을 제안한다. C-TSNet은 영상 간 패턴 비교를 통한 불량 검출 방법과 단일 영상 불량 검출 네트워크를 결합하여 어텐션 모듈을 활용하지 않으면서도 비교 과정에서의 과검출을 최소화한다. 또한, 색상 변화에 보다 강인한 네트워크를 위해 연산 시간에 큰 영향을 주지 않는 채널 어텐션 모듈을 추가한 검사 방법을 제안한다. 실험 결과, 제안된 방법은 기존 방법 대비 검사 성능을 향상시킬 뿐 아니

라 빠른 추론 속도를 보여 실제 제조 공정에서 더욱 효과적으로 활용될 수 있을 것이라 기대한다.