7

An Overview of the Automated Optical Inspection Edge AI Inference System Solutions

Claudio Cantone[1] and Alberto Faro[2]

[1]High Technology Systems H.T.S. srl, Italy
[2]Deepsensing, DEEPS, Italy

Abstract

The aim of this chapter is to provide an overview of automated optical inspection (AOI) edge artificial intelligence (AI) inference system solutions in the digital industry by considering if, and how, they enable manufacturers to reach a satisfactory trade-off between customer needs and production costs. Numerous solutions can address customer and factory needs, from inspection machines to testing boards equipped with cameras installed near the conveyor belt. In all the considered solutions we can implement effective defect detection algorithms, such as the latest You Only Look Once (YOLO) variants based on deep learning (DL), to obtain high key performance indicators (KPIs), i.e., mean average precision, adequate process capability and high throughput yield. Parallel implementations of edge test systems allow us to further improve production yield, while repeated tests performed in sequence can allow us to approach the precision required for zero defect practice. The comparison of available solutions using KPIs, functional requirements (FRs) and non-functional requirements (NFRs) highlights that the advantage of using inspection machines is that they are equipped with user interface and data analysis which helps workers and managers to ensure high quality production process and effective order management. Their weakness is the high cost of purchase and energy consumption, whereas solutions that use computing boards for defect testing at the edge are featured by lower costs. A demonstrator to evaluate the effectiveness of edge AI solutions based on the test boards available on the market and those developed by the EdgeAI project is outlined.

Keywords: automated optical inspection, key performance indicators, functional, non-functional requirements, deep learning, PCB defect detection, edge computing, online and continual learning, process capability.

7.1 Introduction

The advent of cyber physical systems (CPS), i.e., IoT systems equipped with computational capabilities, is affecting the control systems in every industrial and service sector [28], [11]. CPSs allow computational systems to reside ever closer to the production process, reducing latency and increasing throughput yield (TPY), one of the most important KPIs in production processes.

This trend towards edge computing-based inspection systems is particularly evident in the AOI of industrial products. This is our field of research interest in the EdgeAI project.

In this context, we are faced with two different evolutions of the AOI. On the one hand traditional AOI systems that operate at the operational level are being rethought as intelligent systems to be coupled to the production line. On the other hand, CPSs equipped with camera are increasing their computational capacity to achieve effective AOI systems using local or cloud DL algorithms.

The aim of this chapter is to provide an overview of AOI edge AI inference system solutions by discussing if, and how, they allow manufacturers to reach a satisfactory balance between customer needs (mainly in terms of product quality and on-time delivery of ordered lots) and production costs. Section 2 provides the context for us to classify edge AI inference system solutions for AOI, where specific products and prototypes are highlighted to flesh out the discussion. Section 3 compares the leading AOI solutions identified in Section 2 using both KPIs and functional/non-functional requirements. Section 4 outlines the demonstrator we are setting up in the EdgeAI project to improve optical inspection in the digital industry. This will allow us to highlight how edge AI solutions can outperform or complement conventional AI-based inspection machines for AOI.

7.2 Overview of the Main Edge AI Solutions for AOI

Traditional AOI machines are tipically designed to support Surface Mount Technology (SMT) for mounting and interconnecting electronic components on printed circuit boards.

Figure 7.1a outlines the process by which an empty printed circuit board (PCB) is gradually filled with all the components to obtain a fully functional printed circuit board (PCBA). We note that in the figures of this chapter it is assumed that PCBAs are inserted into the AOI system to detect PCB or PCBA defects, i.e. defects relating respectively to the printed circuits or to the component assembly process. Also it should be noted that AOI machines are used online in two stages of the SMT process: at the exit of the Pick and Place process and after the reflow oven to detect almost any surface defect. AOI machines are primarily dedicated to discovering 2D and 3D PCBA defects and making Coordinate Mounting Measurements (CMM). There are multifunctional machines on the market that perform not only AOI and CMM but also Solder Paste Inspection (SPI) [22]. After AOI and X-ray Inspection (AXI) discover surface and internal defects respectively, an electronic test phase consisting of In Circuit Test (ICT) and Functional Test (FCT) is performed.

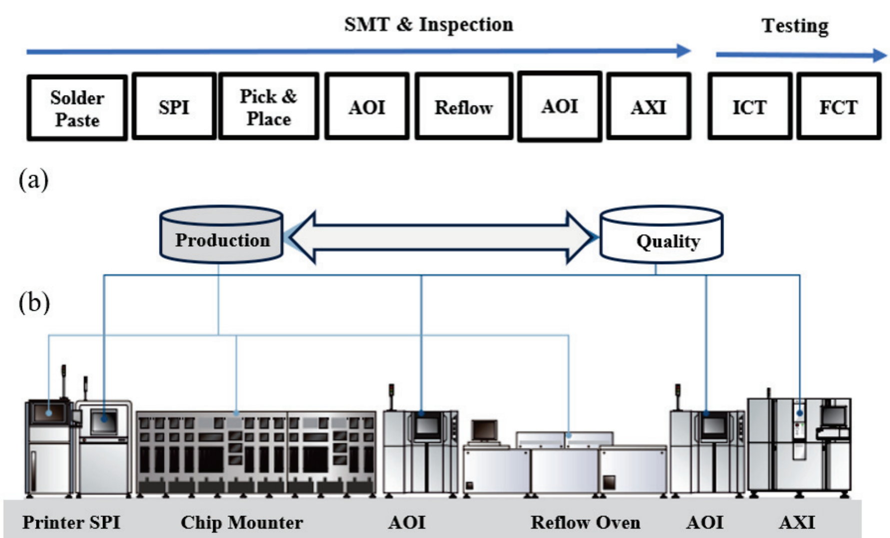

Figure 7.1 a) Printed Circuit Board (PCB) Assembly Process, and b) Typical Implementation of the SMT Production Line, where production data are taken from Printer, Chip Mounter and Reflow, whereas quality data are taken from SPI, AOI and AXI.

An example illustrating this way of using AOI is shown in Figure 7.1b which shows how OMRON proposes to use AOI to discover surface defects using 2D/3D optics and internal defects using X-ray machines [23]. This last control is increasingly widespread, as underlined in [39]. Sometimes the AOI machine is only placed after the reflow oven. In principle this solution is less expensive, although finding defects after reflow oven costs the manufacturer much more to rectify.

The above pattern is followed by mass production factories. In fact, Electronic Manufacturing Services (EMS) factories that produce small manufacturing lots featuring high technology for New Product Introduction (NPI), often adopt offline solutions to avoid changing the path of the conveyor belt in the production site. For this reason, the role of optical inspection machines can be schematized within the production line in two main ways: as control systems not necessarily close to the conveyor belt (Figure 7.2a), and as integrated control systems in the production line (Figure 7.2b) to ensure high production yield, especially in the case of mass production.

Inspection machines have recently been equipped with DL algorithms to improve the accuracy of the defect detection process, such as the Omron VT-S1080. Modern optical inspection machines can be roughly viewed as intelligent edge computing solutions for AOI, whose main problem remains the high purchase and power consumption cost and the constraints they impose on the conveyor belt layout.

A further weakness of inspection machines concerns the AI algorithms used. In fact, if they improve throughput by switching from statistical algorithms to DL inspection-based algorithms, their performance may not reach

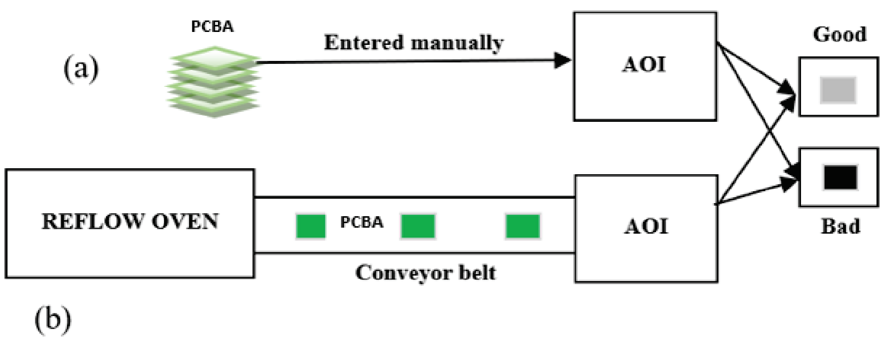

Figure 7.2 Main Inspection Machine Configurations for AOI in the Digital Industry

the high accuracy of 99.8% reported in [23] or the very low rate of defective products reported in [26], if they are not equipped with:

a) Online Learning (OL) to use experimental data to optimize the initial learning model typically obtained from data available in the literature or from similar cases, and
b) Continual learning (CL) to use experimental data to extend the learning capacity of the algorithm to discover further defects without forgetting the previous ones.

Therefore in this chapter, by OL and CL we mean a learning technique that uses experimental data to improve defect detection accuracy and to learn additional defects respectively. Regarding OL, we have to note that in the global industry, online learning should be used with caution if a test program executed on one site is expected to produce the same results as that implemented on the other sites [18]. This implies that global manufacturers should check whether OL improves defect inspection equally across their different sites. In this case, or in the case of tests on local production lines, it is useful for the learning model to be continuously updated from the images captured by the cameras to optimize the discovery of defects or to deal with different types of defects.

In principle, inspection machines can be equipped with OL and CL, but this will increase their cost as it requires the machines to be equipped with a powerful processing CPU or powered by an additional GPU due to the high computational load required by such algorithms [27].

For this reason, AOI solutions have recently appeared on the market consisting of powerful workstations, possibly equipped with GPU boards, and equipped with a high-resolution camera installed near the belt, such as those proposed in [7] using the Neousys technology (Figure 7.3a) and those proposed by Advantech [2], ADLINK [8] and AAEON [1]. Advantech and AAEON solutions are shown in Figure 7.3b and 7.3c.

The hardware architectures shown in Figure 7.3 allow us to highlight that the Neousys and AAEON solutions use a powerful workstation able of both training and testing, while Advantech uses a processing unit for testing and a GPU workstation for learning and for the classification of defects. ADLINK's solution can be achieved by replacing MIC-720 with EOS-I6000-M which is an AI vision system suitable for testing and classification, while learning takes place on cloud server.

Although these solutions support both online and continual learning and real-time verification of product defects, the problem remains of the high

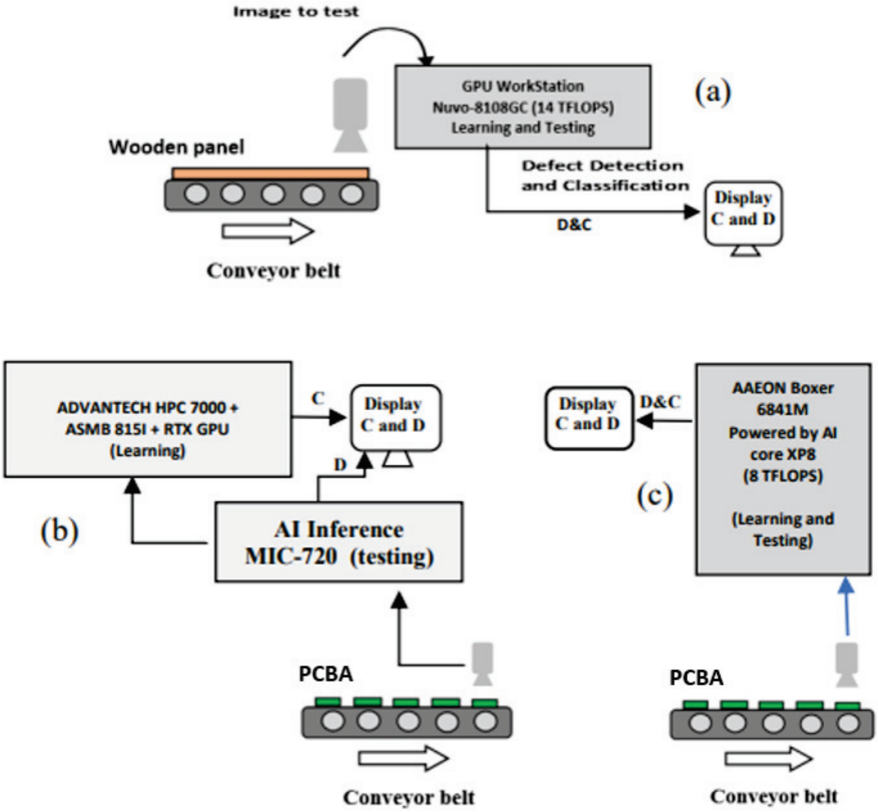

Figure 7.3 Edge AI AOI solutions from Neousys (a), Advantech (b), and AAEON (c) for Defect Detection (D) and Classification (C). The model is Pre-trained on the Workstation.

purchase and power consumption cost, as well as a certain difficulty of installing such systems near to the production line due to their size and conditioning constraints.

Alternatively, a solution where visual testing is done at the edge and learning in the cloud can reduce purchase and power consumption costs without increasing latency, as shown in Figure 7.4. This solution can be obtained by replacing the MIC-720 unit with a NVIDIA board in the Advantech proposal shown in Figure 7.3. In Figure 7.4 a Jetson board is adopted for edge tests, for example Jetson TX2 as proposed in [29]. In the latter case, learning is on the cloud but several tests suitable for highlighting groups of defects can be performed in parallel by competing boards thus decreasing latency.

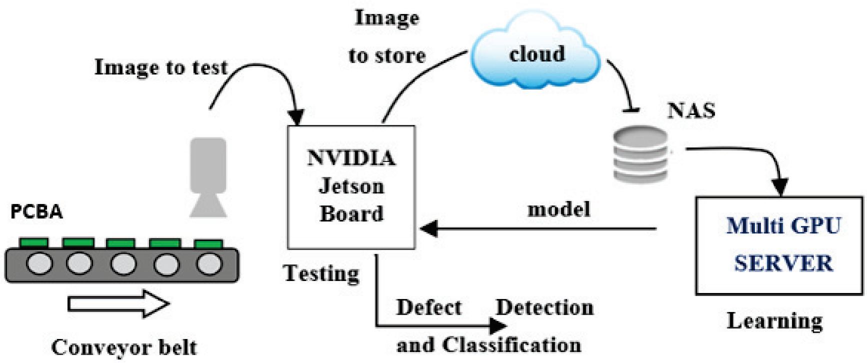

Figure 7.4 AOI Solution Consisting of an Edge Board for Testing and a GPU Server for Learning.

7.3 Comparing EdgeAI solutions for AOI using Relevant KPIs, NFRs and FRs in Digital Industry

In the previous section we outlined three main EdgeAI solutions for AOI, namely the one based on an inspection machine (see Figure 7.1), hereinafter referred to as IS, the one that makes use of a GPU workstation equipped with high-precision cameras (see Figure 7.3a and 7.3c), called GS, and the one based on a test board near the conveyor belt that sends images to the cloud server for online and continual learning (see figures 7.3b and 7.4), called ES. In the chapter we also consider a fourth solution consisting of cameras sending images to a cloud server for testing and learning, which we will call CS.

The discussion of such solutions was mainly based on cost and flexibility aspects and suggested to take into great consideration both CS and ES. In this section, we compare these solutions by considering Key Performance Indicators, Functional and Non Functional Requirements.

7.3.1 Comparison using KPIs

KPIs mainly deal with cost effectiveness, efficiency (precision) of the discrimination process and its productivity (speediness) as suggested in [13] to evaluate the performance of every digital system. Efficiency of the discovery process is usually evaluated, as in every information retrieval system, using precision and recall that may be easily obtained by the confusion matrix related to the adopted discovery algorithm [15]. The confusion matrix together with the precision and recall formulas are shown in Figure 7.5.

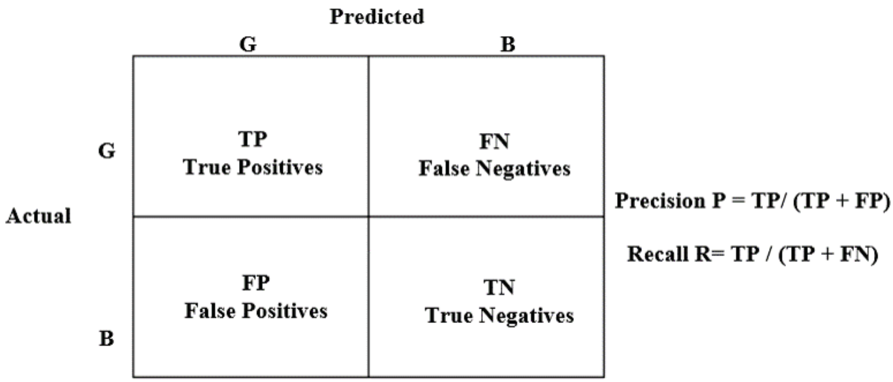

Figure 7.5 Confusion Matrix and Precision/Recall Formulas

In some cases, the most important characteristic is recall, for example, if we are interested in finding out all, or almost all, defective PCBAs, it is reasonable to increase the false positive checking effort, while in other cases it may be better to use precision, for example, if one is interesting that the discrimination process only outlines not defective PCBAs although high accuracy may increase false negatives.

At first glance, one might think that precision may be the most important feature in AOI of PCBAs, especially in cases where the requirement for near-zero defects should be adopted, for example in the aerospace sector or recently in the automotive industry. But precision alone can cause many good products to be discarded, thus increasing production costs. For this reason, efficiency indicators that combine precision and recall are used in the digital industry such as the mAP defined as the mean of the average precisions, and the F-measure defined as the weighted harmonic mean of precision and recall. The following balanced F-measure is often used, denoted as F1, which equally weighs precision and recall:

$$F1 = 2 \cdot P \cdot R / (P+R)$$

We use mAP as it is widely adopted to measure the performance of defect detection algorithms in industrial manufacturing. The meaning of mAP can be understood by introducing the notion of Intersection Of Union (IoU) [37], a measure from 0 to 1 of the similarity between the bounding box containing a possible defect and the one relating to a real one (the ground truth). According to [37] IoU is used as a threshold for whether an object having a defect-like image should enter the defective class (i.e., class consisting of defective

PCBAs) or not. Other rules can be found in the literature for whether a PCBA defect should be predicted as real, for example, in [32] a possible defect contained in a bounding box is predicted as a real defect if both IoU and another coefficient, namely the confidence coefficient, calculated by the detection algorithm are greater than 0.5.

Thus, choosing a high IoU will increase the percentage of really good items compared to those predicted good by the algorithm, but even numerous really good items may be discarded (that is, false negatives increase). Conversely, a low IoU will decrease false negatives, but defect discovery is characterized by low precision thus increasing false positives as the chosen similarity is not sufficient to discriminate good from bad elements. Consequently, to reduce the AOI alarms for possible false positives, e.g., the PCBAs featured by IoU $>$ 0,5 and whose confidence coefficient is close to 0,5, it is advisable to increase precision by adopting a most performing algorithm or to increase IoU since this generally implies an increase of the confidence too, even this is not desired since it implies an increase in false negatives.

The mAP is obtained by evaluating the average precision of the controls performed for each IoU value from 0.5 to 0.9 with a step of 0.1 and performing the mean of these averages [32]. To simplify, in the work we will use $mAP_{0.5}$, i.e. the precision of the discovery process for IoU = 0.5. Therefore $mAP_{0.5}$ = 0.99 does not mean that we will have 1% error, but that the error of the predicted good items is close to 1% with a reasonably low number of false negatives, i.e., few good products will be discarded from the ones for customers.

The above considerations justify why the efficiency of the discovery process is evaluated using mAP. We recall that the mAP depends not only on the efficiency of the discovery algorithm but also on the type of defect to be found. Typical defects to be discovered on the PCB are missing hole, mouse bite, open circuit, short circuit, spurious copper, spur. But measurements of the relevant metrological data of the PCBA are also useful, such as component height, lift, tilt, missing or incorrect component, incorrect polarity, flipped component, OCR inspection of 2D code, component offset (X / Y/rotation), fillet (e.g., end joint width, wetting angle, side joint length), exposed zone, foreign material, zone error, cable offset, cable posture, cable presence, sphere of weld, weld bridge, distance between components and component angle.

Several algorithms have been proposed in the literature to manage the problems listed above. A study highlighting different algorithms to manage either PCB or PCBA defects can be found in [12] where it is demonstrated that $mAP_{0.5}$ ranges from 95% to 98%. In this chapter we update this study considering the best performing algorithms for typical PCB defect detection. From the literature we found that these are mainly optimized versions of the DL-powered YOLO algorithm [35]. The $mAP_{0.5}$ of such algorithms increased from 95.7% proposed in 2018 [5] to higher values using the best performing DL algorithms developed from 2018 to present. For example, $mAP_{0.5}$ is 99% in the algorithm proposed in [38], 99% in [25], 98.7% in [36], 99% in [39]. Such values go beyond 99% more recently, i.e. 99.17% in [20], 99.5% in [11] and 99.71% in [40].

Although this comparison has only an indicative value since the mentioned precision values were not achieved using the same data set [33], we can reasonably assume that the solutions denoted with IS, GS and CS can be equipped with a DL algorithm whose defect discovery precision could increase from 98% to 99.7%, and that this could be further improved by online learning to 99.8%, as stated in [23].

The feasibility of implementing YOLO-based algorithms on ES has been recently shown in literature thus confirming that ES can also be equipped with such an algorithm, e.g., in [30] a YOLO implementation on NVIDIA Jetson TX2 is illustrated in characterized by satisfactory precision performance, that is, $mAP_{0.5}$ = 98%. We are currently working on solving two open problems: a) to what extent more accurate algorithms can be implemented on ES and b) how to implement such algorithms on less expensive boards (e.g., Jetson Nano and Raspberry PI4) by extending the DL algorithms proposed in [14] and [34].

However, although the theoretical accuracy of the optimized YOLO versions has reached a very high value, it may not be sufficient for the quality control of PCBAs to be used in applications where the constraint of near-zero defects is required, such as in the automotive industry [4]. In fact, 99.8% of $mAP_{0.5}$ approximately implies that the delivered defective products are about 2000 per million, whereas 1000 per million defective parts is a typical expected value in automotive products satisfying the near-zero defect constraint [24].

Note that the former failure rate, known as defects per million opportunities (DPMO) [16], measures all PCBA possible failures, i.e., defects of components or due to the assembly process. If each PCBA consists of approximately 100 components, this means that the DPMO is 2000 defective PCBAs per million if the PCB is filled with components with a failure rate of 20ppm. The DPMO in the industrial sector is used as a relevant KPI to measure the process capability, i.e., how well the process yield meets customer expectations in terms of acceptable defective products. This capability can also be expressed by a percentage (called Yield) or by a coefficient named Cpk, i.e., a statistical coefficient between 0 and 2 where Cpk = 2 means that there are no defective PCBAs leaving the production process, while Cpk = 0 means that the quality process does not detect any fault, so all the faulty boards are still in the leaving products. A conversion table is available in the literature to pass from DPMO to Yield or to Cpk and vice versa, e.g., in [31].

In the semiconductor industry, DPMO = 6000 is an acceptable value if the near-zero defect constraint is not required. Using the conversion table, we can find that this corresponds to Cpk = 1.33 and Yield = 99.40%. Consequently, if we aim to have DPMO = 6000 for both defective components and surface defects, using the conversion table we obtain that we must use 99.55% of non-defective components plus an instrument, such as AOI for example, obtaining 99. 80% accuracy to discriminate between good and bad products coming out of the SMT process. The latter accuracy in detecting surface defects can be achieved by recent versions of the AI-YOLO algorithm, but applications characterized by the zero-defect constraint require a Yield of 99.98% which can be achieved using an AOI of 99.95% of precision.

Therefore, while waiting for more performing algorithms, it is reasonable to carry out two or three repetitions of the AOI checks of the products classified as good to improve the accuracy as proposed in [9]. Indeed, this is a reasonable procedure only if the AOI checks are statistically independent. as claimed in [6] due to the noise superimposed on the images when they are taken by the cameras, for example due to faded colours or weaknesses in the lighting system. Consequently, reproducing the control using the same AOI can eliminate the uncertainty due to noise issues thus allowing the AOI to approach its maximum theoretical accuracy calculated using literature data.

Also, as pointed out earlier, one could reduce the volume of bad products delivered as good (i.e., to reduce false positives) by increasing the IoU, but this usually also increases false negatives. In fact, this can lead to consider good products the ones that are close to the boundary between the "good" and "bad" classes and which are affected by the maximum uncertainty of classification.

Therefore, checking the PCB using another AOI system, i.e., not reproducing the measurement but replicating it using a different AOI, can be useful to improve accuracy without changing the recall. This repetition could add some defective elements to the "bad class" as suggested in [21] which, hopefully, could coincide with the few defective products that were not detected by the first test. This is also stated in [9] where it is underlined

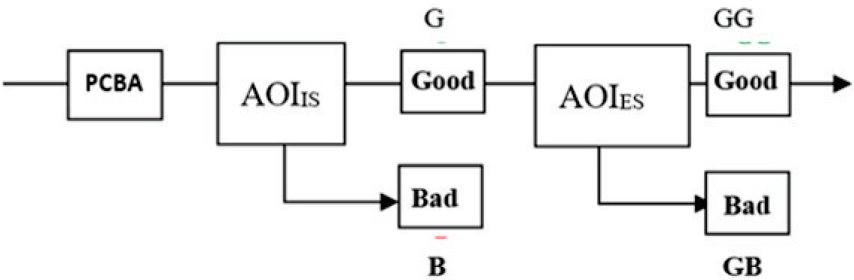

Figure 7.6 Repeating the AOI Check.

that repetition improves accuracy in the electronics industry even if beyond a certain threshold repetition is not cost effective due to the increasing cost of adding a further check.

This consideration suggests evaluating in our project the possibility of adding a check after the inspection machine using an ES check to try to satisfy the constraint of near-zero defects. In fact, the hypothesis of adding a further control to the one currently carried out without modifying the layout is an opportunity given the low cost and the high flexibility of the ES. Figure 7.6 shows how the repetition scheme proposed in [9] can be reworked to improve the mAP of AOI. Test repetition to avoid bad products reaching customers could be carried out, even manually, only for testing the few PCBAs that passed the first test but were classified close to the border between good and bad clusters.

In addition to mAP and process capability to evaluate process efficiency, another important KPI is the productivity of an AOI system, also known as throughput yield (TPY), to measure good PCBAs at optical control output in the unit of time. Since in all considered IS, GS and CS the test phase is performed on GPU machines, the comparison can be made considering the latency due to the algorithm and the camera system used to acquire the images of the PCBAs on the belt. Latency mainly depends on the implementation of the algorithm and is often not indicated in the literature. It can be measured indirectly by the speed, in frames per second (FPS), at which the proposed algorithm is able to process the images.

A general comparison of the FPS achievable in the available methods for defect discovery can be found in [30] where the authors pointed out that their version of the YOLO algorithm is able to reach about 90 FPS. This value is also confirmed in other studies, for example we found that the FPS goes from 33 FPS in [40] to 90 FPS in [20]. Lower but satisfactory FPS characterize

3D defect detection, for example 19 FPS in [Du, 2023]. Regarding ES, we found from the literature that even in ES the implementation of YOLO algorithms can achieve high throughput, for example, in [30] it is proved that the DL-based YOLO algorithm implemented on Jetson TX2 can process 22 FPS [30].

Therefore, using a TX2 board, it can be expected that a 25 cm^2 PCB can be inspected in about 90 msec, if each image taken by the camera is about 5 cm^2. This means that the AOI production per hour obtainable using ES could be around 3250 boards per hour (bph) which is a value comparable with the value of 4189 bph obtained using IS reported in [26]. Let us note that such values refer to the number of PCBA exiting from the optical inspection phase (see fig.1.1a) . Indeed other electrical checks may decrease such thoughput, e.g., the ones dealing with the determination of the safe operating area of PCBs to be used in power applications. In [26] the authors proposed other relevant KPIs beyond hourly production, i.e., precision of detected defects, working time and delivery times from order to shipment.

The accuracy of defect discovery can be calculated using mAP as shown above, while the last two proposed KPIs depend on the organization of work. Therefore, they can only be analysed by knowing the factory organization structure, order volume and rate. It is out of the scope of the chapter. However, they suggest us that mAP and FPS alone are not sufficient to measure the impact of AOI on the SMT process. In fact, cycle time and takt time should also be included in the KPIs at least to verify that the AOI production system can meet the time constraints due to customer orders. A general discussion may be found in [17]. For the paper, it is sufficient to include the following parameters in the KPI list:

- Cycle time (CT), i.e., the time required to produce a lot of PCBs requested by the customer divided by the number of PCBs.
- Takt time (TT), i.e., the time interval during which the production line is available in the time interval required by the customer to produce the PCB lot divided by the number of PCBs to be delivered to the customer. In other words, it is the maximum time interval for producing one PCBA to meet the customer time constraint considering the availability of the production resources and the number of PCBAs of the lot.

Knowing CT and TT we can verify the condition necessary to satisfy the customer's demand, i.e., CT $<$ TT. This means that CT (the inverse of throughput yield) is a very important KPI that should be appropriately scaled down to meet overall customer demand in due time. This can be achieved: i) by increasing the FPS of the AOI unit, ii) by using more than one AOI unit in parallel, or iii) by implementing more than one production line. The first two conditions can be obtained more conveniently by ES than by IS since its low cost allows adopting many cameras to work in parallel. In fact, the CT of a production line can be improved by passing from a solution in which a camera sends images to a testing board as illustrated in Figure 7.7a. to the one proposed in [3] made up of several cameras possibly equipped with a testing board (Figure 7.7b). In both cases, the images are sent to a server to update the pre-trained model.

To get an idea of the cost savings using ES in both cases illustrated in Figure 7.7 let us consider the market cost of CS, ES, and IS. Assuming one CS as a unit cost, from the market cost we found that this cost becomes 2 for ES, 6 for a WS provided with GPUs and from 25 to 50 for IS depending on if the IS is a low-cost machine or a professional one. Therefore, the purchase costs are as follows: n+6 for CS, 2n+6 for ES and 25 or 50 for IS where n is the number of cameras and related testing boards.

Using these values, Figure 7.8a compares the costs of ES and CS with the cost of a low cost IS proposed by Saki in [29]. This comparison is feasible since they have the same configuration, i.e., they are all equipped with a camera which, thanks to a telecentric lens system (Figure 7.7a), takes pictures of PCB slices of about 5 x 25 cm while it is placed on the conveyor belt.

Instead, to evaluate the cost savings by using multiple cameras and boards, we compare the CS and ES with the OMRON professional solution, i.e., VT-S1080, assuming that the CS (ES) is equipped with 5 camera positions (5 camera positions plus 5 testing boards) as in Figure 7.7b so that the whole PCB can be inspected as it is transported on the conveyor belt and at the same time the OMRON AOI captures all images of the PCB using a robotic system that moves the camera over the PCB inside the machine. The comparison is shown in Figure 7.8b.

Figure 7.8 clearly shows that in both cases CS and ES are less expensive than IS. The cycle time using CS and ES in case 3.3a. is greater than that of IS, then CS and ES are suggested only if a relatively high CT is acceptable. If a lower CT is required, the parallel implementation is recommended. Furthermore, we should mention that both mAP and FPS could be further improved in ES by using more performant testing boards like the ones proposed in [19] where it is stated that object classification can be performed at hundreds of FPS. This is for further study.

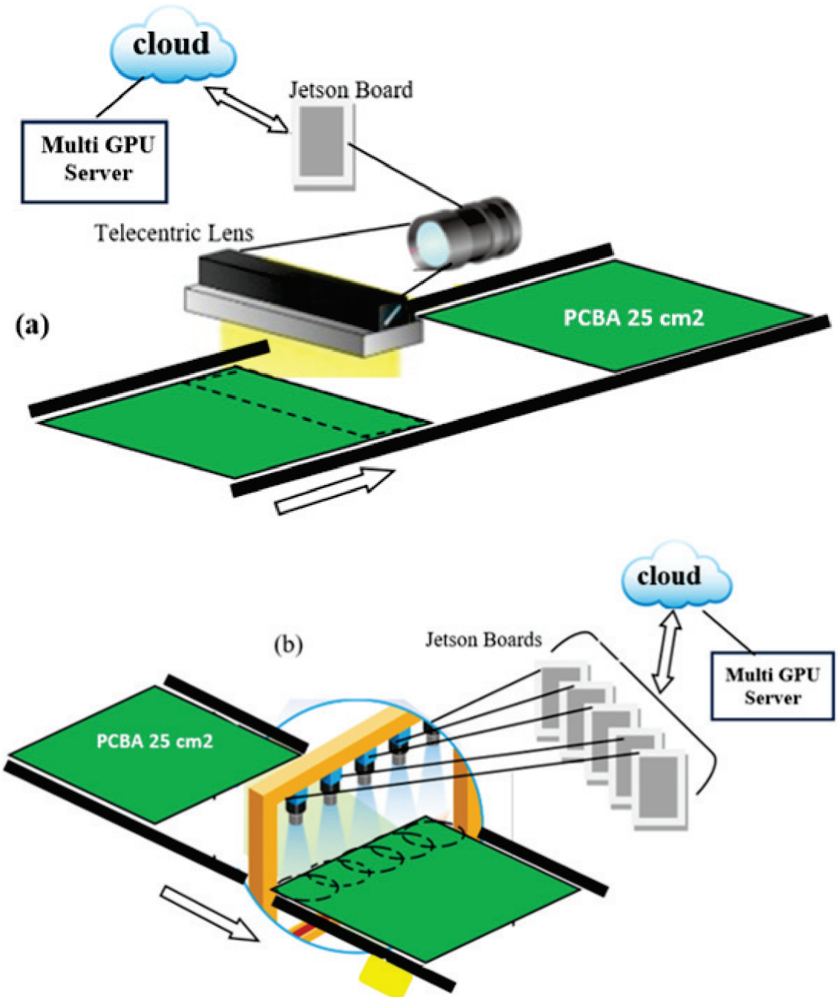

Figure 7.7 a) A Camera Equipped with a Testing Board Which Sends the Image of a PCB Slice of about 5 X 25 cm Using a Telecentric Lens to a Testing Board, b) A Set of Five Cameras Equipped with Testing Boards. Images Are Sent to a Server to Update the Pre-Trained Model. The Server Periodically Sends the Updated Model to the Edge Testing Boards.

7.3.2 Comparison using NFRs

In addition to the mentioned KPIs, further quality requirements, so-called non-functional requirements, should be considered to compare the different solutions. Below we indicate some NFRs that consider those

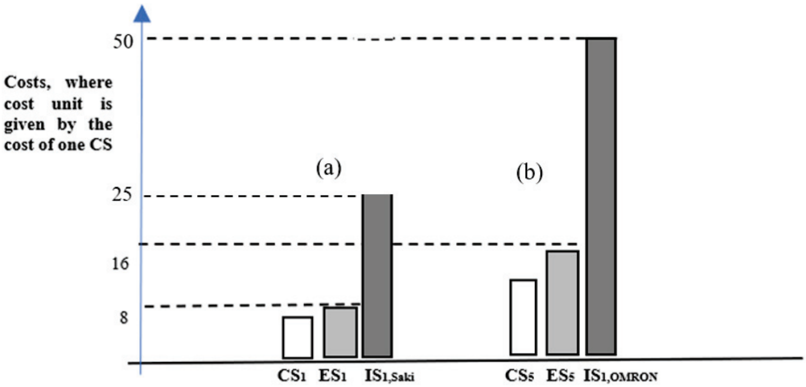

Figure 7.8 An Approximate Comparison of the Purchase Costs of CSs and ESs Equipped with one Camera Versus the Low Cost 2D Saki AOI (a), and the Purchase Costs of CSs and ES Equipped with Five Cameras Seats Versus the Professional 2D/3D OMRON AOI (b).

proposed in [26] for the support of workers, i.e., the adopted solution should:

- Enable efficient use of workers' time through automation.
- Improve control capability through real-time data feedback.
- Explain the defects at least by locating the defects found on the PCB or PCBA, which allows workers to improve the production process.

All the above NFRs can be satisfied by CS and ES provided that appropriate user interfaces are developed that help workers interpret and manage data from optical inspection.

NFRs are proposed in [26] dealing with planning tools, i.e., managers should be supported by suitable planning tools, based on data from AOI and other IoT monitoring systems, to meet takt time and to verify more generally that the overall time including the purchase of the raw material and the delivery of the products to customers (i.e., the lead time) is compatible with the customer's demand. This implies that lead time should also be included in the KPI list above.

Since the current IS and GS provide effective operator interfaces and planning tools for managers, these solutions, despite the high cost, can maintain some advantage over ES until ES is equipped with the mentioned worker interface and management tools.

This can be facilitated by the fact that ESs can take full advantage of parallel technology and cloud computing. In fact, the worker interface, usually installed near the production line, could be implemented by adding another board to the edge testing system, while data analysis tools could be implemented on the network server available to plant managers.

This issue should be addressed carefully in ES and is a challenge for any project aiming to use AOI testing boards at the edge.

7.3.3 Comparison using functional requirements

To complete the comparison, the main FRs must also be considered. The following list consists of five FRs, of which the first two are mandatory while the last three are highly recommended. Such FRs require that any AI solution for AOI:

a) it should have a high mAP suitable to support the process capability required by the industrial sector of interest of the producers, for example CPk = 1.33 for the semiconductor industry. Based on the discussion in this section, all solutions could meet this requirement using OL-based DL defect detection algorithms.
b) it should be able to detect PCB defects in real time as the PCBs are transported on a conveyor belt. Based on the discussion made in this section, all solutions can meet this requirement due to their relatively high FPS value.
c) it must have an adequate feedback loop with the machine controls. This requirement also belongs to the NFR list mentioned, but here it is understood as the requirement that the solution has a minimum set of functions to help workers and managers optimize the PCB production. IS and GS usually satisfy such FR, while it is acceptable for CS and ES to provide at least some defect location functionality to explain the causes of the defect.
d) it should learn to discover defects by exploiting the data available in the literature. In principle, this requirement is satisfied by GS, CS and ES as they are usually open systems, while ISs are usually designed as closed solutions which are not provided with network attached storage system on the cloud to include data from the literature to improve the accuracy of learning or to handle new defects.

7.3.4 Advantages of ES with respect to the other approaches

The comparison of the AI based AOI solutions using the main KPIs, NFR and FR has highlighted that ES is a promising technology provided it is equipped with adequate operational and management interfaces.

Some points that encourage the effort to equip ESs with such interfaces are not only their low cost and parallel processing that allow them to achieve better KPIs for the detection of multiple defects simultaneously, but also the possibility for ESs to take full advantage of the cloud technology not only to use the cloud to better build the mentioned user interfaces, but also to enable small companies to use AOI-based control remotely.

7.4 Edge AI Solutions Demonstrator

Given the different solutions available for optical defect detection, a demonstrator can be useful to evaluate if and how an AOI solution can help in practice manufacturers to reach a satisfactory compromise between the quality required by customers (in terms of acceptable number of defective items and takt time) and factory costs.

Currently we are activating such demonstrator equipped with the following technologies:

- An IS machine, supplied by HTS, i.e., OMRON AI-AOI VT-S1080, to measure mAP and FPS achievable during the PCBA test and to verify that it is able of achieving using Deep Learning the high accuracy required by the industrial sector of interest, i.e., semiconductor or automotive sectors.
- A workstation, provided by DEEPS, equipped with a 7 TB storage system, an AMD Ryzen^{TM} Threadripper 3970 CPU and two NVIDIA RTX 6000 GPUs. This workstation currently acts as a server on the local network so it will allow us to simulate CS ed ES but could be connected via a fast channel to cameras to simulate GS as well.
- Several NVIDIA boards, namely Jetson Nano, Jetson TX2, Xavier and ORIN, to host the algorithm trained on the GPU server at the edge and a NAS (Network Attached Storage) system to store the images taken by the cameras to allow the server to online update the pre-trained model.
- High-resolution Basler cameras to take images of PCBs as they are being transported on a conveyor belt. These images will be sent to the server's NAS or workstation near the belt or passed through a fast channel to the NVIDIA cards.

The components from b) to d) will allow us to activate the demonstrator using the same platform illustrated in Figure 7.4, to measure the most significant KPIs and to evaluate the NFR for defect detection. In this way, commercially available ES and new EdgeAI AOI solutions, such as the one based on the low-power board to be developed by the EdgeAI project, could be compared with GS, CS and IS.

We note that the main purpose of the demonstrator is not to support designers in developing DL defect detection algorithms that outperform the current ones, even if this test can also be performed using the platform, but to demonstrate that: a) the DL-based defect models obtained from the pre-training phase on the server can be implemented on the edge boards to obtain test performance comparable to that of IS and GS but at a lower cost as required mainly by mass production companies, and b) the AI Edge solution can be equipped with extremely precise defect discovery and defect explainability algorithms to support the improvement of the production process and in the identification of possible critical components as required mainly by NPI.

Furthermore, the conditions suggesting the combination of different solutions can be studied. For example, if a low throughput yield is acceptable, this may justify CS over the others. In addition, an ES-based remote solution will be tested to support small companies in implementing a simple and cost-effective solution where the testing board is installed close to the conveyor belt and the learning powered by OL and CL is done by a cloud server.

7.5 Conclusion

An overview of the available solutions for AI-based optical defect inspection of PCBAs has been made from an engineering point of view, i.e., emphasizing whether and how they can support a satisfactory trade-off between product quality and production costs.

From the overview it emerged that it is possible to adopt different solutions to meet the needs of the factory and customers, from inspection machines to testing boards equipped with cameras installed near the conveyor belt. Generally, in all the considered solutions it is possible to implement effective defect detection algorithms, such as the latest DL-based YOLO versions, to obtain the suitable mean precision, i.e., mAP, to support the required process capability.

The main advantage of using inspection machines is that they have data analysis tools that support managers to ensure high quality and effective management planning. Their weakness, i.e., the high cost of purchase and energy consumption, is the strength of solutions that use processing boards for defect testing at the edge.

Parallel implementations of edge solutions, using suitable optical systems, improve latency and the number of PCBAs that may be classified as good or bad products per time unit, while repeated tests carried out by a test board installed after the inspection machine, allow us to approach the process capability required in industry sectors characterized by the near zero defects constrain. This can be achieved without decreasing recall, thus avoiding an increase in false negatives.

It was discussed how a solution can achieve a low cycle time that can meet takt time and lead time to satisfy customer demand, emphasizing that using ES this can be achieved by increasing the FPS of the AOI and activating, if necessary, parallel AOI units in the production line.

A suitable platform was also presented to evaluate the most suitable solutions using experimental data. This will help us demonstrate the efficiency, productivity, and cost-effectiveness of a solution in practice and test whether coprocessing units, such as the recent neuromorphic boards, can improve discovery algorithms. It will allow us also to demonstrate how small companies can use the platform to perform defect detection using local testing boards supervised by a remote server.

Acknowledgements

This research was conducted as part of the EdgeAI "Edge AI Technologies for Optimised Performance Embedded Processing" project, which has received funding from KDT JU under grant agreement No 101097300. The KDT JU receives support from the European Union's Horizon Europe research and innovation program and Austria, Belgium, France, Greece, Italy, Latvia, Luxembourg, Netherlands, and Norway.

References

[1] AAEON, "AI@Edge: AI Vision in Automated Optical Inspection", 2023, Available at:https://newdata.aaeon.com.tw/DOWNLOAD/application/BOXER-6841M%20AOI%20Story.pdf

[2] Advantech, "AI AOI Selected Success Story", 2023, Available at:https://advcloudfiles.advantech.com/membership/upload/ee3b76aa/AI-AOI Success-Story-CollectionFinal.pdf

[3] Advantech, "Building AOI Technology to Accelerate Yield in precision manufacturing", March 2023, Available at:https://www.advantech.com/en-eu/resources/case-study/building-aoi-technology-to-accelerate-yield-in-precision-manufacturing

[4] AEC-Q004: Zero Defects Guideline. Automotive Electronics Council, 2020, Available at:http://www.aecouncil.com/Documents/AEC_Q004_Rev-.pdf

[5] L, Chang, "A CNN Based Reference Comparison Method for Classifying Bare PCB Defects", *The Journal of Engineering*, 2018 (16)

[6] S. Chan, "How Repeatable Is Your AOI? An Easy and Quick Way To Find Out", 2017, Available at:http://www.linkedin.com/pulse/how-repeatable-your-aoi-easy-quick-way-find-out-steven-chan

[7] L.C. Chen, M.S. Pardeshi, WT Lo, et al., "Edge-glued wooden panel defect detection using deep learning". *Wood Science Technology* **56**, 477–507 (2022).

[8] Y. Chia-Wei, "How Edge AI Can Improve the Visual Inspection Process", *Quality Magazine*, September 30, 2020

[9] Y. Chung-Huang, C. Jwu-E, "Application of Three-Repetition Tests Scheme to Improve Integrated Circuits Test Quality to Near-Zero Defect", *Sensors* 22 (4158), May 2022

[10] A. Costanzo, A. Faro, D. Giordano and C. Spampinato, "An ontological ubiquitous city information platform provided with Cyber-Physical-Social-Systems", *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA, 2016, pp. 137-144. Available at:https://doi.org/10.1109/CCNC.2016.7444746

[11] Y. Du, et al., "An automated optical inspection (AOI) platform for three-dimensional (3D) defects detection on glass micro-optical components (GMOC)", *Optics Communications*, to appear on Volume 545, 15 October 2023

[12] A.A.R.M.A. Ebayyeh and A. Mousavi, "A Review and Analysis of Automatic Optical Inspection and Quality Monitoring Methods in Electronics Industry", *IEEE Access*, vol. 8, pp. 183192-183271, 2020

[13] Eurocontrol, KPI Drafting Group, "Cost Effectiveness and Productivity KPIs", October 2001, Available at:http://www.eurocontrol.int/sites/default/files/2019-05/cost-effectiveness-and-productivity-kpis-2001.pdf

[14] R. Faro, "Object Detection and Semantic Segmentation models for Defect Detection in Wood Production", Undergraduate thesis, Department of Electrical, Electronics and Computer Engineering University of Catania, 2023

[15] T, Fawcett, "An Introduction to ROC Analysis", *Pattern Recognition Letters*, 27 (8): 861–874. 2006

[16] K. Feldeman, “Driving Quality Improvement with DPMO: A Roadmap to Process Excellence”, July 17, 2023 Available at:https://www.sixsigma.com/dictionary/defects-per-million-opportunities-dpmo/
[17] E. Fogg, “Takt Time vs. Cycle Time vs. Lead Time: definitions and calculations”, *Machine Metrics, Manufacturing Analytics*, September 24, 2020
[18] D, Haigh, “Repeatability Is AOI Watchword for Volume SMT Production”, 1999, Available at:http://www.edn.com/repeatability-is-aoi-watchword-for-volume-smt-production/
[19] Hailo, “Delivering Unprecedented Performance to a Diverse Range of Edge AI Applications”, 3 August 2023, Available at:https://www.edge-ai-vision.com/companies/hailo/
[20] J.Y. Lim, J.Y.1 Lim, V.M. Baskaran, X. Wang, “A deep context learning based PCB defect detection model with anomalous trend alarming system”, *Results in Engineering*, Volume 17, March 2023
[21] R.J Mackenzie, “Repeatability vs. Reproducibility”, *Technology Networks*, March 25, 2019
[22] Nordson, SQ3000^{TM}+ multi-Function for 3D AOI, SPI & CMM, online 2021
[23] OMRON, “VT-S1080, Next generation 3D AOI”, 2023, Available at:https://inspection.omron.eu/en/products/vt-s1080#features
[24] R. Oshiro (2018), Fundamentals of AEC-Q100: “What Automotive Qualified Really Means”, Available at:https://media.monolithicpower.com/mps_cms_document/w/e/Webinar_Fundamentals_-_of_AEC-Q100-6Nov2018.pdf
[25] J.H. Park, Y.S. Kim, H. Seo H, Y.J. Cho. “Analysis of Training Deep Learning Models for PCB Defect Detection”, *Sensors* (Basel). 2023 Mar 2;23(5):2766
[26] M. Park, J. Jeong, “Design and Implementation of Machine Vision-Based Quality Inspection System in Mask Manufacturing Process”, *Sustainability* 2022, 14, 6009. Available at:https://doi.org/10.3390/su14106009
[27] A. Prabhu, et al, “Computationally Budgeted Continual Learning: What Does Matter?”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2023, pp. 3698-3707
[28] M. Ryalat, H. El Moaqet, M. Al Faouri, “Design of a Smart Factory Based on Cyber-Physical Systems and Internet of Things towards Industry 4.0”. *Appl. Sci.* 2023, *13*, 2156. Available at:https://doi.org/10.3390/app13042156

[29] Saki, "2D AOI technology", Available at:https://www.sakicorp.com/en/company/technology/2daoi_tec/

[30] S. Shaojun, J. Junfeng, H. Yanqing, S. Mingyang, EfficientDet for fabric defect detection based on edge computing, *Journal of Engineered Fibers and Fabrics* Volume 16: 1–13, 2021.

[31] Six Sigma Material, "Conversion Tables", Available at:https://www.six-sigma-material.com/Tables.html

[32] J. Solawetz, "What is Mean Average Precision (mAP) in Object Detection?" Roboflow, May 6, 2020, Available at:https://blog.roboflow.com/mean-average-precision/

[33] G, Spadaro, et al., "Towards One-Shot PCB Defect Detection with YOLO", Ital-IA 2023: *3rd National Conf. on Artificial Intelligence*, organized 000by CINI, Pisa, IT May 29–31, 2023

[34] A. Strano, "Chip Surface Defect Classification: A Benchmark Analysis of Deep Learning Architectures", Undergraduate thesis, Department of Electrical, Electronics and Computer Engineering, University of Catania, 2023

[35] J. Terven, D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond", August, 2023, Under review in ACM Computing Survey, Available at:https://doi.org/10.48550/arXiv.2304.00501

[36] Y. Yang, K. Haiyan, "An Enhanced Detection Method of PCB Defect Based on Improved YOLOv7", *Electronics* 12, no. 9: 2023 Available at:https://doi.org/10.3390/electronics12092120

[37] S. Yohanandan, "What is Mean Average Precision (MAP) and how does it work", Xailent, June 5, 2020, Available at:https://xailient.com/blog/what-is-mean-average-precision-and-how-does-it-work/

[38] C. Zhang, W. Shi, X. Li, H. Zhang, H. Liu, "Improved bare PCB defect detection approach based on deep feature learning", *The 2nd 2018 Asian Conference on Artificial Intelligence Technology*, (ACAIT 2018), *J. Eng*., 2018, Vol. 2018 Is. 16, pp. 1415-1420, 2018.

[39] Q. Zhang, et al., "Deep learning-based solder joint defect detection on industrial printed circuit board X-ray images", *Complex & Intelligent Systems* 8:1525–1537, 2022

[40] H. Zhu, L. Xing, H. Fan, T. Wu, "New PCB Defect Identification and Classification Method Combining Mobile Net Algorithm and Improved YOLOv4 Model", Research Square, 2022