

Received 27 January 2023, accepted 8 February 2023, date of publication 14 February 2023, date of current version 21 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3245093

 SURVEY

Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey

QIN LING¹ AND NOR ASHIDI MAT ISA¹

School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, Nibong Tebal, Penang 14300, Malaysia

Corresponding author: Nor Ashidi Mat Isa (ashidi@usm.my)

This work was supported by the Geran Penyelidikan Pemandanan Universiti Sains Malaysia-Industri through the Project titled “Defect Assessment of Solder Joint Based on X-Ray Images Using Integration of Adaptive Image Processing and Fuzzy Based Look-Up-Table (LUT) Approaches” under Grant 1001.PELECT.8070024.

ABSTRACT Printed circuit boards (PCBs) are a nearly ubiquitous component of every kind of electronic device. With the rapid development of integrated circuit and semiconductor technology, the size of a PCB can shrink down to a very tiny dimension. Therefore, high-precision and rapid defect detection in PCBs needs to be achieved. This paper reviews various defect detection methods in PCBs by analysing more than 100 related articles from 1990 to 2022. The methodology of how to prepare this overview of the PCB defect detection methods is firstly introduced. Secondly, manual defect detection methods are reviewed briefly. Then, traditional image processing-based, machine learning-based and deep learning-based defect detection methods are discussed in detail. Their algorithms, procedures, performances, advantages and limitations are explained and compared. The additional reviews of this paper are believed to provide more insightful viewpoints, which would help researchers understand current research trends and perform future work related to defect detection.

INDEX TERMS Defect detection, PCB, image processing, machine learning, deep learning.

I. INTRODUCTION

The printed circuit board (PCB) is one of the most vital units in the electronic industry [1]. It plays a key role in electronic devices, mechanically holding up and electrically connecting various electronic parts together. PCBs are used in almost every kind of electronic equipment, from electronic watches, smart phones to computers, communication electronic devices and military weapon systems, as long as integrated circuits and other electronic components are present. Benefitting from the development of integrated circuit and semiconductor technology, the size of electronic device components has shrunk down to a tiny scale [2]. PCBs supporting these components are becoming increasingly complicated, diminutive and delicate. Thus, they must be manufactured at a high quality to meet customer demands.

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai¹.

During the Fourth Industrial Revolution (Industry 4.0) [3], PCB manufacturing is facing new challenges and opportunities. The core premise of Industry 4.0 is automated industrial processing with high quality, precision and reliability [4]. Therefore, the producing process of tiny complex PCB boards is required to be more stable and reliable with higher speed, meaning that PCB development must be expeditious [5]. Thus, defect detection in PCBs is crucial. If repeated PCB defects cannot be detected fast and precisely, then many produced boards are likely to be scrapped eventually, which is both wasteful and costly. Currently, achieving real-time, high-precision defect inspection and quality control in PCBs to improve yield and profit is vital for manufacturers [6].

Quality control in the manufacturing process of PCBs is usually challenging because a variety of defects occur inevitably due to mishandling or technical faults. Fig. 1 shows common defects in bare PCBs, such as breakout, open circuit, under-etching, mouse bite, spur, short, spurious copper, over-etching and broken hole. For printed circuit board assembly

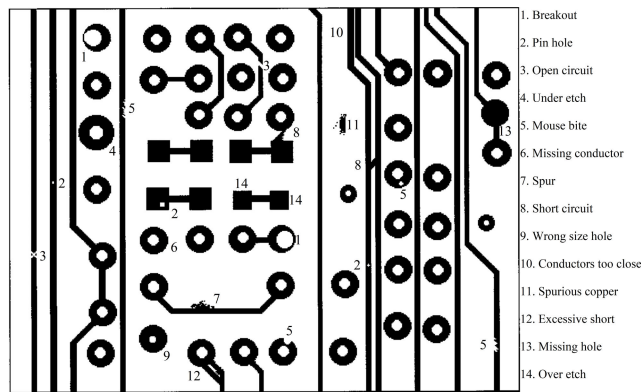


FIGURE 1. Example of defective PCB patterns.

(PCBA), the process of manufacturing includes screen printing, component placement and soldering. Risks of defects are possible, such as missing or misalignment of the component and improper solder joints, containing bridge (i.e. the solders are melted together), empty (i.e. insufficient solder on the solder point), excess solder (i.e. overflowed solder on the solder point) etc. Examples of defective PCBA solder joints are shown in Fig. 2. All these defects could cause the instability of the board or even damage the entire board. Therefore, an efficient, highly accurate automatic detection module needs to be implemented to inspect diverse defects during the PCB manufacturing process.

Previously, basic electric tests, including defect detection, were commonly achieved by manual vision, which requires a few experienced labourers, thus requiring factories to invest heavily in recruiting or training them. However, even the most experienced worker might make mistakes during inspection. Conventional manual inspection was eliminated because components and solder joints are becoming smaller, and possible defects are increasing.

Recently, an automated optical inspection (AOI) technique has been widely used to achieve defect detection on the PCB production line [7]. AOI has numerous advantages over conventional manual inspection. For instance, it is able to scan and detect the board quickly while maintaining high accuracy. During the past decade, although many researchers proposed different image processing methods embedded in the AOI system to make it more efficient and intelligent, the methods could still be classified into three categories: reference comparison, non-reference inspection and hybrid inspection methods [8]. Reference template image-based methods are most widely applied in the manufacturing industry. In this method, the correlation or difference between the template images and the inspected images is acquired to help people estimate defects in PCBs. Furthermore, considering the non-reference approach, a pre-defined mathematical algorithm is proposed to detect PCB defects. The hybrid approach combines reference comparison and non-reference methods.

In addition to traditional image processing methods, machine learning algorithms, such as support vector

machines (SVMs) [9], neural networks (NNs) [10], genetic algorithm (GA) [11] and decision tree (DT) [12] have been integrated into the AOI system to improve the detection accuracy and speed. However, image processing operation still needs to be combined with the machine learning algorithm, because the processing of input images is the foundation of subsequent feature extraction and defect classification. Reference images are also required in these proposed detection or classification methods, most of which are used to perform the template matching [13] operation to extract features or locate regions of interest.

Although the AOI system can help factories release a certain number of labourers, PCB industries still need to devote adequate manpower for quality inspection in cooperation with the AOI system. Thus, to increase the detection accuracy and speed, which helps decrease labour costs, many researchers focus on building advanced traditional rule-based image processing or machine vision-based detection algorithms. In recent years, especially since 2012 when AlexNet was proposed [14], some researchers have tried to apply convolutional neural networks (CNNs) [15] to extract features for defect detection in PCBs.

CNN-based models have achieved outstanding results in many visual detection tasks such as image classification, object detection and semantic segmentation compared with traditional feature extracting algorithms. It has the ability to capture defect features precisely without requiring additional information even if shadows or reflections are present. With these obvious strengths, the CNN-based object detection algorithm keeps refreshing historical records in almost all object detection competitions and progressively occupies the mainstream position in the field of object detection [16]. Hence, this review will discuss and compare traditional image processing-based, machine learning-based and CNN-based algorithms in PCB and PCBA defect detection.

In this paper, we present a detailed review based on the state-of-art PCB defect detection works, filling the gap that few related reviews exist. This review paper is designed to cover broader scopes of PCB defect detection methods to provide good suggestion to researchers or any other intended readers with different competency levels.

Initially, the research methodology, including adoption of database, choice of keywords, searching and classification of research articles, analyses of selected articles is presented for readers to have a clear cognition of the source of selected articles. In addition to a brief introduction of outdated manual defect detection methods, the recent advances of traditional image processing-based methods, machine learning-based methods and deep learning-based methods for PCB defect detection are also covered up to 2022, where the advantages and limitations of each method are critically discussed. Furthermore, our review paper describes the principles and implementation steps of each method, and analyses the performances by comparing their detection criteria and metrics. The comprehensive description conducted in our review paper is expected to offer more new insights to related

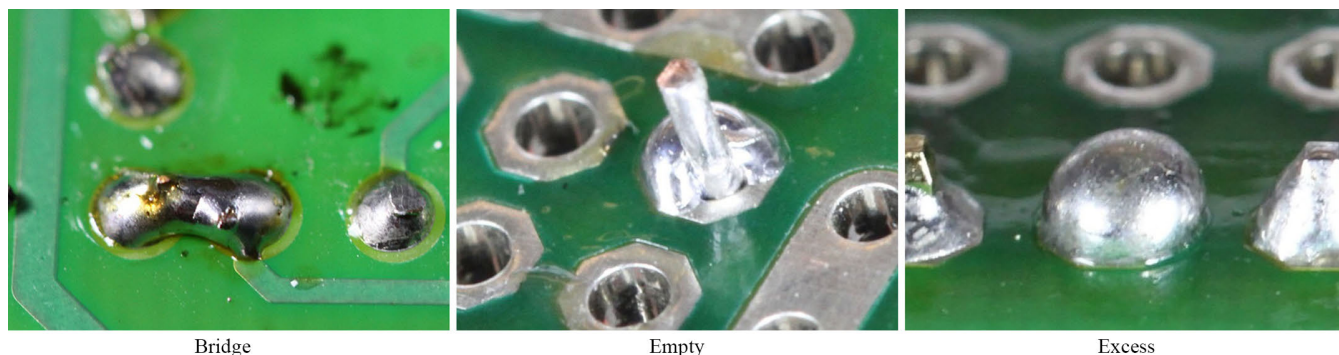


FIGURE 2. Examples of defective solder joints.

researchers or readers in understanding the latest trends of PCB defect detection research, hence assisting them to outline appropriate strategic planning for their future works.

The contributions of this review paper of PCB defect detection can be summarised as follows:

- Detailed descriptions are given of the systematic methodology used for data collection of related articles, including the research definition, choice of search keywords and online databases, classification and verification processes used for selecting relevant papers and research analyses necessary to summarise the discoveries from existing PCB defect detection works.
- A comprehensive review of PCB defect detection methods in past 20 years covering traditional image-based, machine learning-based and deep learning-based is presented. To the best of the authors' knowledge, few reviews for PCB defect detection methods is published except antique ones. Furthermore, recent related reviews only list some PCB defect detection works mainly focusing on traditional image processing and machine learning, without presentation of deep learning based approaches.
- Detailed analyses including principles and implementation procedures, are conducted in our review paper to investigate the strengths and limitations of previous works, which was not considered in previous finite review papers.
- The performances of previous works in terms of detection speed, detection precision are also discussed and compared with each other. Especially, our review paper focus on the inference time and speed simultaneously of each PCB defect detection method instead of detection precision simply, as in real PCB related industries, detection speed is a significant factor that must be considered for application and deployment. Previous review papers only paid close attention to detection precision without practicability.
- Open research challenges from different perspectives are elaborated by referring to the detailed analyses performed on the state-of-art PCB defect detection works

published in recent years. Meanwhile, the corresponding further research directions which are potential to work out each open research challenge are also unveiled in our review paper.

The comparison between our survey and previous published review papers of PCB defect detection is concluded in 1, showing the improvements of our survey.

The remaining sections of this paper are presented as follows: Section II presents the research methodology used to conduct a systematic literature review of PCB defect detection. Section III introduces the manual inspection method. Section IV discusses the image processing-based defect detection methods. Section V discusses machine learning-based defect detection methods. Section VI discusses deep learning-based defect detection methods. Sections VII and VIII present the open research challenges and future research directions, respectively. Lastly, Section IX concludes this paper. The important abbreviations used throughout the paper are listed in Table 2.

II. RESEARCH METHODOLOGY

This section presents the methodology utilised to review PCB defect detection methods. As shown in Fig. 3, a research framework is presented in four main steps: research definition, paper search, paper verification and research analysis [19].

First step: Research definition

Step 1.1: Define the research area. The research areas are PCB defect detection by traditional image processing, machine learning and deep learning methods, as shown in Fig. 4.

Step 1.2: Define the research goal. The research goal is to provide a systematic research scheme and propose ideas to improve the existing PCB defect detection methods.

Step 1.3: Define the research scope. The research scope involves a review of papers on PCB defect detection works, describing their principles and implementation procedures, comparing their performances and discussing their strengths and limitations.

Second step: Paper search

TABLE 1. Comparison between published review papers and our survey.

Reference of the review papers	Year of publication	Contributions	Summary
[8]	1996	- A classification tree of the algorithms is presented - A summary of the commercial PCB inspection system is presented - Various PCB defect types are described	The author presents a comprehensive and detailed review paper about PCB inspection, but it is outdated for the moment.
[17]	2017	- Instead of PCB only, PCBA defects are also involved - Component or solder joint defect detection methods are discussed	The author presents a very brief review paper for PCB and PCBA inspection. But limited methods are introduced, and there is no detailed description for these methods.
[18]	2022	- Deep learning based techniques for PCB defect detection are discussed - Recent inspection models for single layer and multilayer PCBs are included	This recent review paper for PCB inspection is still too brief without adequate references, and there has no detailed description for detection techniques.
Our survey	A comprehensive review of PCB and PCBA defect detection methods covering traditional image-based, machine learning-based and deep learning-based is presented in our survey. Detailed analyses are discussed to investigate the strengths and limitations of previous works. Especially, our review paper focus on the inference time and speed of the algorithms instead of detection precision simply. Open research challenges and the corresponding further research directions are also unveiled. Thus, this paper are believed to provide more insightful viewpoints about current research trends and help researchers perform future work related to defect detection.		

TABLE 2. List of important abbreviations used throughout the paper in alphabetical order.

Abbreviation	Definition	Abbreviation	Definition
AOI	Automated optical inspection	Industry 4.0	Fourth Industrial Revolution
AROC	Area under receiver operating characteristic curve	IoU	Intersection over union
BP	Backpropagation	LD-PCB	Lightweight detection network for PCB components classification
CCD	Charge-coupled Device	LVQ	Learning Vector Quantization
CCL	Connected component labelling	mAP	Mean average precision
CHT	Circle Hough Transform	MLP	Multi-layer perceptron
CNN	Convolutional neural network	MR	Morphological reconstruction
CR-PCB	PCB component character recognition network	NCC	Normalised cross-correlation
CSPNet	Cross Stage Partial Network	NIPS	Conference and Workshop on Neural Information Processing Systems
CS-ResNet	Cost-sensitive residual convolutional neural network	NN	Neural network
CVPR	IEEE Conference on Computer Vision and Pattern Recognition	PANet	Path Aggregation Network
DT	Decision tree	PCB	Printed circuit board
ECCV	European Conference on Computer Vision	PCBA	Printed circuit board assembly
Faster R-CNN	Faster Region-based Convolutional Network	ResNet	Residual Neural Network
FCOS	Fully Convolutional One-Stage	RoI	Region of interest
FPN	Feature Pyramid Network	RPCA	Robust principle component analysis
GA	Genetic algorithm	SSD	Single Shot MultiBox Detector
GARPN	Guided Anchoring Region Proposal Network	SURF	Speeded-up robust feature extraction
GPU	Graphic Processing Unit	SVM	Support vector machine
GW	Geometric-wavelet	TDD-net	Tiny defect detection network
IC	Integrated circuit	VGG	Visual Geometry Group
INCC	Improved normalised cross-correlation	WVT	Wavelet transform

Step 2.1: Define the search terms. The main references were published from 2000 to 2022. Six key terms were used to search related articles, as shown in Fig. 4. All research areas are the main focuses of the review process presented in Section III-VI.

Step 2.2: Search in databases. The online journal databases used for searching research papers based on the search terms are listed as follows:

- IEEE Xplore
- Web of Science
- ScienceDirect
- Springer
- Wiley
- Scopus

Step 2.3: Select only English language papers. Only related articles written in English were reserved. A total of 124 articles were selected for the survey.

Third step: Paper verification

The articles collected in the second step were thoroughly verified by all co-authors. Only the articles related to PCB defect detection were selected.

Step 3.1: Filter and remove irrelevant papers. Articles that generally do not have any significant relation with PCB defect detection were discarded except for several general articles used to introduce basic concepts of different algorithms.

Step 3.2: Double verification process for each paper. The remaining selected articles were determined if they lie in the areas of focus. In this review, our focused parts were architectures, procedures, detection accuracy, inference time and robustness of a detection algorithm, as shown in Fig. 5. Other outstanding articles were used as extra references, thus promoting the value of this review. The articles were mainly obtained from reputable journals, and top-level conference papers in related fields such as IEEE Conference on Computer Vision and Pattern Recognition (CVPR), European

Conference on Computer Vision (ECCV) and Conference and Workshop on Neural Information Processing Systems (NIPS) were selected.

Step 3.3: Classify and store the papers. The gathered articles were divided into 5 specific sub-folders according to their different focus topics: manual inspection, traditional image processing-based, machine learning-based, deep learning-based and other defect inspection. This process was necessary as it ensured the information related to our research was well organised.

Final step: Research analysis

Step 4.1: Analyse the selected papers. We analysed the selected articles to acquire important information such as the frameworks and procedures proposed in the PCB defect detection methods, the improvements compared with original algorithms, the detection performance metrics, detected defect types, inference times and limitations.

Step 4.2: Summarise the papers and make critiques for the proposed methods. We listed the detection criterion, detection performances, advantages and limitations of the proposed approaches. Discussion was conducted to systematically analyse the common shortcomings of existing manual inspection methods, traditional image processing-based methods, machine learning algorithm-based methods and CNN-based methods.

Step 4.3: Identify future research directions and further modifications. Finally, we discussed the potential future development directions for PCB defect detection based on the current research trends and open research challenges observed from the current review analyses.

III. MANUAL DEFECT DETECTION METHODS

Before the AOI technique was applied to detect PCB defects, all the electric tests including defect detection of PCB boards were accomplished through manual vision inspection in most manufacturers. Operators always use a magnifying glass or calibrated microscope to determine whether the PCB is failed or qualified, and to determine when a correction operation is necessary [17]. The tools they utilise are visual inspection and qualified experiences. In this way, experienced inspectors can successfully identify flaws and the types of defects precisely.

Another benefit is that this method has low upfront costs and does not require test fixtures. However, this method highly depends on the skill of operators, who are expected to be practised. This means that factories must spend a large amount of money for labour recruitment. However, even the most experienced expert can make detection mistakes because of his subjective estimation [20]. The disadvantages of this method include high long-term cost, discontinuous defect detection and difficulty in data collection [21]. At present, as PCB production continues to increase and the size of PCB components continues to decrease, many tiny possible defects are emerging. Thus, this outdated manual inspection method, with its low detection speed and efficiency, has become increasingly impractical and has been eliminated [18].

IV. IMAGE PROCESSING-BASED DEFECT DETECTION

Different from manual inspection, the AOI system can eliminate subjective errors and provide a quick, quantitative estimation. This system can work all the time as it does not get tired and burned out, allowing a factory to save a large amount of labour costs. Recently, to promote detection accuracy and speed, which means better quality and lower costs, many researchers have introduced machine vision technique using traditional rule-based image processing method to detect PCB defects. In this section, image processing-based defect detection methods or algorithms for PCBs or PCBA are described.

A. TECHNIQUES FOCUSING ON SURFACE DEFECTS IN PCB

Pal et al. [22] applied the image subtraction method using machine vision to detect bare PCB defects in 2011. This image subtraction approach [8], [23] which is categorised as a reference comparison method, was performed in several steps. Firstly, a reference image has to be obtained among selected good templates. Then, the reference template must be buffered to make it acceptable for subtraction. Next, the inspected image is loaded. To obtain the PCB defects or errors, pixel-by-pixel XOR logic operation, which is also referred to as image subtraction, was implemented between the reference image and inspected image. The resultant image acquired from this subtraction operation should exhibit several defects if the inspected board was defective.

The authors in [22] achieved typical defect detection such as over-etchings (i.e. opens), under-etchings (i.e. shorts) and holes. By subsequent image processing procedures, the precise positions and sizes of defects also were given. However, this method reached its limitations because the reference image needs to have the same size and orientation as the inspected image in terms of pixels, which is very hard to be implemented in real applications. A tiny shift in pixel can cause many false positive targets because the image subtraction is achieved through logical operation pixel-by-pixel, which will give rise to huge financial loss for PCB manufacturers. Moreover, acquiring a real qualified reference image is challenging and time consuming as the illumination environment is always changing all the time. Firms need to prepare different templates corresponding to different PCB boards. Subjected to these prerequisites, the accuracy of this method was not desirable. Missed inspection for bare PCB defects, including open, short, incorrect dimensions and missing conductor occurs frequently by this method. Besides, achieving a high processing speed is impossible for this method due to several additional image processing operations such as image thresholding and image particle analysis.

Ma [24] proposed a similar method to detect defects in bare PCBs using the comparison between the reference image and inspected image. The modification was the standard image was an average of a series of qualified PCBs instead of using only one image. The histogram analysis of the difference

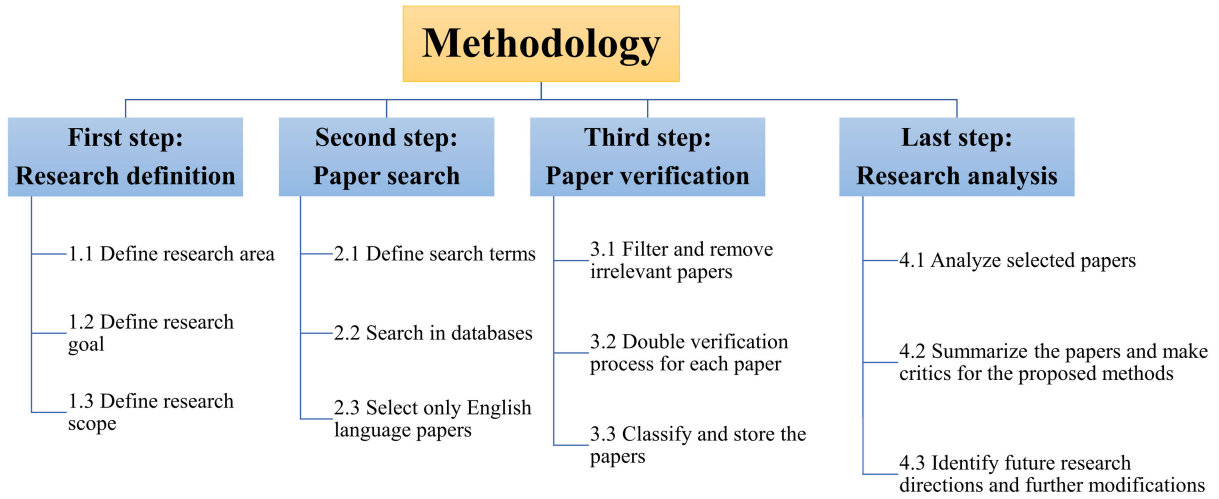


FIGURE 3. Research framework.

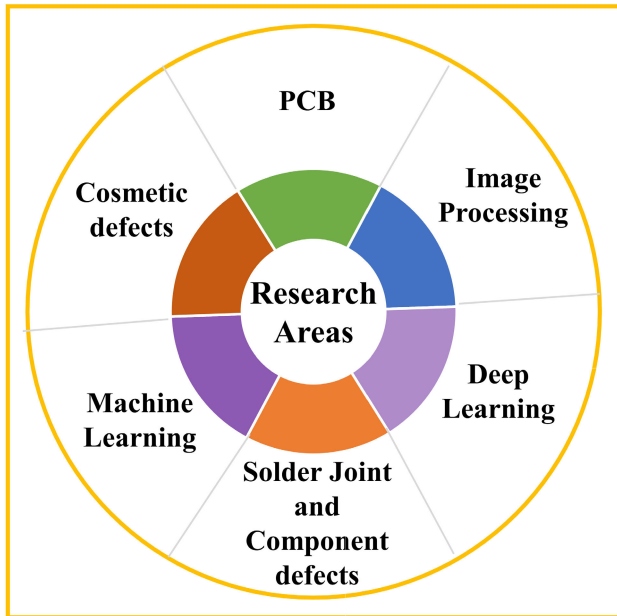


FIGURE 4. Research areas.

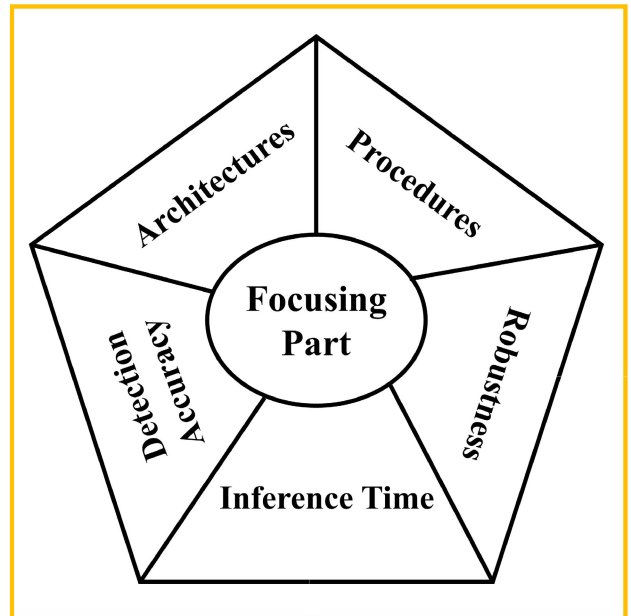


FIGURE 5. Focused review parts.

image was used to determine an appropriate threshold to distinguish potential defect regions. Furthermore, a filter based on target region contours was proposed to capture the real defects from these possible regions, which was reported to outperform those of depending on simple image thresholding. A modified region growing method was applied to obtain the complete zones of the real defects. The defect types were determined by the changing times of the peripheral boundary's grey value.

Method [24] could detect 90% of the defects, including spurious copper, spur, short, mouse bite and open circuit in 100 target images. Compared with [22], this method addressed the template offset problem by averaging and

updating the standard image dynamically. In addition, more image processing operations including bilateral filtering, histogram analysis, region contours filter and region growing were applied to improve the detection precision. So this method can recognize the actual type of defects instead of over-etchings and under-etchings only. Nevertheless, in the operation of updating standard image, the choice of temporal distance between frames becomes a tricky question, which depends on the size and speed of the moving object. Furthermore, many processing procedures cost huge computation, resulting in a slow speed, especially for high resolution images. Another limitation of this method is that the recognition rate for defect types is not desirable. The information

of region contours is not enough to decide the type as some defects have similar outlines and sizes. The colour change of surrounding environment could be considered as an important clue to improve the recognition rate.

Melnyk and Tushnytsky [25] proposed a clustering-based comparison method to detect defects in bare PCBs. A small shift of routing paths and contacts would be considered defects in the traditional subtraction method. To avoid this problem, they applied clustering to the standard and inspected images before the comparison. When the etalon image and defect images were clustered, Equation (1) was applied to obtain the PCB image with marked defects, where $I_r(x, y)$, $I_e(x, y)$, $I_d(x, y)$ are the pixel intensity of the resulting image, the etalon and the defect PCB image, respectively. Tol is a tolerance value to control a difference between an etalon PCB and the controlled sample.

The authors in [25] confirmed the robustness of their method by setting different tolerances such as 50, 150 and 200. The results proved that a comparison of clustered PCB images can ignore small shifts of routing paths and contacts and simultaneously capture real defects. Apart from common limitations of image comparison method presented previously, the difficulties of this method lie in the chosen tolerance and the cluster parameters which need experienced users and many attempts. Improper tolerance can miss real defects as PCB defects always occupy tiny areas in a high resolution image. Hence, this method was not generalized and not convincing in some cases, and the final decision should be made by an expert, which is time consuming and laboured for real application.

$$I_r(x, y) = \begin{cases} I_d(x, y), & \text{if } |I_e(x, y) - I_d(x, y)| < Tol \\ RGB(|I_e(x, y) - I_d(x, y)|, 00), & \text{else} \end{cases} \quad (1)$$

Instead of concentrating on the defects directly, Wenbin and Su [26] focused on the detection and localisation of circular holes on PCBs, which are vital for positioning or defect detection. In this work, a Canny detector was applied to create the corresponding edge maps [27]. Then, the author filtered these short arcs by using a threshold called the step of simpler and faster edge thinning and cleaning. Next, the pre-processed image was transformed by using distance transform to acquire the distance space map. In order to position the centres of these holes correctly, two more steps for searching hole centres—a process called rapid circle centre search—were proposed in the distance space.

The method proposed in [26] was compared with an improved circle Hough Transform (CHT) based method reported by Atherton and Kerbyson [28] to prove its better results. The execution times for Atherton's method and the proposed method were 2.4196 and 1.3420 seconds, respectively, and the F-scores were 0.8083 and 0.9167, respectively. This paper provides a method to position holes on a PCB with a good detection performance than traditional hole detection algorithms. However, this method is still time consuming

for real coloured samples, and a balance between speed and accuracy should be considered to choose a suitable threshold to thin the edges. Moreover, the significance of this paper is limited as it only focused on PCB holes, which merely can help detect or position defects related with holes in PCBs. In the real environment, it is impossible for workers to predict the defect types to choose specified detection method. So this method is not able to be deployed commonly in the real industry with high efficiency.

Recently, Jakkrit and Jakkree [29] proposed an algorithm that can detect and classify 14 PCB defect types [8]. In this algorithm, colour images were firstly converted into greyscale images, and median filtering was used to mitigate the noise. Then, the median-filtered images were scaled, rotated and translated by the operation of affine transform. Fuzzy c-means clustering was applied to segment images before image subtraction for defect detection. In the defect detection stage, image subtraction was applied to obtain the difference between the segmented and the template images. The results for the subtraction were negative, positive and zero, which indicated excess copper, missing copper or non-defective PCB, respectively. In addition, arithmetic and logic operations such as CHT, morphological reconstruction (MR), and connected component labelling (CCL) were used in defect categorisation. The CHT algorithm transformed a variety of feature points from the image space into a series of accumulated votes in parameter space to detect and classify circular patterns. The MR algorithm categorised features such as holes, pads and traces of the segmented images successively, and then converted the connected background pixels into foreground pixels until the object boundaries were reached. CCL [30] scanned the segmented image, divided the pixels into components based on their connectivity and assigned labels to each pixel until the boundary is reached.

The algorithm proposed in [29] achieved 100% defect detection and 99.05% defect classification accuracies for all 14 known PCB defect types and determined defect characteristics such as location, area and nature, showing noticeable superiority than previous works. However, the input images used in this paper were PCB circuit layout schemes in black and white instead of real PCB images. Even for these low resolution images, the computation was large (i.e. from 1.31 seconds to 5.56 seconds per image). The reason is that from [29], about 20 steps were still needed after the pre-processing, segmentation, subtraction and operations of CHT, MR, CCL. This condition means that the computer processing unit must be upgraded to achieve an acceptable detection speed for the real PCB images. Besides, the processing speed also depended on the image segmentation algorithms used including global thresholding, adaptive thresholding, k-means clustering and fuzzy c-means clustering. This indicates that this approach is not efficient and generalized for fast PCB defect detection because users have to decide the type of image segmentation algorithm used for different applications.

B. TECHNIQUES FOCUSING ON COMPONENT AND SOLDER JOINT DEFECTS

Unlike previous methods on detecting surface defects in PCBs, Sundaraj [23] used the background subtraction method to inspect the missing or misaligned components in PCBA. The author provided a solution that was able to monitor a larger region in a PCB at any moment. This automatic visual inspection system was considered efficient at decreasing the inspection time and increasing the handling capacity. In this work, Mixture-of-Gaussian distribution [31] was chosen for background modelling. Then, three stages—background learning, parameter estimation and pixel classification—were implemented to build a Gaussian density function. During the experiment, the non-defective PCBA was firstly fed to the learning process to acquire the template or background image by capturing about 100 frames of the PCBA board. Subsequently, the image of a defective PCBA with missing or misaligned components was captured. The defective image was subtracted from the background image to obtain a binary mask that presents foreground and background pixels. A large area of foreground pixels indicated some kind of defect on the position.

The method reported in [23] could detect missing or misaligned capacitor and chip with an accuracy of more than 90%. However, this method has several limitations. The process of background learning to build a reference image cost huge time (about 3 to 4 seconds), which makes this technique not suitable for real-time detection. Similar to previous works, image subtraction requires both the reference image and tested image must be placed in the same orientation precisely, which is trivial to be achieved in PCB inspection industry. Besides, the accuracy highly depends on hardware such as a rigid optical table, an advanced camera with a macro zoom lens and a fixed lighting with a suitable illumination angle. Furthermore, this system was captious for inspected images. It failed to obtain good results on the component whose colour was the same as that of the background. Same colour presents same or similar pixel values, which will be considered as background and missed in the subtraction operation.

Conventional normalised cross-correlation (NCC) was used to detect defects by comparing the reference image and defect image in [32], but this method is time consuming as it computes the NCC depending on two-dimensional images. Therefore, Annaby et al. [33] proposed an improved normalised cross-correlation (INCC) method to detect integrated circuit (IC) defects such as misaligned components in PCBs. As to the standard image and defect image, considering two corresponding pixels that they contain, blocks with the same size, in which the top left corners are the two pixels, were created. Then, these two-dimensional block images were converted into one-dimensional feature vectors in the matching process. Next, these vectors were augmented by spatial statistical features following discrete cosine transform. After this process, the feature descriptors were obtained. As in the

classical NCC method, the correlation coefficient between the corresponding one-dimensional feature descriptors in the standard and defect images was then calculated. A pixel with a correlation coefficient smaller than a specific threshold is classified as a defect pixel in the testing image. Otherwise, the pixel is a defect-free pixel.

The INCC algorithm proposed in [33] was compared with other popular methods, including conventional NCC, Yoo2010 [34] and Fouda2015 [35] using PCB image datasets. The results proved that INCC performed better than others. INCC have fewer false alarms than other two methods, and the defect regions are much more coincident than the methods whose defect regions are smaller than those of the standard model. INCC detected defect regions at least four times faster than NCC when the image size was bigger than 400×500 . In addition, this method was robust to the changing illumination environment. However, this method has own limitations. Firstly, it was proved to be useful on images contained only one missing or defective component, which is not suitable for real PCB industry. Secondly, although it cut down the computation cost compared with traditional NCC method, it was still inefficient. The core theory of INCC is NCC calculation represented by the dot product of two vectors transformed from standard and detection image, which must cost huge computation. For a large image, it takes several or even more than 10 seconds during the detection process. Hence, it is not able to meet the real-time demand for industry production testing.

Hassanin et al. [36] developed an automatic defect detection method in PCBs on the basis of speeded-up robust feature extraction (SURF) [37] technology and other image processing operations. The reference image and tested image, which had been transformed into binary images, should be pre-processed by a spatial smoothing filter and then the median filter to remove noise. Then, to eliminate the rotation and translation shifts between them, image alignment was performed by using SURF. The rotation angle and translation distance of misalignment can be calculated by SURF and geometric correction was applied depending on the SURF results. After pre-processing, the images were segmented to obtain the foreground pixels using the Otsu thresholding method. The XOR operation was performed to compare the difference between the segmented template image with the segmented inspection image to obtain the defects. Dilation operation was also implemented to reshape the defect regions. Finally, the features of the subtraction image were extracted by SURF, and the feature datasets of each standard component were combined into a matrix. The Euclidean distance between individual defect features with the known components in the feature dataset was evaluated to identify the defect type.

The work in [36] presented statistical characterisation such as the equal error rate estimated based on correct and incorrect matching results and the area under receiver operating characteristic curve (AROC) to prove the good performance. The introduce of SURF given this method noticeable

TABLE 3. Summary of traditional image processing methods.

Work	Concept or approach proposed	Detection criterion	Performance	Advantages	Limitations
Works focusing on surface defects					
[22]	XOR subtraction	NA	NA	Detects typical defects, including over-etchings, under-etchings and holes	- Reference image is required - Other common defects cannot be detected - Same size and orientation of reference and inspected image is required
[24]	Dynamic update of reference image and improved region growing	Detection rate	90%	Weaken the threat from template offset	- Reference image is required - The recognition rate for defect types is not high - Hard to control the averaging template procedure - Time consuming
[25]	Clustering based image comparison	NA	NA	Decreases the influence of small shifts of routing paths	- Reference image is required - Parameters should be chosen - The decision should be made by an expert - Time consuming
[26]	Detection of circular holes	F-measure	90%-100%	Faster and more accurate than [29]	- Still time consuming - The equilibrium between speed and accuracy should be considered - Only focus on holes
[29]	Integration of fuzzy c-means clustering, arithmetic and logic operations, CHT, MR and CCL	Detection accuracy	>99.05%	Classify 14 PCB defect types	- Reference image is required - Huge computation; time consuming - Detects PCB circuit layout schemes in black and white instead of real PCB images - Image segmentation algorithm must be chosen for different datasets
Works focusing on solder joints or components					
[23]	Background subtraction focusing on missing or misaligned components	Detection accuracy	>90%	- Can monitor a larger region - Increase the handling capacity	- Reference image is required - Fails to detect the component that has the same colour as the background - Same size and orientation of reference and inspected image is required - Time consuming
[33]	INCC for misaligned components	Detection accuracy	>96%	- Smaller false alarms - Four times faster than conventional NCC - Weaken the influence of changing illumination	- Reference image is required - Each image contained only one missing component - Time consuming, takes several seconds for a large image - Same size and orientation of reference and inspected image is required
[36]	SURF technology for missing components	AROC curve	>95%	Localise and identify missing components regardless of PCB position, rotation	- Reference image is required - Time consuming, the inference time for one image can reach 6-7 seconds - Deeply depends on local pixel information

* Note: NA indicates that information is not available in the references.

improvement in detection accuracy. In fact, this method is far from real-time detection. The inference time for one image can reach 6-7 seconds, of which SURF register processing takes more than 2 seconds. Other image processing operations such as dilation, binarization and XOR comparison also consume a lot of time. Furthermore, SURF algorithm deeply depends on the gradient direction from local pixels, which may lead to inaccurate main registration direction. For tiny surface mounted PCBs, large amount of components crowd on the surface to result in a more complicated feature space. Only local pixel information instead of global contexts may not able to correct image shifts and extract enough features, which will give rise to a bad detection performance.

According to previous description, traditional image processing-based PCB defect detection works are tabulated in Table 3. Although most of them achieved PCB defect detection successfully, there are large room for improvements. In traditional methods, feature extraction mainly relies on artificially designed extractors, which requires professional knowledge and complex parameter adjustment process. Meanwhile, each method is targeted at specific

applications, with poor generalization ability and robustness. The main limitations can be summarised as follows:

- Most traditional machine vision-based methods for defect detection in PCBs used image subtraction or image matching to obtain defect regions. This means the qualified template image is essential. The general problem they must face is that a perfect reference image is always very difficult to acquire. For different PCBs, different reference images have to be prepared, which is time consuming and expensive.
- The detection performance highly depends on the quality of the reference image. Hence, they are not robust to the changing environment such as illumination and noise. Even small shadows can give rise to false alarms. Besides, they are not generalized to different samples. For components whose colour are similar with background or surroundings, the detection accuracy could be terrible.
- Both of them did not reach real-time detection at all, especially for high resolution inspection images from

real PCB industry. Many additional image processing procedures except image subtraction are necessary to obtain a high accuracy, which inevitably decrease the inference speed.

- The size and orientation between reference and inspection image must be same in pixels entirely, which commonly needs highly precise calibration equipment and trained manpower. Some works have introduced novel approaches such as averaging inference image, clustering pre-processing and SURF to eliminate the affection caused by image shifts, but the whole detection algorithm get more complex and entailed larger computation costs and more parameter choices than before.

It is difficult for traditional image processing to build a complete rule to achieve automatic detection. Sometimes there is a complex combination of various algorithms with too many parameters and manual adjustment. However, seeking the right and suitable parameters is too tedious. One possible solution may be machine learning that computers learn a set of automatic judgment criteria from enough data by extracting a certain number of deep features. Then, the computer will select a best “hypothesis” from the hypothesis set through what it has learned. Hence, the basic image processing operations such as edge detection, morphological analysis, Hough transform, various image thresholding techniques and other feature representation techniques, can be used as a front end for machine learning classifiers (i.e. SVM [9], K-Nearest Neighbor [38], NNs [10]) to construct more in-depth detections. The integration of traditional image processing and machine learning can build a more robust detection algorithm with high inference accuracy and speed than single one of them.

V. MACHINE LEARNING-BASED DEFECT DETECTION

Image subtraction-based traditional image processing methods can achieve high detection accuracy for PCB defects. However, using these methods is always time consuming, and many reference images are essential and can affect the detection results. Many researchers have proposed several machine learning-based algorithms to overcome these difficulties. Many classical machine learning algorithms have been developed and used in PCB defect inspection. In this section, several machine learning algorithm-based methods that mainly focus on solder joints and components are introduced and discussed.

Yun et al. [39] used SVMs and a tiered circular illumination technique to inspect the PCB solder joints. The illumination equipment consists of a Charge-coupled Device (CCD) camera with three circular lamps whose colours were blue, red and green. The solder joint surface has the same nature as the mirror surface. Three lamps illuminate the surface of solder joints at different angles, and then the CCD camera can receive reflected light with different intensity patterns due to the difference in surface slope. The visual information of the solder joints surface can be inferred from the highlighted

patterns and images. Then, six characteristic features based on average intensity value and percentage of highlights of images from three colour frames were extracted. Finally, each solder region represented by a six-dimensional feature vector was fed into an SVM layer that consisted of four SVM classifiers to output one of several pre-defined types, including excessive, good, insufficient and no solder.

In [39], 402 solder joints were collected to train and test the model. All the insufficient and no solder defects were detected correctly, and for the excessive class and good class, the detection rates were 96.07% and 98%, respectively. This model was compared with other methods, including K-means classifier [40] and backpropagation (BP) classifier [41], to show its better performance. However, this paper just achieved solder joints classification without localisation as SVM is a classifier only. The solder joint images with a size of 50×100 pixels instead of whole PCB images were fed into the model. This means that in real PCB defect applications, solder joints images must be extracted firstly, which is a tedious and inefficient process. Moreover, how to choose the hyperparameter C and kernel in SVM is a tricky problem. Meanwhile, the SVM is difficult to be implemented for large-scale training samples, because it uses quadratic programming to solve the support vector, in which the calculation of M-order matrix will be designed. When the order of matrix is large, it will consume huge machine memory and operation time. In addition, this method required complex three-colour circular illumination system which is not easy to operate.

Ko and Cho [42] combined an Learning Vector Quantization (LVQ) neural network [43] and fuzzy logic [44] scheme to inspect solder joints. The solder images were also obtained by a three-colour circular illumination system. Firstly, the obtained solder images were divided into three sub-images (right, centre and left columns in the longitudinal direction of a solder joint), which were each fed into three LVQ classifiers. The purpose of this operation was to decrease the burden and computation, as increased class number decreases the classification performance markedly [45], [46]. The output values of nodes for each classifier were marked from 1 to 10 for right and left columns, and from 1 to 8 for the centre column based on previous studies [47], [48]. Then, the input images were labelled by an expert into five classes based on their experience. Compared with the linguistic approach, the LVQ classifier has difficulty distinctly categorising complex boundaries among various categories. Thus, a fuzzy rule was proposed based on experts' knowledge to obtain better classification results.

The performance of this method in [42] was tested by using 96 images. The total success rate achieved 95.83% among five solder joints types, which was much higher than that of the original LVQ classifier. Similarly, limited by the computing power, this method also got single solder joint images instead of entire PCB images as inputs, which is definitely not suitable for real PCB industry. A complex three-colour circular illumination system was implemented to obtain color

patterns as well. Besides, this method was neither a fully automatic nor efficient detection system, because it highly depended on the expert's experience to make a decision in the process of fuzzy logic scheme. Furthermore, this work only classified the solder joint types without localisation yet because of the original limitation of classifiers.

Belbachir et al. [49] proposed a PCB defect inspection system using wavelet transform (WVT) [50] and multi-layer perceptron (MLP) [51] NNs. Complicated illumination lamps were replaced by a common lighting source. The CCD camera was moved instead of the PCBs so that the system was more flexible and could be placed on an automatic production line. A databank that stored all the components and corresponding defect images to detect was built. A reference board was obtained by assembling images of the components from the database based on the circuit layout. A defective board was obtained by overlapping the defect image with the qualified images of components. For the training path, a series of NNs associated with specific components was trained using a set of images from the databank to recognise the corresponding defects that could appear. For the testing path, template matching was applied to identify the reference point, where the CCD was moved to frame the region of interest (RoI). Then, the WVT was implemented on each captured image to extract features, which were fed into a corresponding trained NN. The outputs of the NN displayed the defect class.

With the system in [49], the authors achieved missing component defect classification and faulty solder joint defect classification. Compared with previous works, this method did not require complex illumination system at all. However, a database storing all the components and corresponding defect images must be constructed firstly, which has to take large amount of time and labour. The training dataset was synthesized by components in the database according to the given circuit scheme. The defective circuits (i.e. missing component) could be obtained by overlapping the component with an area having the same color as the background. All of these complicated pre-processing operations make the detection not easy to be deployed in real PCB defect inspection industry. In addition, template matching was implemented in this method leading to the delay of detection speed. More concrete classification metrics and comparison with other methods were not exhibited in the paper.

On the basis of MLP NN and geometric-wavelet (GW) feature extraction, Acciani et al. [52] presented a novel solder joint defects inspection method. The method consists of four steps: image acquisition, image pre-processing, feature extraction and classification. The image that contained more than one board was acquired by a CCD camera and then processed by three image segmentations. To extract the target central board, template matching with a real reference image was applied using the horizontal energy and the vertical energy vectors of a greyscale image. Then, the central board from the previous procedure was matched with

a real IC image to obtain the IC region in the target image. Similarly, the solder joint images were acquired by template matching using the corresponding reference image. In the feature extraction step, geometric features [53] and wavelet features [54], [55] were used together to form the third feature type named GW features. Finally, these features were sent into the MLP NNs to obtain the classification results.

The method in [52] achieved overall 98.8% classification rate among five solder joint defect types, which was bigger than that of the method that used single geometric feature [56] or wavelet feature. The performance of this method was also progressive compared with that of the LVQ NN with GW features. However, this method used reference PCB images to get target solder joints through several image matching steps, which represents a mass of computations, causing a low detection speed. Obviously, it must have the limitations same as previous traditional image processing-based methods that require reference images. Besides, it is very difficult for MLP to select the number of hidden nodes. The authors applied n-fold cross-validation strategy to decide the hidden neurons without regard to the time cost. Furthermore, only defects classification without localisation was executed in this work.

Crispin and Rankov [57] achieved PCB component detection by integrating GA and template matching. A generalised template of the resistor was formed by linearly combining six template images of the same size to average their statistical variation of grey-level intensity values. The common operation of template matching involves sliding the template image on every pixel in the target image to calculate the similarity using normalised NCC, which is computationally time consuming. To reduce the computation, the authors applied Canny [58] edge filtering algorithm to the target image so that the search was constrained to edge positions to decrease the sliding space. Histogram equalisation was applied before edge detection and forming template images. Then, GA was used to search template matches. As GA is skilled at providing a single solution, the author designed a mechanism to allow it to capture more targets. When a region in the target image was considered matched, the position was replaced by an inverse template image so that future correlation values in this position were very poor.

The method proposed in [57] achieved resistors detection successfully, and the running time was about 39.5 s with a population size of 160, recombination percentage of 75% and mutation percentage of 35%. Reference images were required in this method. Template matching or NCC was applied in this method, indicating a huge computation cost and low detection efficiency. Moreover, the selection of parameters for genetic algorithm seriously affects the quality of the solution, but at present the choice of these parameters is complicated and mostly depends on experience, resulting in a long processing time. Besides, it is easy for GA to fall into precocity situation and to converge to the local optimal solution, which means the detection ability of the algorithm for new PCB data space is limited.

Cai et al. [59] formulated the solder joint inspection issue as an optimisation problem using robust principle component analysis (RPCA) [60], [61]. IC solder joint images were obtained using a three-colour circular illumination system similar as that in [39] and [42]. On the basis of RPCA theory, the hue channel of a colour solder joint image, which contained major information for colour perception [62], was decomposed into a low-rank component and an error component. An appearance model was built by averaging low-rank components of qualified solder joints images. Then, according to prior human knowledge and appearance, defect score was defined to evaluate the quality of inspected images. A discrimination scheme was applied to make the final decision for a qualified or unqualified image by setting a threshold. The threshold was determined by the maximal value among the defect scores obtained at the iterative stage of building the appearance model.

In the work [59], the best performance of the inspection of unqualified solder joints was achieved when at least 350 qualified solder joint images were used to build the appearance model. In that case, the precision, recall and F-measure were 95.65%, 100% and 97.77%, respectively. However, this method only achieved qualified and unqualified classification without localisation. In addition, the building of an appearance model was essential in this method, resulting in a very slow speed because of the large computation costs of singular value decomposition for the data matrix at each iteration in the RPCA process. It is sensitive to very small rotations, as this method was based on the hypothesis that solder joint images of the same type have the same size without rotation. Furthermore, a complex three-colour illumination system was required, and prior human knowledge was incorporated into the defect score to achieve better inspection performance. All of these limitations indicate that this method is not generalized and not suitable for real PCB inspection applications.

Hongwei et al. [63] used improved AdaBoost [64] and DT to classify six types of solder joint defects. The whole inspection process consisted of the training stage and the test stage. In the training stage, the solder joint image was firstly divided into several sub-regions, which had different colour distributions for defective solders. 30 geometric features based on three colours were extracted in each sub-region. Then, an improved AdaBoost was proposed to seek three necessary optimal features according to their classification performances. Finally, the classifier was built for each sub-region after training. In the detection stage, three steps (sub-region division, critical feature extraction and classification for regions) were executed based on the training results. For defect diagnosis, a new DT that combines classification and regression tree was implemented to obtain the final defect class.

The method proposed in [63] achieved an overall classification rate of 97.3% without localisation. The average time for one chip was 8.6 milliseconds, which was slower

than image comparison method. In this method, only the chip images with a small size instead of whole images were fed into the classification algorithm, and one component contained only one solder joint defect. Thus, this method will be still time consuming when detecting an entire PCB image with various types of components. Similar to the limitation of previous works, a three-colour hemispherical circular illumination was required to infer the characteristics of the solder shape. The improved AdaBoost was used in this work to select fewer but optimal features. But two thresholds for each weak classifier were introduced and a threshold of relevance information was also applied to guarantee the selected classifier brought new information. Extra threshold parameters caused more computation cost, resulting in a slow classification speed.

Wu et al. [65] from the same group in [63], achieved solder joint classification using Bayes and SVM. Similarly, a three-colour illumination system was used to obtain solder joint images. Six colour features and one template matching feature were acquired in the component body and solder pad, respectively. Then, feature selection on the basis of information gain was used to identify the optimal features from these 14 features. The selected features were fed into a Bayesian classifier [66] to determine whether the solder was qualified or unqualified. For the unqualified solder joints, several SVM classifiers were applied to obtain the defect type.

In the work [65], the best performance of 100% classification rate was achieved using 7 features for both Bayesian and SVM. However, this method was a two-stage classification algorithm, including Bayesian for binary classification and SVM for multi-classification. The detection focused on only one component image with one defect instead of a PCB board. Besides, the template matching was used by calculating the NCC between template and search images. Due to previous limitations, it must be huge computation, especially for a whole real PCB image. Similarly, this method only accomplished defects classification without positioning yet.

Different from [65], the same authors, Wu et al. tried to combine BP network and GA to inspect solder joints in [67]. The processes of solder joint images acquisition and features extraction were the same as in [65]. Differently, GA was used to select 6 appropriate features, which were sent into the BP [68] network for training by calculating mean square error.

The overall classification rate of 98.46% for 6 types of solder joints was achieved by the algorithm proposed in [67]. This method was compared with DT and SVM to prove its good performance. Same as [63] and [65], a three-colour hemispherical circular illumination was necessary in this method, and the detection focused on only one component image with one defect instead of a PCB board. Moreover, the template matching was also used by calculating the NCC between template and search images. Thus, this method was still distant from real-time detection. In addition, the classification rates for cold solder were not high enough at only about 90%.

Wu and Zhang et al. [69] used colour grads and Boolean rules to inspect solder joint defects. The solder joint images were acquired by the same illumination as [63], [65], and [67]. On the basis of the colour distribution and the solder land size, several regions were defined in the solder joint image. Then, the area features, the barycentre features, the distribution features and the colour grads' features were defined and extracted. Models of 8 different solder joint types, namely, acceptable solder joint, lacked solder, excessive solder, pseudo-solder, component shifted, component lift and tombstone, no solder and missed component, were built according to their appearances and previous feature analysis. The relationship between the solder joint types and regions was derived, and Boolean rules were built to classify solder joints.

The method proposed in [69] achieved an average classification rate of 97.7%, which was higher than that of other approaches in [52], [70], [71], and [72]. However, the prerequisite is that judging criteria must be constructed according to the solder joints' appearances and various feature analysis. This process deeply depends on the expert experience and prior knowledge, which is time consuming inevitably. Thus, this method is not suitable for large-scale applications in real PCB inspection industry. Furthermore, the requirement for illumination system is critical to the construction of rules as different illumination situation will bring large errors into the forming of criteria. The running time reached 11 seconds for one PCB, which was also much faster than other references in the paper, but it was still far from real-time detection.

Li et al. [73] proposed a semantic segmentation method to inspect components in the PCB by using a random forest pixel classifier. In this work, depth images of PCBs were used as inputs due to its high robustness to the influence of the illumination and texture of the target [74]. Depth images were applied to detect human body pose [75], [76], [77] and hand gesture [78], [79], [80], [81] by pixel classification approaches. A three-dimensional model of the PCB was used to synthesise the training set by computer graphic rendering, which contained the depth images and corresponding colour-labelled images. Pixel samples in the depth images were selected randomly, and their depth difference features were extracted. Pixel classification labels were extracted from the colour-labelled images.

Then, the difference features were used to train a random decision forest classifier [82], [83], which built a mapping relation between the difference features and the pixel classification labels. Finally, the semantic segmentation of PCB components was achieved by the trained random forest pixel classifier. Missing or misplaced components were also detected by this method. The prediction of the depth images with missing or misplaced components was compared with the colour-labelled images of qualified PCB. Then, the missing or misplaced component can be detected by calculating and analysing the accuracies of each component.

For the method in [73], the overall component recognition accuracies were 98.96% and 83.64% on the synthetic and real depth images, respectively. This method performed poorly for real PCB depth images, maybe because the real PCB images have more complicated layout and various noises, which can make the random forest pixel classifier overfitting. Additionally, this method could only classify defective or non-defective caused by missing or misplaced without the exact defect types. Another limitation is that several experiments must be performed to choose the optimal training parameters and random seed for the random forest classifier. Hence, this method must be ineffective for real component defects recognition.

Machine learning algorithm-based defect detection works are tabulated in Table 4. Most of them focused on solder joints or components defects detection and achieved good performances successfully. However, large gaps between them with real PCB inspection application in the industry still exist. In machine learning-based methods, features sent into the machine learning algorithms are mainly extracted by traditional image processing operations such as edge detection, morphological processing, various image thresholding techniques. Hence, the detection performance of machine learning-based methods deeply depends on the quality of previous extracted features. Meanwhile, each method was designed for special applications, indicating a poor generalization ability and robustness. The main limitations can be summarised as follows:

- All the proposed methods only achieved classification without localisation. Then main reason is that most machine learning algorithms were developed for classification issues. The localisation of defects can be achieved by adding several image post-processing operations.
- Image pre-processing were required before the final classification process, and some methods required a qualified template to locate the reference positions or regions of interest. Similar to traditional image processing-based methods, they also have difficulty acquiring reference images, which is time consuming and complicated. Their performances were deeply affected by the reference image and prior image processing.
- Another common limitation is that some of them used a complex three-colour illumination system to obtain target images with abundant features, which was hard to implement and thus expensive.
- Due to the influence by many image processing steps and the choices of various parameters in machine learning algorithms, all of these methods cannot perform real-time inspection of a PCB board automatically, although some of them achieved high-speed inspection for one component.

Obviously, machine learning-based PCB defects detection methods cannot discard traditional image processing as a

front end. We still need to identify and manually code application features based on data types (i.e. pixel value, shape, orientation). Moreover, many machine learning methods cannot handle large-scale data rapidly. Some traditional machine learning algorithms need to solve constrained optimization problems or complex matrix calculations, leading to high time complexity. Some algorithms need to calculate the distance between all sample points, which have high space-time complexity. More importantly, many traditional learning algorithms are difficult to combine with Graphic Processing Unit (GPU) parallel computation. So they cannot deal with large-scale data efficiently such as real complicated PCB images with high resolution.

To overcome these limitations, deep learning-based algorithms is a good choice. Various CNN models have been used in object detection, image segmentation etc. with desirable detection accuracy and speed simultaneously. Features are not extracted by manual designed extractors, but extracted by CNNs automatically. In addition, they are robust to the environment and noise. Thus, PCB defects inspection also can be considered as an object detection issue in nature, which is possible to be overcome by deep learning methods.

VI. DEEP LEARNING-BASED DEFECT DETECTION

Recently, deep learning-based methods, especially CNNs, have been widely used in image processing [84], object detection and segmentation [85]. Unlike traditional machine vision methods, CNN-based approaches are able to automatically extract image features, simplifying the image pre-processing process so that the detection accuracy and speed can be promoted effectively. In addition, CNN-based methods are robust to the environment and noise. Even though shadows or reflections exist, significant object detection outputs can still be achieved because of these methods' multi-level feature extraction ability. With these advantages, CNN-based object detection algorithms have outperformed competing algorithms on different datasets and have become the primary driving force for the development of object detection. Therefore, deep learning algorithms have attracted the attention of researchers for PCB defect detection and achieved good detection performance.

The most frequently used and modified CNNs to detect PCB defects are Faster Region-based Convolutional Network (Faster R-CNN) [86] and You Only Look Once (YOLO) series. Faster R-CNN is a two-stage object detection model that was developed based on R-CNN [87] and Fast R-CNN [88]. YOLO series is a one-stage object detection network that has both a high detection speed and good accuracy. In this section, some CNN-based methods proposed by researchers for defect detection in PCBs are presented and discussed.

A. CNN-BASED METHODS FOR SURFACE DEFECTS

Ding et al. [89] proposed a tiny defect detection network for PCBs named tiny defect detection network (TDD-net) based on Faster R-CNN. As the PCB images acquired by industrial

camera always have a high resolution and the defect regions always occupy a very small portion compared with the whole image, the traditional Faster R-CNN, which produces anchors using 3 scales and 3 different ratios, is not suitable for tiny defect detection. Thus, inspired by YOLOv2 [90], the author applied k-means clustering to the PCB training dataset bounding boxes to seek reasonable anchor scales automatically. Then, data augmentation methods, including adding Gaussian noise, changing light, image rotating, flipping, random cropping and shifting, were implemented. To maintain the features that have low semantic level, the architecture of Feature Pyramid Network (FPN) [91] was adopted. In FPN, the features with low resolution and strong semantics are connected with features that possess high resolution and weak semantics by lateral connections from top to down, and it predicts features on each level of the pyramid network. To mitigate the impact on the detection accuracy caused by the small and imbalance dataset, online hard example mining [92] was applied in the whole training phase to improve the quality of RoI proposals.

In the work [89], 693 images with 6 different defects, namely, missing hole, mouse bite, open circuit, short, spur and spurious copper were synthesized. One image contained several defects of the same type. After training and testing, TDD-net achieved a better performance with 98.9% mean average precision (mAP) than other state-of-the-art methods, including Faster R-CNN with different backbones and FPN. This work achieved excellent detection mAP benefited by the fusion of multiscale features using FPN. However, this algorithm is a two-stage frame and has huge parameters, meaning that the detection speed is obviously slower than that of one-stage network. The use of the full connection layers took up a large amount of parameters. RoI pooling was still used in this work, resulting in the loss of translation invariance of subsequent network features, which would affect the final positioning accuracy. Moreover, the whole retained RoI proposals by the RoI pooling layer will go through the full connection layer and be calculated separately instead of shared calculation between them, leading to huge repetitive computations. All of these shortcomings makes this method not able to achieve real-time detection in real PCB industry.

Adibhatla et al. [93] used the YOLO model to detect defects in PCBs. In order to achieve fast and accurate detection simultaneously, tiny YOLOv2 [94] network modified from YOLOv1 [95], was introduced in this method. The backbone of YOLOv2 is a Darknet-19 network with 19 convolution layers and 5 MaxPooling layers, which is smaller than traditional Visual Geometry Group (VGG) [96] network with comparable accuracy. The floating-point computation of Darknet-19 is reduced to about 1/5 of VGG-16 to allow for faster arithmetic. In tiny YOLOv2, the Darknet network is clipped to less convolutional layers to get faster inference speed. Besides, batch normalization layers are added behind each convolutional layers to make the model converge rapidly.

TABLE 4. Summary of machine learning-based methods.

Work	Concept or approach proposed	Detection criterion	Performance	Advantages	Limitations
[39]	SVMs and a tiered circular illumination system	Classification rate	>96%	High classification rate	- Achieved classification without localisation - Complex illumination - Choices of the kernel and hyperparameters in SVM - Focused on small solder joint images only
[42]	Combined LVQ NN and fuzzy logic scheme	Classification rate	95.83%	A promotion than original LVQ	- Achieved classification without localisation - Complex illumination - Focused on small solder joint images only - Highly depended on the expert's experience
[49]	Combined WVT and MLP NNs	NA	NA	Common lighting source	- Only classified missing components and faulty solder joints
[52]	MLP NN and GW feature extraction	Classification rate	98.8%	High classification rate	- Required template matching and reference images - A database must be constructed firstly - Required several image matching and inference images
[57]	Combined GA and template matching	NA	NA	Achieved resistor detection successfully	- Achieved classification without localisation - Hard to choose the number of hidden nodes in MLP
[59]	RPCA	Precision, recall and Fmeasure	95.65%,100% and 97.77%	High classification rate	- Required template matching and reference images - Time consuming, running time of 39.5 s - Selection of parameters - Complex illumination - Sensitive to very small rotation - Appearance model must be built, time consuming
[63]	Combined improved AdaBoost and DT	Classification rate	97.3%	Higher success rate than original AdaBoost	- Achieved classification without localisation - Time consuming - Complex illumination - Only focused on the chip images with a small size, and one component contained only one solder joint defect
[65]	Combined Bayes and SVM	Classification rate	100%	High classification rate	- Classified one component with one defect, instead of a PCB board - Complex illumination - Required template matching and reference images - Achieved classification without localisation - Time consuming
[67]	Combined BP NN and GA	Classification rate	98.46%	High classification rate	- Classified one component with one defect, instead of a PCB board - Complex illumination - Poor performance for cold solder - Required template matching and reference images - Time consuming - Achieved classification without localisation - Complex illumination
[69]	Used colour grads and Boolean rules	Classification rate	97.7%	Classified 8 types of solder joint defects	- Time consuming, reached 11 s for one PCB - Highly depended on expert experience and prior knowledge
[73]	Random forest pixel classifier for component recognition	Detection accuracies	98.96% or 83.64%	High recognition rate for synthetic images	- Poor performance for real images - Only classified defective components without the exact defect types - Hard to choose the optimal parameters and random seed - Time consuming

^a Note: NA indicates that information is not available in the references.

For the model in [93], 11000 images with 11 defect types were collected for training. The detection accuracy of 98.82% for PCB cosmetic defects was accomplished. This work just used tiny YOLOv2 to execute PCB defects detection without modification or innovation for the algorithm. In addition, all the defect types were labelled as a single type of defect, which means that the regions were labelled as either not good or good, reducing the detection difficulty. More performance metrics such as F-score or mAP were missing in this work, and the good detection accuracy only was achieved by small local images cropped from whole PCB images, which make it not very convincing for real PCB inspection application.

Hu and Wang [16] proposed a PCB surface defect detection method based on improved Faster R-CNN and FPN. Faster R-CNN with the backbone of Residual Neural Network (ResNet50) [97] was adopted for the input image as the detector. FPN was used to fuse deep and shallow features, which is considered a powerful way to promote the detection accuracy of small objects. Shuffle V2 [98] residual units were implemented to decrease the computation of the whole

network. The Guided Anchoring Region Proposal Network (GARPN) algorithm was used to produce more accurate anchors, and then RoI pooling layers were executed to obtain proposals. Lastly, the fully connected layers were used to classify and perform bounding box regression to accomplish the final defect detection.

Data augmentation was used in [16] to obtain a total of 12000 defect images. This method achieved a detection mAP of 94.2% which was better than that of other state-of-the-arts including Faster R-CNN [86], RetinaNet [99] and YOLOv3 [100]. Although the introduction of Shuffle network reduced the computation cost, but the running time of this method reached 0.078 seconds per image which was longer than YOLOv3 with the backbone of MobileNet [101]. The reason is that this model was still a two-stage detection model based on Faster R-CNN. Furthermore, this model struggled to detect open defects due to the open defect in the training set appeared in diverse characteristics. This model accomplished faster detection than original Faster R-CNN by some useful tricks, but it did not reach the real-time detection yet.

Liao et al. [102] used YOLOv4 [103] with an improved backbone network to detect surface defects in PCBs. In this model, Darknet53, which has higher memory consumption in YOLOv4, was replaced by the lightweight network MobileNetV3 [104] to speed up the detection. The MobileNetv3 uses depth-wise separable convolution to capture feature maps and includes two steps: depth-wise convolution and pointwise convolution. The number of parameters in depth-wise separable convolution is much less than that of normal convolution, making the whole parameters is about 40% less than original YOLOv4.

The model proposed in [102] produced the highest detection accuracy of 98.64% mAP than that of Faster R-CNN [86], RetinaNet [99], Single Shot MultiBox Detector (SSD) [105], YOLOv3 [100] and YOLOv4 [103]. Significantly, the inference speed was satisfactory at 56.98 frames per second by RTX3080 GPU. However, the numbers of each defect type were maintained manually to be approximately equal so that the dataset was balanced in samples, which is contributed to the good performance. Imbalanced samples from real PCB industry usually cut down the detection metrics of common CNN models. Besides, each image in the dataset contained only one defect. Thus, the ability of the method to detect various defects in one image was not proved. Furthermore, the real-time inference speed was achieved for cropped images with small size of 416×416 pixels, instead of a whole PCB images. The size of a whole PCB image in this work is about 70 times larger than cropped one. Hence, the detection speed of an entire image must be brought down sharply.

Adibhatla et al. [106] applied YOLOv5 [107] to detect PCB defects because of its good structure, high speed and smaller size compared with previous YOLO versions. In this experiment, three different size YOLOv5 models (small, medium and large) were used. Mosaic data augmentation, splicing four images by random scaling, random cropping and random arrangement, was implemented in this method to rich the dataset and decrease the load of GPU with a small batch size. The backbone of this model was improved by inserted the Cross Stage Partial Network (CSPNet) [108] into Darknet53, making it lightweight while maintaining accuracy. Additionally, Path Aggregation Network (PANet) [109] was combined with FPN to fuse various level features, which should improve the detection precision, especially for small size targets.

In the work [106], all the three different size YOLOv5 models achieved high detection precision. The large YOLO-v5 model had the best detection accuracy of 99.74%. Nevertheless, this work just applied YOLOv5 models to carry out PCB defects detection task without modification or innovation for the algorithm. Only defective or non-defective targets were distinguished. The exact types of defects were not classified and the defect confidence scores were not high. Moreover, similar to previous works, in order to reduce the computation cost, this article only detected defects from cropped images with small size of 400×400 pixels, instead

of a whole PCB images. Hence, inspection engineers must seek a balance point between the detection precision and speed when the model is applied for a whole PCB image with high resolution.

Zhang et al. [110] modified the Fully Convolutional One-Stage (FCOS) [111] network to achieve surface defect detection in PCBs. The authors replaced the backbone network of ResNet101 [97] with a light network of MobileNetV2 [112] to decrease model parameters. The convolutional block attention module [113] was applied in MobileNetv2 to promote the feature extraction performance. PANet-based FPN was implemented instead of traditional FPN in the model of FCOS. Besides, the bounding box regression loss function was replaced with distance-intersection over union (IoU) loss to boost the detection accuracy for smaller defects.

This model proposed in [110] achieved a detection mAP of 44.3%, which was much higher than that of the traditional FCOS model. The detector achieves proposal free and anchor free, significantly reducing the number of parameters. In addition, by eliminating the anchor frame, the detector completely avoids complex IoU calculations and matching between the anchor boxes and the ground truth boxes during training, which reduces the total training memory footprint by about two times. Thus, the speed was increased by 30% to 37.6 milliseconds inference time compared with that of Faster R-CNN in this work. But, only 1455 images based on the dataset in [89] were collected for training by this model. As a result, this model had a bad mAP of 44.3% for high resolution image, which was 5% lower compared with Faster R-CNN due to the abandon of anchors and the lack of data.

In the real PCB industry, the defect class distribution is always imbalanced. Few previous deep learning models concentrated on this situation. Zhang et al. [114] proposed a cost-sensitive residual convolutional neural network (CS-ResNet) to overcome the problem of imbalanced data. In the study, a cost-sensitive adjustment layer was applied to the fully connected layer of ResNet [97]. After the softmax function was applied, a weighted loss function was obtained based on the imbalanced degree of the training dataset. The weight w_t was calculated by Equation (2), where α is an adjustment factor, which is a constant greater than or equal to 1.0, cnt_r is the number of real defects, cnt_p is the number of pseudo defects. Then, the weight w_t was involved in the calculation of softmax cross-entropy loss function, which means that the imbalanced degree of the dataset was considered part of the loss function.

$$w_t = \alpha * (cnt_p)/(cnt_r) \quad (2)$$

The introduction of weight w_t in [114] mitigated the effects of imbalanced dataset. The proposed CS-ResNet achieved a better performance than standard ResNet and other CNN-based methods. The sensitivity of CS-ResNet was 0.89, which is a 12% increase over that of ResNet. One of the drawbacks of this method is that a suitable adjustment factor α needs to be chosen for different datasets with various imbalanced degrees. It performs binary classification successfully,

which is not guaranteed for multi-type defects detection. For real PCB inspection tasks, there will be different kinds of defects. The parameter w_t based on total real defects and total pseudo defects cannot represent the imbalanced degrees of each type, making it hard to get a better performance.

B. CNN-BASED METHODS FOR SOLDER JOINT AND COMPONENT DEFECTS

To overcome the issue that supervised learning algorithms always need thousands of annotated images, Dai et al. [115] proposed a novel method based on active learning and semi-supervised learning for solder defect detection. YOLOv3 was used to localise the solder joints. The pre-trained VGG-16 model was used as the features extractor. Principal component analysis was implemented to reduce the dimensions of extracted features. Then, an SVM classifier was applied to the labelled data. K-means clustering was implemented to analyse the clustering structure of the whole images. Through a combination of these two diverse learning rules, a section of the unlabelled images can be annotated both automatically and manually. Then, the classifier was executed again based on the augmented labelled dataset until the stop criteria, resulting in a good classification performance finally.

In work [115], the YOLOv3 model achieved solder joints localisation with a mAP of 95.1%. Then, during the classification phase, a partition of unlabelled samples were annotated automatically with a small error rate of less than 1.5%. Overall, this model achieved a detection precision of 95.16% and 87.68% for insufficient and shifting solder datasets, respectively. However, this method decomposed the detection task into localisation and classification respectively, which consumed more time than one completed CNN-based model. The localisation speed was about 0.34 seconds per image with 1024×1024 pixels, which was not fast enough. The classification phase by SVM and clustering also was not efficient. Besides, this method concentrated on only two defect types of solder joints: insufficient solder and shifting solder, which is not appropriate to real-world solder joint inspection tasks. Furthermore, the number of selected clusters was a tricky problem in this method, making it not generalized to other datasets perfectly.

Li et al. [116] proposed a deep ensemble model combining hybrid YOLOv2 and Faster R-CNN to achieve PCB solder defect detection. The authors developed the retraining process, which was able to utilise the exception data, such as false positive and false negative data, to train the previous network and produce a renewed model with better performance. After several cycles, the model can decrease the false alarm rate significantly to 20%–30%. The rule was that only the defects detected by both Faster R-CNN and hybrid YOLOv2 are accepted as defects and highlighted. This was achieved by setting the IoU between two predicted bounding boxes larger than the IoU threshold.

The ensemble method proposed in [116] achieved a 97.45% detection rate and a 10–15 seconds inference time

for a high resolution image successfully. When this model was applied to another production line, it still obtained good performance after seven retraining processes. However, the retraining cycles must cost huge time and computation resource. The inference speed was also considerably slow, which was far from real-time detection. Besides, during the training process, engineers had to confirm, judge and label the exception data to obtain the retraining data, which means that this method is non-automatic and artificially costly.

Sezer and Altan [117] combined the deep learning network with image matching to detect solder paste defects. Image matching was used to identify the qualified solder joints and unqualified solder joints. Meanwhile, the solder joint pads were extracted by image matching to build the dataset for CNN classifier. For unqualified solder joint pads, containing uncorrectable solder, missing solder, excess solder, short circuit and undefined object, CNN model with the rectified linear unit activation function and AdaMax [118] optimisation method was applied to distinguish them. Common data augmentations such as flipping, rotating, shifting, blurring image were applied expand the dataset.

After image matching, the CNN model achieved 96.4% classification accuracy in 100 epochs, representing good performance. This method was designed for special PCB defects and a direct integration of image matching and CNN classifier. The detection performance highly depended on the output from image matching. In addition, templates were essential for image matching, which is tedious and demanding to be prepared.

Shen et al. [119] proposed a lightweight detection network for PCB components classification (LD-PCB) and a lightweight PCB component character recognition network (CR-PCB). In LD-PCB network, the stem block and dense block from PeleeNet [120] were used to extract features. Inspired by Hu et al. [121], the context-aware RoI pooling method was applied to improve the detection accuracy for small objects. In CR-PCB network, spatial transformer network was applied to reduce the deformation of the raw images, feature sequence extraction layer was utilised to extract features of characters, recurrent neural network layer was implemented to predict the label distribution and attention mechanism converted label distribution into the final recognition result. Based on previous outputs, the information bank that contains component names and characters was established. Then, a defect analysis module counted the information of PCB components and executed template matching or image comparison was implemented to identify defects, such as missing insertion, incorrect insertion, wrong text and missing text.

In the method proposed in [119], the LD-PCB achieved a detection mAP of 85.8% within 0.036 seconds per image. The CR-PCB achieved a recognition mAP of 96.7% within 0.018 seconds per image. The whole system achieved a defect detection accuracy of 95%. This detection strategy was a three-stage algorithm containing component classification, text recognition and image comparison successively. More

TABLE 5. Summary of CNN-based detection models.

Work	Concept or approach proposed	Detection criterion	Performance	Advantages	Limitations
Works focusing on surface defects					
[89]	TDD-net	mAP	98.9%	- High mAP - Detects 6 defects	- One image contained only one type of defects - Not real time yet - Two-stage methods
[93]	Used the YOLO model to detect defects	Detection accuracy	98.82%	High accuracy	- Not sufficient for unbalanced datasets - Only located defective regions without classification - Focused on small local images cropped from whole PCB images
[16]	Improved Faster RCNN and FPN	mAP, Recall	94.2%, 82.5%	One image contained several kinds of defects	- Not real time yet - Cannot obtain a good result for open defects - Low recall
[102]	YOLOv4-MobileNetv3	mAP, F1-score	98.64%, 97.83%	- Real time - High mAP	- Each image contained only one defect - Focused on cropped images with small size from whole PCB images - Designed balanced dataset manually
[106]	Applied YOLOv5 to detect PCB defects	Detection accuracy	>99%	- Real time - High accuracy	- Only located defective regions without classification - Only detected defects from cropped images with small size
[110]	Modified FCOS	mAP	44.3%	Almost real time for high resolution image	- mAP was lower than that of Faster R-CNN and other models - Small volume data
[114]	CS-ResNet	Sensitivity	89%	- Reduce the influence of imbalanced dataset - Greater sensitivity than ResNet	- Difficult to choose a suitable adjustment factor α for different datasets - Performed binary classification only
Works focusing on solder joint or components					
[115]	YOLOv3 and clustering-based active semisupervised classification	Detection precision	90%	Reduce the workload of labelling	- Not real time - Only detected two defect types - The number of selected clusters was a tricky problem
[116]	Hybrid YOLOv2 and Faster R-CNN with retraining process	Detection precision	99.99%	High precision	- Time consuming, 10–15 s inference time for a high resolution image - Need engineers to be involved - More training time
[117]	Deep learning network and image matching to detect solder paste defects	Classification accuracy	96.4%	Detected six types of solder joints with high accuracy	- Template matching was still required - Two-stage method - Not real time
[119]	LD-PCB and CR-PCB to detect component defects	Detection accuracy	95%	High accuracy for component defects	- Template marching or image comparison was required - Three-stage method - Not real time

* Note: NA indicates that information is not available in the references.

stages made the method performing a high detection accuracy, but increased the processing time considerably. The whole running time must exceed the range of real-time detection as the combination of LD-PCB and CR-PCB had already cost 54 milliseconds per image. The template matching or image comparison also was time-consuming.

The common approaches used by previous works are CNN networks such as Faster R-CNN and the YOLO series. CNN models are able to automatically extract image features and simplify image pre-processing so that the detection accuracy and speed can be promoted effectively. Furthermore, compared with traditional image processing methods and machine learning methods, CNN-based methods are robust to the environment and noise due to their multi-level feature extraction ability. Moreover, in most cases, reference images can be discarded in deep learning models. Deep learning-based defect detection works are tabulated in Table 5. However, these CNN-based methods still need to be improved. The main limitations are summarized as follows:

- Most of them without template matching or image comparison cannot accomplish real-time detection and produce high accuracy at the same time, especially for high resolution PCB images. The main reason is

CNNs should contain more layers so that they are deep enough to extract deep features for precise detection. However, a deeper model means many parameters and heavy computations, leading to much more inference time. A balanced point needs to be achieved or some targets should be sacrificed to attain high detection accuracy or speed. For example, methods of [16] and [89] based on Faster R-CNN achieved good performances in mAP, but they are far away from being able to perform real-time detection, which is important in real industries. References [93], [102], and [106] based on the YOLO series obtained real-time detection with excellent detection accuracy, but the datasets they used are designed specially, or they only achieved easy tasks such as binary classification instead of real detection and localisation.

- CNN-based models always need high volume dataset to obtain good detection performance, especially good generalized ability. But PCB defects are difficult to be acquired as most factories must adopt various measures to ensure quality control. The PCB defect types are also diverse, meaning that a dataset containing all types of defects is impossible to be constructed. Moreover,

labelling the dataset needs experienced expert knowledge and can expend great effort and time.

- For CNN models, a huge volume dataset means longer training time ranging from several hours to dozens of hours using a general PCB dataset. Thus, some methods utilized powerful GPUs to decrease the training time significantly. However, powerful GPUs are costly and is unusual in common PCB inspection equipment.
- Some CNN-based methods only achieved binary classification of defective and non-defective. The best one achieved 6 types of defects detection, which is not enough for real PCB inspection applications.
- CNN-based methods always have large amount of parameters, which are too weight to be deployed in real PCB inspection applications. Hence, the size of CNN models need to be decreased without losing detection accuracy.

CNN models have achieved PCB defects detection with stunning results successfully. In the real applications, the most important issue is how to compress and deploy them in the embedded devices used for inspection. The commonly used methods for model compression and acceleration can be roughly divided into four categories: parameter pruning and quantization, low-rank factorization, transferred/compact convolutional filters and knowledge distillation [122], [123]. Parameter pruning and quantization mainly reduce redundant parameters in the model. Low-rank factorization uses tensor factorization to estimate the parameters of neural networks. The transfer/compression convolution filters are designed with a special structure, which can reduce the parameter space and save memory. Knowledge distillation involves training a larger model and then a smaller model to achieve the same results as the larger model. These four methods are independently designed, but they complement each other. In practical applications, they can often be used together to further compress or accelerate the model for PCB defects detection.

VII. PERFORMANCE COMPARISON AMONG EXISTING METHODS

To make the performance comparison among existing CNN-based methods, a new PCB dataset was built based on 10 high resolution PCB images from the public HRIPCB dataset [89]. Similarly, various defects are produced in each image by Photoshop to build a new dataset containing 800 PCB images with an average resolution of $2,777 \times 2,138$ pixels. Then, the original high resolution images are cropped into 1981 sub-images with the size of 640×640 pixels. Compared with [89], the difference is that at least 2 types of defects exist in each image of our dataset. There are maximum 6 types of defects in each image, including missing hole, mouse bite, open, short, spur and spurious copper. Data augmentation techniques such as flipping, rotation etc. are applied to this small dataset. Finally, our training set with 11093 images and validation set with 2774 images are built, respectively.

TABLE 6. Performance comparison among existing CNN models.

Models	Parameters (M)	Weight (MB)	mAP	Inference speed (frames per second)
Faster RCNN-ResNet50-FPN	41.15	315	90.1%	2
Faster RCNN-ResNet101-FPN	60.14	460	90.6%	1.4
RetinaNet- ResNet50-FPN	36.21	277	94.4%	1.9
RetinaNet- ResNet101-FPN	55.2	422	96.2%	1.4
SSD	24.41	186	86.7%	1.4
SSD-Lite	3.1	24.1	86.0%	9.8
YOLOv3-tiny	8.68	16.6	99.4%	19
YOLOv5-small	7.03	13.8	99.4%	14

Using our PCB defect dataset, we compare the detection performances of existing state-of-the-art models, including Faster R-CNN, RetinaNet, SSD, YOLOv3 and YOLOv5. The comparison results are reported in Table 6. The experiments are run on a general computer with one NVIDIA GeForce GTX 1660 Super GPU. All the detection results are acquired at the input image resolution of 2240×2240 pixels.

According to Table 6, Faster R-CNN with backbone of ResNet101, as a representative of the two-stage object detection models, achieves a mAP of 90.6%. RetinaNet with backbone of ResNet101 performs a mAP of 96.2%, which is 5.6% more than Faster R-CNN. However, the detection speed of RetinaNet has no marked increase. SSD-Lite has the least parameters, resulting in the significant sacrifice in precision. YOLOv3-tiny obtains the highest mAP of 99.4% and fastest speed of 19 frames per second, demonstrating the desirable superiority in PCB defects detection. But all of them do not reach the real-time detection speed, and models are still too weight to be applied in real PCB inspection industry.

VIII. RESEARCH CHALLENGES

A. TRENDS IN PCB DEFECT DETECTION

As stated earlier in this paper, PCB defect detection methods can be mainly divided into four strategies. The first detection strategy is manual visual inspection, which allows experienced experts to capture flaws and types of defects. However, with its low efficiency, high cost and poor robustness, it had become increasingly impractical and is thus being eliminated. After the industrialised AOI machine was developed, many works started to improve the inspection performance of methods based on traditional image processing and machine learning algorithms. More than 60 works were conducted on these two strategies from 1980s to the present because of the ongoing evolution of classical machine learning algorithms. Even recently, some researchers continue to propose novel ideas to modify existing algorithms to enhance the detection accuracy and speed. The last PCB defect detection strategy is deep learning. More than 30 reviews were conducted on this strategy between 2012 and 2022. In the past several years, with the explosion of the use of CNNs for object detection, an increasing number of researchers have become interested in using various modified CNN models for PCB defect detection and other surface inspection of industrial products. The trend of using CNNs to detect defects will continue for a long time due to the rapid development of CNN models and the computation power.

B. RESEARCH CHALLENGES

Traditional image processing-based methods and machine learning-based methods are not fully automatic and real time. Thus, as mentioned in Section VIII-A, deep learning-based PCB defect detection method is an important research focus in the future. Deep learning-based methods completely depend on the computer for automatic detection, which has the obvious advantages of fast detection speed, high recognition accuracy and robust adaptability. Some researchers have applied deep learning methods to obtain acceptable results compared with traditional methods in PCB defect detection, but studies on deep learning methods are still not abundant. The previous review and introduction indicate that several challenges exist for current deep learning methods used for defect detection in PCBs.

1) OPEN PCB DEFECT DATASET

For the supervised deep learning method, few open datasets of PCB defects are available, especially for component defects and solder joint defects. The existing open datasets mainly focus on component detection or cosmetic defect detection with only one type of defect in one image. Having a large amount of images directly contributes to the performance of CNNs. Thus, having numerous defect images is necessary for the training of CNN models. Also, the process of data acquisition and labelling is time consuming.

2) DETECTION WITH BOTH HIGH ACCURACY AND SPEED

For surface defect detection, without subtraction, most deep learning methods classify inputs into two classes only: defective with location identification and non-defective. The specific defect types are not recognised. In other cases, some works are able to detect defect location and type identification, but the dataset was designed specially, which means that one image contained only one defect or several defects belonging to one type. Even though a few models such as TDD-net have achieved defect location and type identification, these models should balance the detection mAP or accuracy and speed. Thus, the difficulty lies in achieving high detection mAP and speed simultaneously. In addition, the number of types is limited to about six, which also needs to be increased.

3) TINY COMPONENT DEFECT DETECTION

With the development of micro-electronic technology, the assembled PCB is becoming increasingly smaller. Hundreds of tiny components can be mounted on a small board, which is difficult for human eyes to inspect. Component defects may include missing and misaligned components. Some researchers have focused on the detection of component types. However, few studies focus on component defect detection only depending on the deep learning method without image subtraction or matching. One of the possible reasons is that collecting PCBA component defects is more difficult than collecting cosmetic defects. An assembled PCB

always has a mass of parts such as resistors, capacitors and chips, thus being a more complicated detection environment for CNN models. Image matching is always applied to reduce the inference time.

4) SOLDER JOINT DEFECT DETECTION

For deep learning-based solder joint defect detection, current methods always rely on outside assistance, such as image matching or registration and manual confirmation or inspection during the detection process, which means that they are not fully automatic. Without assistance, they can only detect the qualified and defective classification with good performances, thus indicating that the detection classes are limited to a small range, which does not coincide with the real industrial manufacturing process.

IX. FURTHER RESEARCH DIRECTIONS

A. OPEN PCB DEFECT DATASET

Having a large number of defect images is indispensable for the training of deep learning algorithms. The acknowledged public synthetic PCB cosmetic defect dataset was published by [89] from Peking University. However, its drawback is that one image contains only one type of defects. Further work, such as using Photoshop, can be performed to enrich this dataset so that one image contains at least two types of defects. In addition, some public datasets of solder joint defects and components defects can be downloaded. After they are modified, labels can be marked by software such as LabelMe [124] or LabelImg [125]. During the collection process, data augmentation should be executed to obtain high-volume datasets that include different kinds of PCB and PCBA defects.

B. DETECTION WITH BOTH HIGH ACCURACY AND SPEED

Previous work indicated that the YOLO series algorithms can achieve precise and real-time detection, thus making them applicable in real industrial manufacturing. The advanced YOLOv5, YOLOx [126] or YOLOv7 [127] can be used to detect several types of defects in one image, and the performance can be verified through a comparison with other state-of-the-art models. The limitations should be analysed and discussed. An improved model focusing on tiny object detection and model compression can be proposed to achieve better PCB defect detection performance.

C. TINY COMPONENT DEFECT DETECTION

As discussed previously, few deep learning models have been used to detect PCB component defects. Deep learning models can be applied to detect missing or misaligned components. Instead of having a simple context environment for cosmetic defects in PCB, component defects are usually surrounded by many chips, resistors and capacitors. Thus, the deep learning algorithm used for detecting component defects should pay more attention to the context environment around targets. Some tiny object detection models can also be introduced.

D. SOLDER JOINT DEFECT DETECTION

For solder joint defect detection, current deep learning models depend on image matching or registration and manual confirmation during the detection process. These methods can only achieve binary classification without additional operations. An end-to-end CNN model or Transformers [128] based model can be proposed to localise and classify solder defects at the same time. Compared with components, solder joint defects are smaller. Thus, the combination of global context ideas and multi-scale feature learning methods can be used to improve the performance. Moreover, because mostly tiny targets are the focus of PCB defect detection, the depth of the model backbone should contain more spatial information instead of semantic information.

E. PCB AND PCBA DEFECT DETECTION

Finally, on the basis of previous ideas that have become reality, a CNN model that is able to detect bare PCB cosmetic defects and PCBA solder joint and component defects simultaneously with acceptable performance can be proposed. This approach will be much more challenging than implementing only one of them, which means that the model must be robust to different environments and backgrounds. Such a model can greatly benefit the PCB-related manufacturing industry.

X. CONCLUSION

Quality control in the PCB manufacturing process is usually a critical problem. Various defects, including surface, component and solder joint defects, inevitably appear due to mishandling or technical faults. In this paper, different defect detection methods based on conventional image processing techniques and deep learning models are reviewed and compared. Traditional image processing-based defect detection methods achieve acceptable detection accuracy, but they are time consuming and sensitive to the environment and the inference image. With its multi-level feature extraction and automatic learning abilities, CNN-based defect detection algorithms have overcome such issues and achieved significant performances in both accuracy and speed. However, deep learning-based methods still has some limitations, such as the collection of datasets, binary classification and small range of defects. On the basis of such shortcomings, some suggestions are proposed and discussed, which are believed to achieve good results in the future.

REFERENCES

- [1] 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.
- [2] T.-M. Bajenescu, "Miniaturisation of electronic components and the problem of device overheating," *Electrotehnica, Electronica, Automatica*, vol. 69, no. 2, pp. 53–58, May 2021.
- [3] U. M. Dilberoglu, B. Gharehpapagh, U. Yaman, and M. Dolen, "The role of additive manufacturing in the era of industry 4.0," *Proc. Manuf.*, vol. 11, pp. 545–554, Jan. 2017.
- [4] N. Karnik, U. Bora, K. Bhadri, P. Kadambi, and P. Dhattrak, "A comprehensive study on current and future trends towards the characteristics and enablers of industry 4.0," *J. Ind. Inf. Integr.*, vol. 27, May 2022, Art. no. 100294.
- [5] Tempo Automation. (2021). *How Industry 4.0 Impacts Your PCB Development*. Accessed: Dec. 23, 2022. [Online]. Available: <https://www.tempoautomation.com/blog/how-industry-4-0-impacts-your-pcb-development/>
- [6] S. Spinzi. (2017). *The Evolution of Industry 4.0, Through the Eyes of the PCB Manufacturer*. Accessed: Dec. 23, 2022. [Online]. Available: <https://www.electronicdesign.com/technologies/test-measurement/article/21207506/the-evolution-of-industry-40-through-the-eyes-of-the-pcb-manufacturer>
- [7] D. B. Anitha and M. Rao, "A survey on defect detection in bare PCB and assembled PCB using image processing techniques," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2017, pp. 39–43.
- [8] 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, 1996.
- [9] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," in *Advanced Course on Artificial Intelligence*, vol. 2049. Berlin, Germany: Springer, 2001, pp. 249–257.
- [10] E. Grossi and M. Buscema, "Introduction to artificial neural networks," *Eur. J. Gastroenterol. Hepatol.*, vol. 19, pp. 1046–1054, Jan. 2008.
- [11] C. R. Reeves, "Genetic algorithms," in *Encyclopedia of Database Systems*. New York, NY, USA: Springer, 1993.
- [12] L. Rokach and O. Maimon, "Decision trees," in *The Data Mining and Knowledge Discovery Handbook*. New York, NY, USA: Springer, 2005.
- [13] M. Seul, L. O'Gorman, and M. J. Sammon, "Template matching," in *Practical Algorithms for Image Analysis: Description, Examples Code*. New York, NY, USA: Cambridge Univ. Press, 2000.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.
- [15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [16] 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.
- [17] K. P. Anoop, N. Sarath, and S. V. V. Kuma, "A review of PCB defect detection using image processing," *Int. J. Eng. Innov. Technol.*, vol. 4, pp. 188–192, May 2015.
- [18] N. Aggarwal, M. Deshwal, and P. Samant, "A survey on automatic printed circuit board defect detection techniques," in *Proc. 2nd Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, Apr. 2022, pp. 853–856.
- [19] M. F. Ahmad, N. A. M. Isa, W. H. Lim, and K. M. Ang, "Differential evolution: A recent review based on state-of-the-art works," *Alexandria Eng. J.*, vol. 61, no. 5, pp. 3831–3872, May 2022.
- [20] A. Suhasini, D. K. Sonal, B. G. Prathiksha, B. S. Meghashree, and H. D. Phaneendra, "PCB defect detection using image subtraction algorithm," *Int. J. Comput. Sci. Trends Technol.*, vol. 3, p. 8887, May 2015.
- [21] A. Mujeeb, W. Dai, M. Erdt, and A. Sourin, "One class based feature learning approach for defect detection using deep autoencoders," *Adv. Eng. Informat.*, vol. 42, Oct. 2019, Art. no. 100933.
- [22] A. Pal, S. Chauhan, and S. Bhardwaj, "Detection of bare PCB defects by image subtraction method using machine vision," *Proc. World Congr. Eng. (WCE)*, vol. 2, Jul. 2011, pp. 6–8.
- [23] K. Sundaraj, "PCB inspection for missing or misaligned components using background subtraction," *WSEAS Trans. Inf. Sci. Appl.*, vol. 6, pp. 778–787, Jan. 2009.
- [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] R. A. Melnyk and R. B. Tushnytskyy, "Detection of defects in printed circuit boards by clustering the etalon and defected samples," in *Proc. IEEE 15th Int. Conf. Adv. Trends Radioelectronics, Telecommun. Comput. Eng. (TCSET)*, Feb. 2020, pp. 961–964.
- [26] W. Zhu, H. Gu, and W. Su, "A fast PCB hole detection method based on geometric features," *Meas. Sci. Technol.*, vol. 31, no. 9, Sep. 2020, Art. no. 095402.
- [27] M. Fornaciari, A. Prati, and R. Cucchiara, "A fast and effective ellipse detector for embedded vision applications," *Pattern Recognit.*, vol. 47, no. 11, pp. 3693–3708, Nov. 2014.

- [28] T. Atherton and D. J. Kerbyson, "Size invariant circle detection," *Image Vis. Comput.*, vol. 17, no. 11, pp. 795–803, Sep. 1999.
- [29] M. A. Haq, A. K. Jilani, and P. Prabu, "Algorithmic scheme for concurrent detection and classification of printed circuit board defects," *Comput. Mater. Continua*, vol. 71, no. 1, pp. 355–367, 2022. [Online]. Available: <http://www.techscience.com/cm/v71n1/45360>
- [30] R. Akmal, N. Z. Othman, A. Bade, M. Rahim, and I. Amin, "Connected component labeling using components neighbors-scan labeling approach," *J. Comput. Sci.*, vol. 6, no. 10, pp. 1099–1107, Oct. 2010.
- [31] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *Proc. Int. Conf. Image Process.*, vol. 5, Oct. 2004, pp. 3061–3064.
- [32] D.-M. Tsai, C.-T. Lin, and J.-F. Chen, "The evaluation of normalized cross correlations for defect detection," *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2525–2535, Nov. 2003.
- [33] 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.
- [34] J.-C. Yoo, B. D. Choi, and H.-K. Choi, "1-D fast normalized cross-correlation using additions," *Digit. Signal Process.*, vol. 20, no. 5, pp. 1482–1493, Sep. 2010.
- [35] Y. Fouda and A. R. Khan, "Normalize cross correlation algorithm in pattern matching based on 1-D information vector," *Trends Appl. Sci. Res.*, vol. 10, pp. 195–206, Apr. 2015.
- [36] A.-A. Hassani, F. Abd El-Samie, and G. El Banby, "A real-time approach for automatic defect detection from PCBs based on surf features and morphological operations," *Multimedia Tools Appl.*, vol. 78, pp. 1–21, Dec. 2019.
- [37] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jan. 2008, doi: [10.1016/j.cviu.2007.09.014](https://doi.org/10.1016/j.cviu.2007.09.014).
- [38] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, p. 218, 2016.
- [39] T. S. Yun, K. J. Sim, and H. J. Kim, "Support vector machine-based inspection of solder joints using circular illumination," *Electron. Lett.*, vol. 36, no. 11, pp. 949–951, May 2000.
- [40] M. Simic. (2022). *K-Means for Classification*. Accessed: Dec. 23, 2022. [Online]. Available: <https://www.baeldung.com/cs/k-means-for-classification>
- [41] Y. Lee and R. Lippmann, "Practical characteristics of neural network and conventional pattern classifiers on artificial and speech problems," in *Proc. NIPS*, 1989, pp. 168–177.
- [42] K. W. Ko and H. S. Cho, "Solder joints inspection using a neural network and fuzzy rule-based classification method," *IEEE Trans. Electron. Packag. Manuf.*, vol. 23, no. 2, pp. 93–103, Apr. 2000.
- [43] C. Borgelt, D. Girimonte, and G. Acciani, "Learning vector quantization: Cluster size and cluster number," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2004, p. 5.
- [44] F. Demonecourt, "Introduction to fuzzy logic," Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 21, Jan. 2013.
- [45] B. Fritze, "Vector quantization with a growing and splitting elastic net," in *Proc. ICANN*, S. Gielen and B. Kappen, Eds. London, U.K.: Springer, 1993, pp. 580–585.
- [46] R. Duda, P. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, Jan. 2001.
- [47] Y. Takagi, S. Hata, S. Hibi, and W. Beutel, "Visual inspection machine for solder joints using tiered illumination," in *Proc. Other Conf.*, 1991, pp. 21–29.
- [48] J. H. Kim, H. S. Cho, and S. Kim, "Pattern classification of solder joint images using a correlation neural network," *Eng. Appl. Artif. Intell.*, vol. 9, no. 6, pp. 655–669, Dec. 1996.
- [49] A. N. Belbachir, M. Lera, A. Fanni, and A. Montisci, "An automatic optical inspection system for the diagnosis of printed circuits based on neural networks," in *Proc. Ind. Appl. Conf.*, vol. 1, Oct. 2005, pp. 680–684.
- [50] J. M. Gallegos, J. R. Villalobos, G. Carrillo, and S. D. Cabrera, "Reduced-dimension and wavelet processing of SMD images for real-time inspection," in *Proc. Southwest Symp. Image Anal. Interpretation*, 1996, pp. 30–36.
- [51] C. M. Bishop, "The multi-layer perceptron," in *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford Univ. Press, 1995.
- [52] G. Acciani, G. Brunetti, and G. Fornarelli, "Application of neural networks in optical inspection and classification of solder joints in surface Mount technology," *IEEE Trans. Ind. Informat.*, vol. 2, no. 3, pp. 200–209, Aug. 2006.
- [53] A. Rosenfeld, "The fuzzy geometry of image subsets," *Pattern Recognit. Lett.*, vol. 2, no. 5, pp. 311–317, Sep. 1984. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0167865584900187>
- [54] T. Taniguchi, D. Kacprzak, S. Yamada, and M. Iwahara, "Wavelet-based processing of ECT images for inspection of printed circuit board," *IEEE Trans. Magn.*, vol. 37, no. 4, pp. 2790–2793, Jul. 2001.
- [55] Z. Ibrahim, S. Al-Attas, O. Ono, and M. Mokji, "A noise elimination procedure for wavelet-based printed circuit board inspection system," in *Proc. 5th Asian Control Conf.*, vol. 2, Jul. 2004, pp. 875–880.
- [56] G. Acciani, G. Brunetti, and G. Fornarelli, "Automatic detection of solder joint defects on integrated circuits," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 1021–1024.
- [57] A. J. Crispin and V. Rankov, "Automated inspection of PCB components using a genetic algorithm template-matching approach," *Int. J. Adv. Manuf. Technol.*, vol. 35, nos. 3–4, pp. 293–300, Dec. 2007.
- [58] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
- [59] N. Cai, Y. Zhou, Q. Ye, G. Liu, H. Wang, and X. Chen, "IC solder joint inspection via robust principle component analysis," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 7, no. 2, pp. 300–309, Feb. 2017.
- [60] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, May 2011.
- [61] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds. Red Hook, NY, USA: Curran Associates, 2009, pp. 1–45.
- [62] N. Cai, J. Lin, Q. Ye, H. Wang, S. Weng, and B. W. K. Ling, "A new IC solder joint inspection method for an automatic optical inspection system based on an improved visual background extraction algorithm," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 6, no. 1, pp. 161–172, Jan. 2016.
- [63] X. Hongwei, Z. Xianmin, K. Yongcong, and O. Gaofei, "Solder joint inspection method for chip component using improved AdaBoost and decision tree," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 1, no. 12, pp. 2018–2027, Dec. 2011.
- [64] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: [10.1006/jcss.1997.1504](https://doi.org/10.1006/jcss.1997.1504).
- [65] H. Wu, X. Zhang, H. Xie, Y. Kuang, and G. Ouyang, "Classification of solder joint using feature selection based on Bayes and support vector machine," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 3, no. 3, pp. 516–522, Mar. 2013.
- [66] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, no. 2, pp. 131–163, Nov. 1997.
- [67] W. Hao, Z. Xianmin, K. Yongcong, O. Gaofei, and X. Hongwei, "Solder joint inspection based on neural network combined with genetic algorithm," *Optik*, vol. 124, no. 20, pp. 4110–4116, 2013.
- [68] K. Li and J.-X. Peng, "System oriented neural networks—Problem formulation, methodology and application," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 20, no. 2, pp. 143–158, Mar. 2006.
- [69] F. Wu and X. Zhang, "An inspection and classification method for chip solder joints using color grads and Boolean rules," *Robot. Comput.-Integr. Manuf.*, vol. 30, no. 5, pp. 517–526, Oct. 2014.
- [70] S.-N. Chiu and M.-H. Perng, "Reflection-area-based feature descriptor for solder joint inspection," *Mach. Vis. Appl.*, vol. 18, no. 2, pp. 95–106, Mar. 2007.
- [71] N. S. S. Mar, P. K. D. V. Yarlagadda, and C. Fookes, "Design and development of automatic visual inspection system for PCB manufacturing," *Robot. Comput.-Integr. Manuf.*, vol. 27, no. 5, pp. 949–962, Oct. 2011.
- [72] G. Acciani, G. Fornarelli, and A. Giaquinto, "A fuzzy method for global quality index evaluation of solder joints in surface Mount technology," *IEEE Trans. Ind. Informat.*, vol. 7, no. 1, pp. 115–124, Feb. 2011.
- [73] D. Li, C. Li, C. Chen, and Z. Zhao, "Semantic segmentation of a printed circuit board for component recognition based on depth images," *Sensors*, vol. 20, no. 18, p. 5318, Sep. 2020.
- [74] R. Li, Z. Liu, and J. Tan, "Exploring 3D human action recognition: From offline to online," *Sensors*, vol. 18, no. 2, p. 633, Feb. 2018.
- [75] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. CVPR*, Jun. 2011, pp. 1297–1304.

- [76] J. Shotton, "Efficient human pose estimation from single depth images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2821–2840, Dec. 2013.
- [77] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 415–422.
- [78] C. Keskin, F. Kirac, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1228–1234.
- [79] C. Keskin, F. Kirac, Y. E. Kara, and L. Akarun, "Hand pose estimation and hand shape classification using multi-layered randomized decision forests," in *Computer Vision ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Germany: Springer, 2012, pp. 852–863.
- [80] T. Sharp, C. Keskin, D. Robertson, J. Taylor, J. Shotton, D. Kim, C. Rhemann, I. Leichter, A. Vinnikov, Y. Wei, D. Freedman, P. Kohli, E. Krupka, A. Fitzgibbon, and S. Izadi, "Accurate, robust, and flexible real-time hand tracking," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 3633–3642, doi: 10.1145/2702123.2702179.
- [81] J. Tompson, M. Stein, Y. Lecun, and K. Perlin, "Real-time continuous pose recovery of human hands using convolutional networks," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 1–10, Sep. 2014, doi: 10.1145/2629500.
- [82] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [83] M. Krzywinski and N. Altman, "Classification and regression trees," *Nature Methods*, vol. 14, pp. 757–758, Aug. 2017.
- [84] A. van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, vol. 2, 2013, pp. 2643–2651.
- [85] O. Avilov, S. Rimbart, A. Popov, and L. Bougrain, "Deep learning techniques to improve intraoperative awareness detection from electroencephalographic signals," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 142–145.
- [86] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [87] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [88] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [89] R. Ding, L. Dai, G. Li, and H. Liu, "TDD-Net: A tiny defect detection network for printed circuit boards," *CAAI Trans. Intell. Technol.*, vol. 4, no. 2, pp. 110–116, Jun. 2019.
- [90] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
- [91] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
- [92] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 761–769.
- [93] 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. [Online]. Available: <https://www.mdpi.com/2079-9292/9/9/1547>
- [94] J. Redmon. *Tiny YOLOV2*. Accessed: Dec. 23, 2022. [Online]. Available: <https://pjreddie.com/darknet/yolov2/>
- [95] 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.
- [96] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, Y. Bengio and Y. LeCun, Eds., 2015, pp. 40–53.
- [97] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [98] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in *Computer Vision ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 122–138.
- [99] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
- [100] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [101] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [102] X. Liao, S. Lv, D. Li, Y. Luo, Z. Zhu, and C. Jiang, "YOLOv4-MN3 for PCB surface defect detection," *Appl. Sci.*, vol. 11, no. 24, p. 11701, Dec. 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/24/11701>
- [103] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [104] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
- [105] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 21–37.
- [106] 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, doi: 10.3934/mbe.2021223.
- [107] Ultralytics. *YOLOV5*. Accessed: Dec. 23, 2022. [Online]. Available: <https://github.com/ultralytics/yoloV5>
- [108] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1571–1580.
- [109] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.
- [110] Y. Zhang, F. Xie, L. Huang, J. Shi, J. Yang, and Z. Li, "A lightweight one-stage defect detection network for small object based on dual attention mechanism and PAFFPN," *Frontiers Phys.*, vol. 9, Oct. 2021, Art. no. 708097.
- [111] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9626–9635.
- [112] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [113] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 3–19.
- [114] H. Zhang, L. Jiang, and C. Li, "CS-ResNet: Cost-sensitive residual convolutional neural network for PCB cosmetic defect detection," *Expert Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115673. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417421010617>
- [115] W. Dai, A. Mujeeb, M. Erdt, and A. Sourin, "Soldering defect detection in automatic optical inspection," *Adv. Eng. Informat.*, vol. 43, Jan. 2020, Art. no. 101004. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1474034619305774>
- [116] Y.-T. Li, P. Kuo, and J.-I. Guo, "Automatic industry PCB board DIP process defect detection system based on deep ensemble self-adaption method," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 2, pp. 312–323, Feb. 2021.
- [117] A. Sezer and A. Altan, "Detection of solder paste defects with an optimization-based deep learning model using image processing techniques," *Soldering Surf. Mount Technol.*, vol. 33, no. 5, pp. 291–298, Aug. 2021.
- [118] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, Dec. 2014, pp. 1–15.

- [119] J. Shen, N. Liu, and H. Sun, "Defect detection of printed circuit board based on lightweight deep convolution network," *IET Image Process.*, vol. 14, no. 15, pp. 3932–3940, Dec. 2020.
- [120] R. J. Wang, X. Li, and C. X. Ling, "Pelee: A real-time object detection system on mobile devices," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2018, pp. 1967–1976.
- [121] X. Hu, "SINet: A scale-insensitive convolutional neural network for fast vehicle detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1010–1019, Mar. 2018.
- [122] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A comprehensive survey on model compression and acceleration," *Artif. Intell. Rev.*, vol. 53, no. 7, pp. 5113–5155, Oct. 2020.
- [123] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," 2017, *arXiv:1710.09282*.
- [124] *Computer Science and Artificial Intelligence Laboratory, MIT. Labelme*. Accessed: Dec. 23, 2022. [Online]. Available: <http://labelme2.csail.mit.edu/Release3.0/index.php?message=1>
- [125] *Lableimg*. Accessed: Dec. 23, 2022. [Online]. Available: <https://pypi.org/project/labelimg/>
- [126] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding Yolo series in 2021," 2021, *arXiv:2107.08430*.
- [127] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
- [128] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 213–229.



QIN LING received the B.Eng. degree in mechanical engineering and automation from Xi'an Jiaotong University, in 2013, and the master's degree in mechatronics engineering from Massey University, New Zealand, in 2022. He is currently pursuing the Ph.D. degree with the School of Electrical and Electronic Engineering, Universiti Sains Malaysia (USM). His research interests include image processing, deep learning, convolution neural networks, and artificial intelligence.



NOR ASHIDI MAT ISA received the B.Eng. degree (Hons.) in electrical and electronic engineering and the Ph.D. degree in electronic engineering (majoring in image processing and artificial neural network) from Universiti Sains Malaysia (USM), in 1999 and 2003, respectively. He is currently a Professor with the School of Electrical and Electronic Engineering, USM. He has published more than 188, 222, and 300 articles indexed in WoS-ISI (H-index of 32), SCOPUS (H-index of 38), and Google Scholar (H-index of 46), respectively. His research interests include intelligent systems, image processing, neural networks, computational intelligence, and medical image processing. Due to his outstanding achievement in research, he gained recognition, both nationally and internationally. He was recognized as the Top 2% Researcher in category—Citation Impact in Single Calendar Years 2019, 2020, and 2021, by Stanford University, Stanford, CA, USA; and the Top Research Scientist Malaysia (TRSM) by Akademi Sains Malaysia (ASM), in 2020.

• • •