MDPI

Article

Advancements in PCB Components Recognition Using WaferCaps: A Data Fusion and Deep Learning Approach

Dmitrii Starodubov [1], Sebelan Danishvar [1,*], Abd Al Rahman M. Abu Ebayyeh [2] and Alireza Mousavi [1]

[1] Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, UK; dmitry_starodubov@hotmail.com (D.S.); alireza.mousavi@brunel.ac.uk (A.M.)
[2] Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK; a.abu-ebayyeh@imperial.ac.uk
* Correspondence: sebelan.danishvar@brunel.ac.uk

Abstract: Microelectronics and electronic products are integral to our increasingly connected world, facing constant challenges in terms of quality, security, and provenance. As technology advances and becomes more complex, the demand for automated solutions to verify the quality and origin of components assembled on printed circuit boards (PCBs) is skyrocketing. This paper proposes an innovative approach to detecting and classifying microelectronic components with impressive accuracy and reliability, paving the way for a more efficient and safer electronics industry. Our approach introduces significant advancements by integrating optical and X-ray imaging, overcoming the limitations of traditional methods that rely on a single imaging modality. This method uses a novel data fusion technique that enhances feature visibility and detectability across various component types, crucial for densely packed PCBs. By leveraging the WaferCaps capsule network, our system improves spatial hierarchy and dynamic routing capabilities, leading to robust and accurate classifications. We employ decision-level fusion across multiple classifiers trained on different representations—optical, X-ray, and fused images—enhancing accuracy by synergistically combining their predictive strengths. This comprehensive method directly addresses challenges surrounding concurrency, reliability, availability, and resolution in component identification. Through extensive experiments, we demonstrate that our approach not only significantly improves classification metrics but also enhances the learning and identification processes of PCB components, achieving a remarkable total accuracy of 95.2%. Our findings offer a substantial contribution to the ongoing development of reliable and accurate automatic inspection solutions in the electronics manufacturing sector.

Keywords: PCB components recognition; capsule networks; convolutional neural networks; data fusion; WaferCaps; image classification

Citation: Starodubov, D. ; Danishvar, S.; Abu Ebayyeh, A.M.; Mousavi, A. Advancements in PCB Components Recognition Using WaferCaps: A Data Fusion and Deep Learning Approach. *Electronics* **2024**, *13*, 1863. https://doi.org/10.3390/electronics13101863

Academic Editor: George A. Tsihrintzis

Received: 8 April 2024
Revised: 1 May 2024
Accepted: 6 May 2024
Published: 10 May 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Recent years have seen a growing demand for placing electronic components and integrated circuits (ICs) on printed circuit boards (PCBs) with greater density and precision. This trend has been driven by rapid advancements in electronic products such as digital cameras, televisions, and other cutting-edge devices [1,2]. As a result, electronic assembly lines play a critical role in manufacturing these products with the utmost accuracy and efficiency. It is of paramount importance that electronic components and ICs are assembled in strict accordance with the designated class and location on the PCBs. Consequently, the positioning and integration of these components must not only be precise but also optimized to meet the escalating demands and ensure seamless operation across production lines [3]. A bill of materials (BoM) refers to a detailed roster of raw materials, components, and instructions utilized for the manufacturing, repairing, or construction of a product. BoMs are typically organized in a hierarchical structure, with the finished product listed at the top and the individual materials and components listed below. Within the BoM, various

components such as resistors, capacitors, ICs, ports, connectors, and more are included. BoMs serve multiple purposes, including electronic board inspection, quality assurance, security, and technical auditing [4,5]. Utilizing computer vision in automatic optical inspection (AOI) can result in cost reduction and improved accuracy when inspecting electronic components and performing ICs positioning tasks. Additionally, it aids in the prevention of various flaws, including missing interposition, component switching, incorrect polarity, and lead lift [6]. AOI can be performed using conventional image sensor devices (e.g., charged coupled device (CCD)) and constant illumination settings (e.g., non-environmental light) for inspecting surface features or X-rays for internal features that are not visible. Both techniques have advantages and drawbacks that are summarized in Table 1 [1].

Table 1. Summary of quality monitoring approaches used in industrial inspection.

Inspection Method	Description	Advantages	Limitations
X-ray	A standard X-ray inspection system is composed of three main elements: an X-ray tube as the source, an X-ray detector, and a fixture that holds and manages the position of the component being inspected. The X-ray source emits X-rays that pass through the object, while the receiver captures the transmitted energy. By analyzing the transmitted energy, the internal characteristics of the inspected component can be identified.	• Very efficient for detecting inner features. • The X-ray inspection results are not significantly influenced by the shape and surface conditions [7].	• Traditional X-ray methods have the potential to cause damage. • Micro-level inspections typically exhibit a relatively low resolution. • The processing time is lengthy, typically on the order of hours.
Image sensor (e.g., CCD)	Detects surface features according to human visual perception.	• Among all quality monitoring approaches, it is considered the least expensive, relatively simple, and most frequently used. • Regarded as a non-contact and non-destructive method. • Capable of identifying surface features and defects. • AOI helps to reduce time and improve the accuracy of detection.	• Not efficient for inner features. • Need for constant light and illumination settings.

As can be seen from Table 1, each technique can acquire different types of information. Therefore, for maintaining the advantages of both, image fusion techniques can be used. Image fusion involves merging the information from multiple images to create a single image. The resulting image contains more comprehensive information compared to any of the individual images being combined [8]. In this study, we propose combining optical and X-ray images for better identification of PCB components and ICs. The combined images are then sent to the inspection algorithm for identifying which component is which. Standard image-processing techniques such as template matching and rule-based classifiers are still used as detection algorithms for PCB component identification. However, with the rapid development of computational powers and GPU, the door is now wide open for using many powerful methods such as deep learning (DL). DL techniques can meet the requirement of high demand on production by performing detection and inspection on a real-time basis [6]. Furthermore, it can save a lot of time by performing the feature extraction within their hidden layers, instead of performing hand-crafted feature extraction as in the case of conventional image processing and machine learning tools. In this study, we propose a data fusion approach followed by capsule neural network WaferCaps [9] for

identifying, recognizing, and classifying the five different groups of components of PCBs: chips, connectors, ports, two-solders, and other components.

2. Related Work

A variety of studies have utilized X-ray and AOI techniques for PCB component detection, leveraging advanced algorithms to aid the recognition and detection process. In this section, we will review some of these articles and list the main contributions of our research. Wu et al. [10] used Normalized Cross-Correlation (NCC) analysis to detect component misplacement in PCBs. They also proposed an accelerated particle swarm optimization (AS-PSO) to conduct the search process and speed up the detection process. This study achieved a recognition rate of 100%, although only resistors were tested and no other components were included. Li et al. [11] proposed an improved algorithm based on YOLO-v3 for PCB electronic component detection. In order to improve the recognition of components, they combined a real PCB picture with a virtual PCB picture accompanied by synthesized data. Experimental results indicate that the improved YOLO-v3 algorithm can achieve a mean average precision (mAP) of 93.07%. However, the improved YOLO-v3 was not efficient in identifying some components such as resistors and capacitors. Schmidt et al. [12] proposed an X-ray imaging system for solder joint inspection in surface mount technology (SMT) electronics production using a convolutional neural network (CNN). They used 2D grayscale images to feed the CNN classifier. The results indicated a significant reduction in the false call rate. However, there is still a dependency on the selection of the appropriate focus level of the image files made by the machine, which requires further development of the proposed solution. Wenwei Zhao et al. [13] underscored the value of combining human-interpretable computer vision algorithms—such as those analyzing color, shape, and texture features—with machine learning approaches to reduce the dependency on large datasets and improve the explainability of the models. The results of the study indicated that color features are promising for the detection of PCB components. In parallel to advancements in PCB component recognition, Huanjie Tao et al. [14] developed novel data augmentation techniques to address the challenges of data scarcity in surface defect inspection. A notable approach involves an erasing-inpainting-based data augmentation method using a denoising diffusion probabilistic model (DDPM), which generates diverse training images from a limited number of samples. This technique mirrors our efforts in PCB assurance, where acquiring extensive defect-specific data can be challenging due to the rapid evolution of PCB technologies. By adapting such innovative data augmentation methods, the PCB component detection models can benefit from enhanced training efficiency and improved generalization capabilities, similar to their successful application in generalized surface defect inspection. Huanjie Tao et al. [15] also introduced the attention multi-hierarchical feature fusion network (AMHNet) designed to improve defect recognition on steel surfaces by effectively fusing multi-level features and enhancing relevant details through a novel gating mechanism. Extensive testing demonstrated that attention multi-hierarchical feature fusion networks set new benchmarks in accuracy and AUC, achieving state-of-the-art results in defect recognition.

3. Methodology

The methodology of this study is crafted to address the intricate challenges of recognizing and classifying components on densely packed printed circuit boards (PCBs), where traditional single-modality imaging techniques fall short. By innovatively integrating optical and X-ray imaging, our approach not only enhances the visibility of both surface and hidden component features but also overcomes the typical limitations—such as the optical method's restricted view of internal component structures and the X-ray's lower resolution for surface details. This dual-modality fusion, facilitated by our advanced data fusion technique, allows for a more detailed and comprehensive dataset, essential for high-accuracy classification. Furthermore, the deployment of the WaferCaps capsule network introduces a significant advancement over traditional convolutional neural networks (CNNs) by lever-

aging its superior capabilities in maintaining spatial hierarchies and utilizing dynamic routing, which enhances the accuracy and reliability of the classification outcomes.

3.1. Image Fusion

The physical limitations of imaging sensors can make it difficult to obtain uniformly good images from a scene. The fusion of images represents one possible solution to this problem. Through this technique, a perfect image of a scene can be created by combining multiple samples that each provide complementary information about the scene. In this paper, images of the same electronic PCB are combined using X-rays and optical images. Through this process, salient and relevant information from hidden parts of images acquired by an X-ray machine (for instance, inside a chip) is fused with details from optical images [16].

The fusion of images takes place on three levels: at the pixel level, at the feature level, and at the decision level. In pixel-level image fusion techniques, input images can be processed directly for further computer processing. The feature level techniques for image fusion involve the extraction of relevant features, such as pixels, textures, or edges, and blending them together to generate supplementary merged features. In decision level fusion, multiple classifiers combine their decisions into a single one that describes the activity that occurred [17]. A variety of fusion methods can be categorized into two groups: traditional algorithms (spatial and frequency domain techniques) and deep learning-based methods [18]. In spite of the high performance of traditional fusion methods, they have some disadvantages. A major problem is that fusion performance heavily depends on the extraction and selection of features, and there is no universal method for obtaining features. To address these drawbacks, deep learning-based fusion methods have been developed. In these fusion methods, deep learning is employed to extract deep representations of the information provided by the source images. Various fusion strategies have been proposed in order to reconstruct the fused image. The fusion strategy can also be designed with deep learning. In this paper, we employ the innovative fusion method proposed by Jingwen Zhou et al. [16] that is presented for infrared and visible image fusion based on the VGG-19 model. In this method, unlike Li et al.'s proposed approach in [19], the source image does not need to be split into basic and detailed parts. This decomposition makes the fusion process too complex and leads to the incomplete extraction of details and salient targets. The Jingwen fusion method uses grayscale images as inputs. Therefore, because the optical images are in color in this research, the IHS (intensity, hue, and saturation) transform is applied for transforming the optical images from RGB (red, green, and blue) to the IHS color space [20]. In the IHS space, intensity indicates the spectral brightness, hue represents the wavelength, and saturation reflects the spectrum's purity. Through the Jingwen fusion method, the intensity component of the optical image is combined with the X-ray image and then the updated intensity, along with the hue and saturation, is converted back into RGB color space. The result of this process is a fused color image. Figure 1 displays the architecture of the whole fusion process. Instead of decomposing images into high and low frequencies, optical images and X-ray images feed into VGG-19 for layer-by-layer feature extraction. In X-ray images, hidden targets are usually visible because of the imaging characteristics of the technique. Optical cameras usually capture more surface details, but the targets are often covered by other objects and there is also no way to see inside the components. With its superior classification and location capability, VGG-19 is well suited to the fusion task because it can extract detailed features and salient targets.

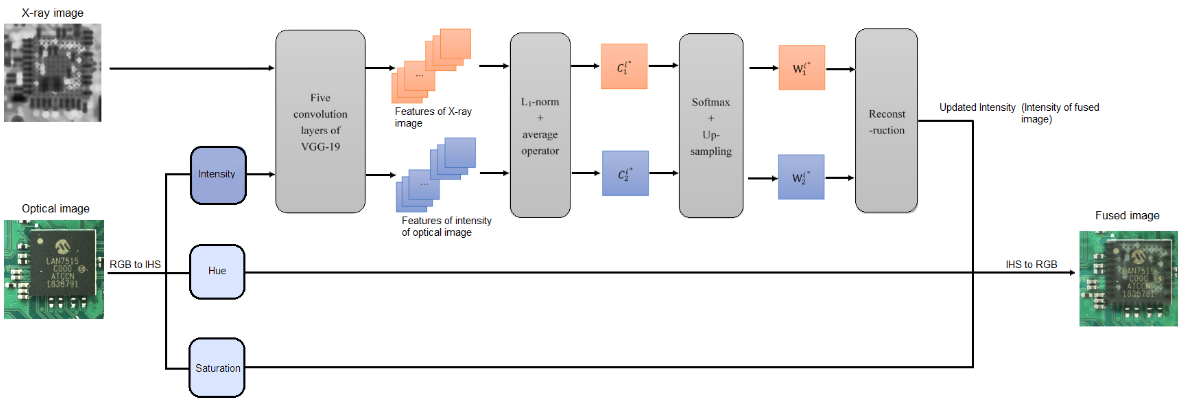

Figure 1. Image fusion approach.

3.2. VGG-19 Network: Features Extraction, Processing, and Reconstruction

VGG-19 is a CNN (convolutional neural network) for which over a million images from the ImageNet database have been used to train. This network is composed of 19 layers and has the capacity to classify images into one thousand object categories. Therefore, the network has learned affluent representations of features for a variety of images. The proposed fusion method by Jingwen entails five layers that are used to extract detailed features and salient targets.

The first two selected convolutional layers are conv1_1 and conv1_2 of VGG-19, which are mainly responsible for extracting details and edges and then must be retained. The third selected layer is conv2_1, which mainly extracts the edges in the picture. Conv3_1, the fourth selected layer, is used to extract the image's prominent targets. The last retained layer is conv4_1, which mainly extracts the salient targets. Using L1-norm and average operators, activity level maps ($C_a^{i^*}$) are derived from the extracted features and targets. Weight maps ($W_a^{i^*}$) are generated using the SoftMax function and upsampling operator. The X-ray and intensity components of the optical images are then convolved with five different weight maps to produce candidates for fusion. In this section, the final fused image is formed by using the maximum strategy to combine the five candidate fused images. Figure 1 demonstrates the process of feature processing. As a result of the L1-norm, intuitive images are translated into objective data distributions, and the average operator ensures robustness to misregistration of the images.

The average operator of the final activity level map ($C_a^{i^*}$) is shown in Equation (1):

$$C_a^{i^*}(x,y) = \frac{\sum_{v=-r}^{r}\sum_{w=-r}^{r}||x_a^{i,1:N}(x+v,y+w)||_1}{(2r+1)^2} \tag{1}$$

where $x_a^{i,1:N}(x,y)$ is an N-dimensional vector of the feature maps of the input image a (since we fuse two input images then $a \in 1,2$), which derived from the i-th convolutional layer. In the i-th layer, N indicates the maximum number of channels. The parameter r represents the size of the average operator and according to [16], it is set to 1. Based on the final activity level map ($C_a^{i^*}$), an initial weight map is calculated by utilizing the SoftMax function. Equation (2) indicates that all weight map values fall within the range [0, 1]:

$$W_a^{i^*}(x,y) = \frac{C_a^{i^*}(x,y)}{\sum_{n=1}^{K} C_a^{i^*}(x,y)} \tag{2}$$

where i denotes the convolutional layer number ($i \in 1,2,3,4,5$) and K shows the number of activity level maps, whose value is 2 since source images contain X-rays and optical images. In VGG-19, the pooling operator reduces the size of the feature maps gradually via subsampling with a stride of two. Due to this, the size of the feature maps in different convolutional layer groups is $1/(2^{a-1})$ of the original image.

As soon as the initial weight maps are obtained ($W_a^{i^*}$), they are up-sampled to ensure that they are in proportion to the source image. In Equation (3), the final weight map has the same dimensions as the source image:

$$W_a^{i^*}(x+p, y+q) = W_a^{i^*}(x, y) \;\; p, q \in \{0, 1, \ldots, (2^{a-1} - 1)\} \tag{3}$$

Eventually, based on Equation (4), each pixel of all five competitor fusion images is determined to be the maximum value as the final result in the fusion image [16]:

$$F(x, y) = \max\left[\sum_{n=1}^{K} W_n^{i^*}(x, y) \times I_n(x, y) | i \in \{1, 2, 3, 4, 5\}\right] \tag{4}$$

where $I_n(x, y)$ are the source images and K is the number of source images. Figure 2 depicts an example of a chip component optical and X-ray images and the resulting fused image.

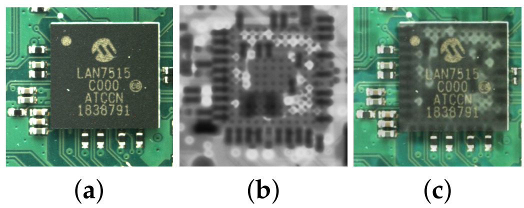

Figure 2. An example of a chip component: (**a**) optical image, (**b**) X-ray image, (**c**) fused image.

3.3. WaferCaps

Many computer vision tasks have extensively utilized convolutional neural networks (CNNs) [21]. Although CNNs have shown remarkable performance in many classification tasks, they still have some drawbacks. One of these drawbacks is the use of pooling layers in CNNs. Pooling layers have an advantage in decreasing the computation requirements by shrinking the sizes of the feature maps during the feed-forward process; however, this comes with the cost of losing many vital features that could be important in the learning process. Additionally, CNNs have limitations in accurately identifying the spatial location of the inspected feature within an image [22].

The capsule network (CapsNet) is a newly proposed neural network that is employed in classification tasks and can overcome the previous drawbacks of CNNs. It was developed in 2017 by Sabour et al. [23] and it was implemented primitively to classify MNIST handwritten digits dataset. CapsNet stands apart from conventional CNNs due to two primary factors: dynamic routing and layer-based squashing [24]. Feature detectors with scalar output are replaced with capsules with vector output in CapsNets. Furthermore, CapsNets use the routing-by-agreement concept instead of pooling layers. In CapsNet, each capsule consists of multiple neurons, where each neuron represents specific features in different regions of an image. This approach enables the recognition of the entire image by considering its individual parts [25].

The initial layer of CapsNet employs a convolutional layer similar to CNNs, but subsequent layers differ in structure. In the second layer, known as PrimaryCaps, each of the 32 capsules possesses an activity vector $\mathbf{u}_i$ to encode spatial information through instantiation parameters. The output of $\mathbf{u}_i$ is then transmitted to the subsequent layer, DigitCaps, where each of the 16 capsules per digit class receives $\mathbf{u}i$ and performs matrix multiplication with the weight matrix $\mathbf{W}_{ij}$. This computation yields the prediction vector $\hat{\mathbf{u}}_{j|i}$, which signifies the contribution of capsule i in PrimaryCaps to capsule j in DigitCaps as indicated by Equation (5):

$$\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij}\mathbf{u}_i \tag{5}$$

Subsequently, the predictions undergo multiplication with a coefficient known as the coupling coefficient c, which signifies the level of agreement between capsules as depicted

in the equation. Consequently, the value of the coefficient c is updated iteratively through an iterative process, giving rise to what is commonly referred to as "Dynamic Routing". This process is determined by utilizing a routing Softmax function, where the initial logits b_{ij} represent the logarithmic prior probabilities of coupling capsule i in PrimaryCaps with capsule j in DigitCaps. These operations can be exemplified through Equations (6)–(9):

$$a_{ij} = \mathbf{s}_j \cdot \hat{\mathbf{u}}_{j|i} \tag{6}$$

$$b_{ij} = b_{ij} + a_{ij} \tag{7}$$

$$c_{ij} = \frac{\exp\left(b_{ij}\right)}{\sum_k \exp\left(b_{ij}\right)} \tag{8}$$

$$\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i} \tag{9}$$

In this context, $\mathbf{s}_j$ represents the weighted sum computed to derive the candidates for the squashing function v_j. The role of the squashing operation is to generate a normalized vector from the collection of neurons present within the capsule. The activation function employed for this purpose can be described by Equation (10):

$$v_j = \frac{\|\mathbf{s}_j\|^2}{1 + \|\mathbf{s}_j\|^2} \cdot \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|} \tag{10}$$

To facilitate the classification process, a margin loss function is established. This function assesses the loss term derived from the output vector of DigitCaps, aiding in determining the correspondence between the chosen digit capsule and the actual target value of class k. The mathematical representation of the margin loss function can be observed in Equation (11):

$$\begin{aligned} L_k = {} & T_k \max\left(0, m^+ - \|v_k\|\right)^2 + \\ & \lambda(1 - T_k) \max\left(0, \|v_k\| - m^-\right)^2 \end{aligned} \tag{11}$$

Here, the label T_k is used to indicate the presence ("1") or absence ("0") of class k. The hyper-parameters of the model, denoted as m^+, m^-, and λ, hold specific values: m^+ is set to 0.9, m^- is set to 0.1, and λ is set to 0.5.

The original CapsNet performed well in classifying the MNIST dataset with 99%; however, it did not provide high accuracy with more complex images such as the CIFAR-10 dataset. Therefore, in this study, we propose a modified version of CapsNet known as WaferCaps that was originally proposed in [9] to classify semiconductor wafer defects and was also used in [26] to classify optoelectronic wafer defects. The structure of WaferCaps is shown in Figure 3 and Table 2. WaferCaps, in comparison to CapsNet, incorporates an additional two convolutional layers with larger kernel sizes, enabling a more effective feature extraction procedure. Additionally, dropout layers are introduced after each convolutional layer to mitigate overfitting.

Table 2. Layers of the proposed WaferCaps.

Layer	Type	Input Size	Kernel Size/Stride	Activation	Dropout	Output Size
1	conv1	[128,128,3]	15/1	ReLU	Yes	
2	conv2	[50,50,256]	15/1	ReLU	Yes	[36,36,512]
3	conv3	[36,36,512]	15/1	ReLU	Yes	[22,22,1024]
4	PrimCaps	[22,22,1024]	9/2	ReLU	No	[4,4,8,128]

Table 2. *Cont.*

Layer	Type	Input Size	Kernel Size/Stride	Activation	Dropout	Output Size
5	WBMCaps	[4,4,8,128]	-	Squash	No	[16,5]
6	FC	[16,5]	-	Softmax	No	[5]

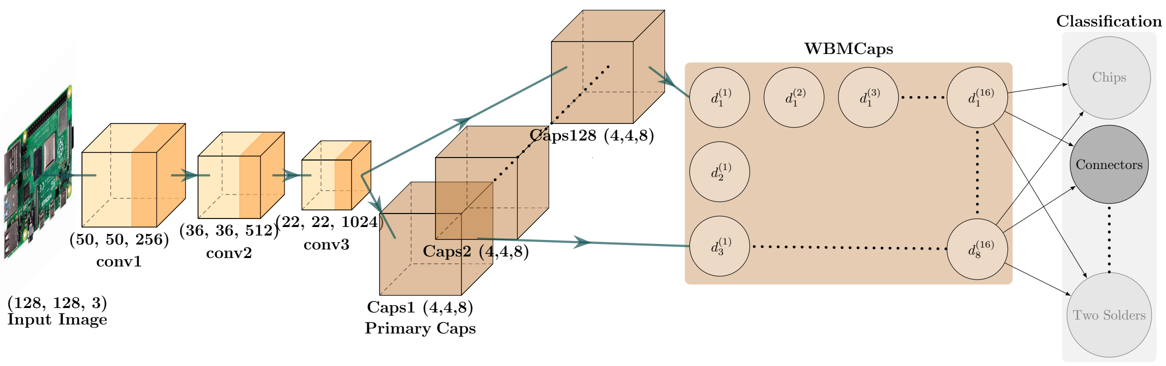

Figure 3. WaferCaps architecture.

3.4. Decision Fusion

The concept of decision fusion refers to a method of combining data from different classifiers into a single decision about the activity that made up the dataset that was merged. It has been shown in many studies that a decision fusion approach has a significant effect on classification accuracy.

A given classification problem may be classified differently by different classification techniques. Using decision fusion, multiple classifiers are integrated into a common explanation of an event and a variety of rules can be integrated in a fully flexible manner. The accuracy of classification can be improved through decision fusion.

Decision fusion techniques can be divided into several types based on their architecture: serial decision, parallel decision and hybrid decision fusion. In serial decision fusion, classifiers are arranged one after another, and their outputs are fed into the next classifier. In parallel decision fusion, several classifiers perform classification simultaneously in parallel, then combine their results. Hybrid decision fusion is a hierarchy-based classification process [27].

In this study, three WaferCaps-based networks are combined to accomplish a parallel decision fusion process. According to Figure 4, these parallel processes use optical, X-ray, and fused images to provide a final decision. Algorithm 1 shows a general view of this process. As will be seen from the results, some networks are able to accurately predict the class of components with a higher probability than others. Therefore, integrating these three networks improves the accuracy of detecting the class of components. Accordingly, the decision fusion approach will lead to an increase in the accuracy of the final classification of all classes. The combined and fused classifier is composed of a selection rule and three individual classifiers.

Three WaferCaps-based classifiers based on three different training datasets are laid out in the first layer. The three training datasets are composed of optical, X-ray, and fused images. In the second layer, selection methods are applied to the outputs of the individual classifiers in order to produce a final classification result. Every classifier produces outputs that represent the probability of each component. In percentage terms, the decimal numbers between 0 and 1 represent confidence levels.

In the first layers of classifiers, O_P, X_P, and F_P represent the probability of the predicted class of components, and O_C, X_C, and F_C represent predicted classes by networks that are trained based on X-ray, optical, and fused images, respectively. In the second layer, Algorithm 1 describes the selection rules. The application of these rules to the output of the three classifiers results in a high level of accuracy due to the blending of the advantages of all three classifiers. Trial and error are used to determine the thresholds.

Algorithm 2 succinctly encapsulates the entirety of the methodology in a structured pseudocode format. This algorithm presents all steps of our approach in a unified sequence, ensuring that the logic and operations are clearly delineated and easily interpretable.

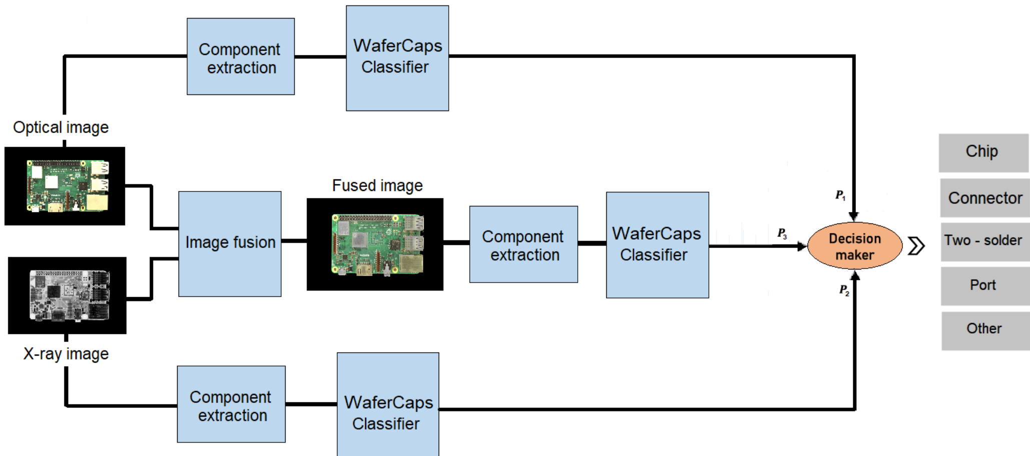

Figure 4. Decision fusion-based classification method.

Algorithm 1 Selection rules.

1: **procedure** SELECTION($O_C, X_C, F_C, O_P, X_P, F_P$)
▷ Oc, Xc, and Fc represent predicted classes by the networks trained on optical, X-ray, and fused images, respectively. Op, Xp, Fp represent the corresponding probabilities of the predicted classes.
2: **if** $(O_P > 0.80)$ **AND** $(X_P > 0.80)$ **AND** $(F_P > 0.80)$ **then**
3: predictedClassLabel $\leftarrow$ $classProbabilities[max(O_P, X_P, F_P)]$
▷ Assume that 'classProbabilities' is a dictionary linking probabilities to class labels
4: **else if** $(X_C \neq O_C)$ **AND** $(F_C \neq O_C)$ **AND** $(O_P < 0.90)$ **AND** $(X_P < 0.90)$ **then**
5: predictedClassLabel $\leftarrow F_C$
6: **else if** $(X_C \neq O_C)$ **AND** $(F_C \neq O_C)$ **AND** $(X_P > 0.85)$ **then**
7: predictedClassLabel $\leftarrow X_C$
8: **else**
9: predictedClassLabel $\leftarrow O_C$
10: **end if**
11: **return** predictedClassLabel
12: **end procedure**

Algorithm 2 PCB component classification.

```
1:  procedure PCB_COMPONENT_CLASSIFICATION
2:      opticalImage ← CAPTUREOPTICALIMAGE
3:      xrayImage ← CAPTUREXRAYIMAGE
4:      opticalIHS ← CONVERTTOIHS(opticalImage)
5:      IHSfusedImage ← FUSEIMAGES(opticalIHS, xrayImage, 'VGG − 19')
6:      fusedImage ← CONVERTTORGB(IHSfusedImage)
7:      opticalComponents ← COMPONENTEXTRACTION(opticalImage)
8:      xrayComponents ← COMPONENTEXTRACTION(xrayImage)
9:      fusedComponents ← COMPONENTEXTRACTION(fusedImage)
10:     opticalClass ← CLASSIFY(opticalComponents, 'WaferCaps')
11:     xrayClass ← CLASSIFY(xrayComponents, 'WaferCaps')
12:     fusedClass ← CLASSIFY(fusedComponents, 'WaferCaps')
13:     finalDecision ← DECISIONFUSION(opticalClass, xrayClass, fusedClass)
14:     return finalDecision
15: end procedure
16: function CAPTUREOPTICALIMAGE
17:     // Capture optical image using a camera setup
18: end function
19: function CAPTUREXRAYIMAGE
20:     // Capture X-ray image using X-ray equipment
21: end function
22: function CONVERTTOIHS(image)
23:     // Convert RGB image to IHS color space
24: end function
25: function FUSEIMAGES(opticalIHS, xrayImage, method)
26:     // Apply image fusion algorithm on intensity component of IHS optical image and X-ray
    image
27: end function
28: function CONVERTTORGB(image)
29:     // Convert IHS image back to RGB color space after fusion
30: end function
31: function COMPONENTEXTRACTION(image)
32:     // Extract components from the image (Single-component images have been extracted from
    the PCB images and labelled already)
33: end function
34: function CLASSIFY(image, method)
35:     // Classification process using the specified method (e.g., WaferCaps)
36: end function
37: function DECISIONFUSION(opticalClass, xrayClass, fusedClass)
38:     // Final decision is made based on Algorithm 1.
39: end function
```

4. Experimental Results

4.1. Experiments Environment and Devices

In this research, we build an image acquisition system that takes pictures from devices using X-ray imaging and an optical camera. For component detection and recognition that require high resolution and quality, the acquired images are suitable. In order to provide constant illumination, the optical system is enclosed in a tent that controls outside light. As a background, various colored papers are employed to help devices stand out from the background. For instance, red colored papers are used when the PCB base is black. In this research, we have 100 various circuit boards that range in size from 70 mm × 50 mm to 350 mm × 250 mm. The acquisition system includes a professional camera 'Allied Vision ALVIUM 1800U-2050C-CH-C' with two high-resolution lenses with two focal lengths of 35 mm and also 12 mm (KOWA LM35SC and KOWA LM12SC), which are mounted above distances of 42.6 cm and 62 cm for lenses 35 mm and 34 cm and 45 cm for lenses 12 mm.

These four postures of lenses and camera positions guarantee that small and large PCBs (such as mainboards) are fully visible in a single image with high spatial resolution. All optical images have a resolution of 5496 $(H) \times 3672$ (V) pixels. During optical image acquisition, the PCBs are sampled in two different exposures, one high and one low, which are controlled internally by the camera software. With these two different exposures, we can both increase our dataset size and have a more accurate interpretation of different components with different materials. An X-ray machine is one of the best ways to inspect the quality of electronic boards. The SCANNA mailroom X-ray machine (SCANMAX 225) with an additive high-resolution panel (80 Micron) offers a spacious scan area of 56 cm $\times$ 42 cm, which is large enough to accommodate most PCBs and includes its own control and analysis software. The X-ray images are made up of 3840 $(H) \times 3072$ (V) pixels. We assume in this paper that the images to be fused have been registered correctly and that sufficient single-component images have been extracted from the PCB images. We process data in RGB format according to five classes: chips, connectors, two solders, ports, and others. Training, validation, and test data are separated into three categories. There are 4000 components images per class in total. The training dataset has 15,000 samples, whereas the validation and test datasets each contain 2500 samples.

4.2. Evaluation Criteria

This section evaluates how well our proposed fused image classification performs in comparison to using non-fused images. Several metrics were used in our evaluation, including confusion matrices, accuracy, precision, F1-score, and recall. Examining the confusion matrix involves studying the outcome predictions for a classification issue. As a result, it illustrates the ways in which a classification model makes mistakes when making predictions. The accuracy determines the number of pictures (components) that are correctly classified. Precision refers to the calculation of the proportion of correctly identified positive samples out of the total number of positive samples. Sensitivity measures the proportion of correctly classified positive samples to all samples in a given class. In the F1-score, the harmonic mean of a classifier's precision and sensitivity is integrated into one metric:

$$Accuracy = \frac{TP + TN}{TN + FN + TP + FP} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Sensitivity = \frac{TP}{FN + TP} \tag{14}$$

$$F1 - score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{15}$$

where TP is the value of the true positive, FP is the false positive, TN is the true negative, and FN is the false negative.

4.3. Ablation Study on Input Image Scenarios

To substantiate the efficacy of our innovative approach, we conducted ablation studies focusing on three distinct models corresponding to different imaging inputs: optical only, X-ray only, and fused images. This investigation aimed to delineate the contribution of each imaging modality to the overall performance of the component recognition system.

(A) Optical images only:

Initially, we trained and evaluated a model using only optical images. This scenario focuses on assessing the capability of the WaferCaps model to classify components based on visible light imaging, which typically captures the external surface features of the components.

(B) X-ray images only:

Subsequently, a separate model was developed using only X-ray images. X-ray imaging provides internal views of the components, which are crucial for identifying embedded or obscured features not visible in optical imaging.

(C) Fused Images:

Finally, we utilized a model that combines both optical and X-ray images through our novel fusion technique. This model aims to leverage the complementary information available from both imaging types, potentially enhancing the model's ability to accurately classify a wider range of component types under various conditions.

For each scenario, we trained the models using the same network architecture and hyperparameters to ensure a fair comparison. The models were then evaluated based on their classification accuracy, precision, recall, and F1-score, with particular attention paid to the improvements observed in the fused image scenario. The results in Table 3 demonstrate that while the models trained on optical-only and X-ray-only images performed commendably, the fused image model exhibited superior performance, confirming our hypothesis that the integration of multiple imaging modalities enhances the system's overall effectiveness. Specifically, the fused image model showed significant improvements in detecting components with complex internal structures or those partially obscured in optical images.

Figure 5 presents the confusion matrices for the three input image scenarios utilized in our study. Each matrix provides a visual representation of the classification accuracy and misclassifications made by the WaferCaps model under different imaging conditions. This ablation study not only underscores the benefits of our image and decision fusion approach but also highlights the robustness of the WaferCaps architecture in handling diverse imaging inputs. These findings are critical, as they validate the design choices made in our methodology and provide a clear justification for the adoption of fused images in PCB component recognition tasks.

Table 3. Comparative performance of WaferCaps model classification across different imaging scenarios.

Dataset	Metric	Chips	Connectors	Others	Ports	Two Solders
Optical images	Recall	0.938	0.942	0.868	0.926	0.962
	Precision	0.956	0.935	0.863	0.965	0.921
	F1-score	0.947	0.938	0.865	0.945	0.941
X-ray images	Recall	0.932	0.944	0.7	0.952	0.938
	Precision	0.949	0.99	0.803	0.969	0.775
	F1-score	0.94	0.966	0.748	0.961	0.849
Fused images	Recall	0.938	0.926	0.876	0.922	0.958
	Precision	0.971	0.957	0.82	0.968	0.916
	F1-score	0.954	0.941	0.847	0.945	0.936

4.4. Comparative Analysis of Deep Learning Models Using Fused Images and Decision Fusion

In order to verify the effectiveness of the proposed approach, we first train and evaluate three different models with three input image groups: optical, X-ray, and fused images. The presented approach is compared with significant popular algorithms such as ResNet-50, Inception-V3, and MLP that are trained from scratch. The results of the classification of WaferCaps and three other different classifiers on optical images only can be identified in the confusion matrix of Figure 6 and metrics in Table 4. It can be observed that recall metrics for the chips and ports components are notable in both Inception-v1 and MLP; however, WaferCaps achieved higher overall accuracy with 92.7% than the other classifiers. In the second place, Inception-v3 is seen with 87.4% overall accuracy. Figure 7 and the metrics in Table 5 show the performance of classifiers on X-ray images only. Comparing

the four networks, WaferCaps achieved the highest level of accuracy with 89.3%, followed by ResNet-50 with 88.4%, then Inception-v3 with 84.9%, and finally MLP with 80.4%.

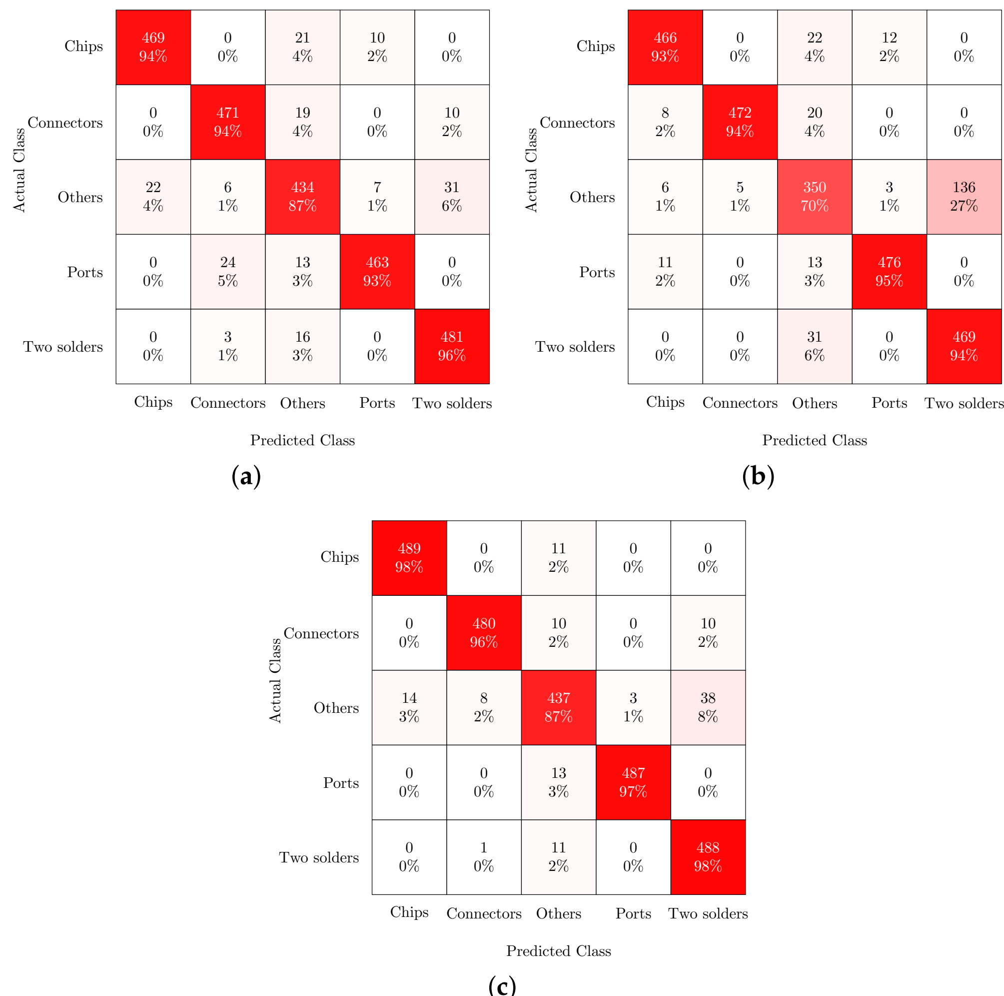

Figure 5. Confusion matrices for optical, X-ray, and fused image classifications using WaferCaps model. (**a**) Optical images; (**b**) X-ray images; (**c**) Fused images.

Table 4. Evaluative metrics for assessing test data from optical datasets across various deep learning models.

Model	Metric	Chips	Connectors	Others	Ports	Two Solders
WaferCaps	Recall	0.938	0.942	0.868	0.926	0.962
	Precision	0.956	0.935	0.863	0.965	0.921
	F1-score	0.947	0.938	0.865	0.945	0.941
ResNet-50	Recall	0.918	0.83	0.732	0.898	0.898
	Precision	0.871	0.8	0.775	0.868	0.966
	F1-score	0.894	0.815	0.753	0.883	0.931
Inception-v3	Recall	0.98	0.886	0.774	0.852	0.876
	Precision	0.939	0.807	0.74	0.962	0.946
	F1-score	0.959	0.845	0.757	0.903	0.91
MLP	Recall	0.894	0.806	0.702	0.978	0.864
	Precision	0.822	0.852	0.785	0.957	0.823
	F1-score	0.856	0.828	0.741	0.967	0.843

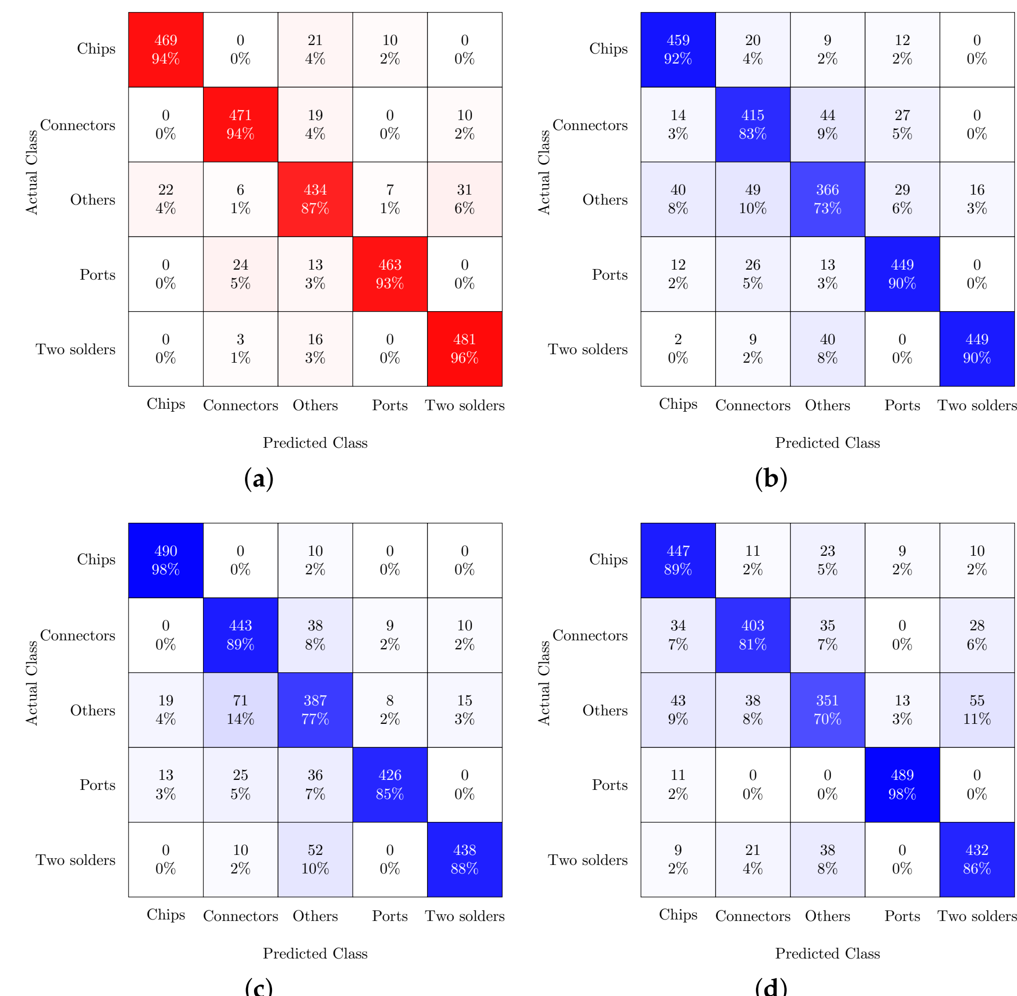

Figure 6. Confusion matrices for the optical dataset using different DL networks. (**a**) WaferCaps; (**b**) ResNet-50; (**c**) Inception-v3; (**d**) MLP.

Table 5. Evaluation of X-ray dataset test data for different deep learning models using metrics.

Model	Metric	Chips	Connectors	Others	Ports	Two Solders
WaferCaps	Recall	0.932	0.944	0.7	0.952	0.938
	Precision	0.949	0.99	0.803	0.969	0.775
	F1-score	0.94	0.966	0.748	0.961	0.849
ResNet-50	Recall	0.884	0.892	0.754	0.852	0.936
	Precision	0.989	0.933	0.709	0.986	0.836
	F1-score	0.933	0.912	0.731	0.968	0.883
Inception-v3	Recall	0.87	0.944	0.536	0.926	0.97
	Precision	0.962	0.973	0.795	0.957	0.655
	F1-score	0.913	0.958	0.64	0.941	0.782
MLP	Recall	0.868	0.856	0.528	0.924	0.842
	Precision	0.921	0.868	0.675	0.945	0.642
	F1-score	0.894	0.862	0.593	0.934	0.728

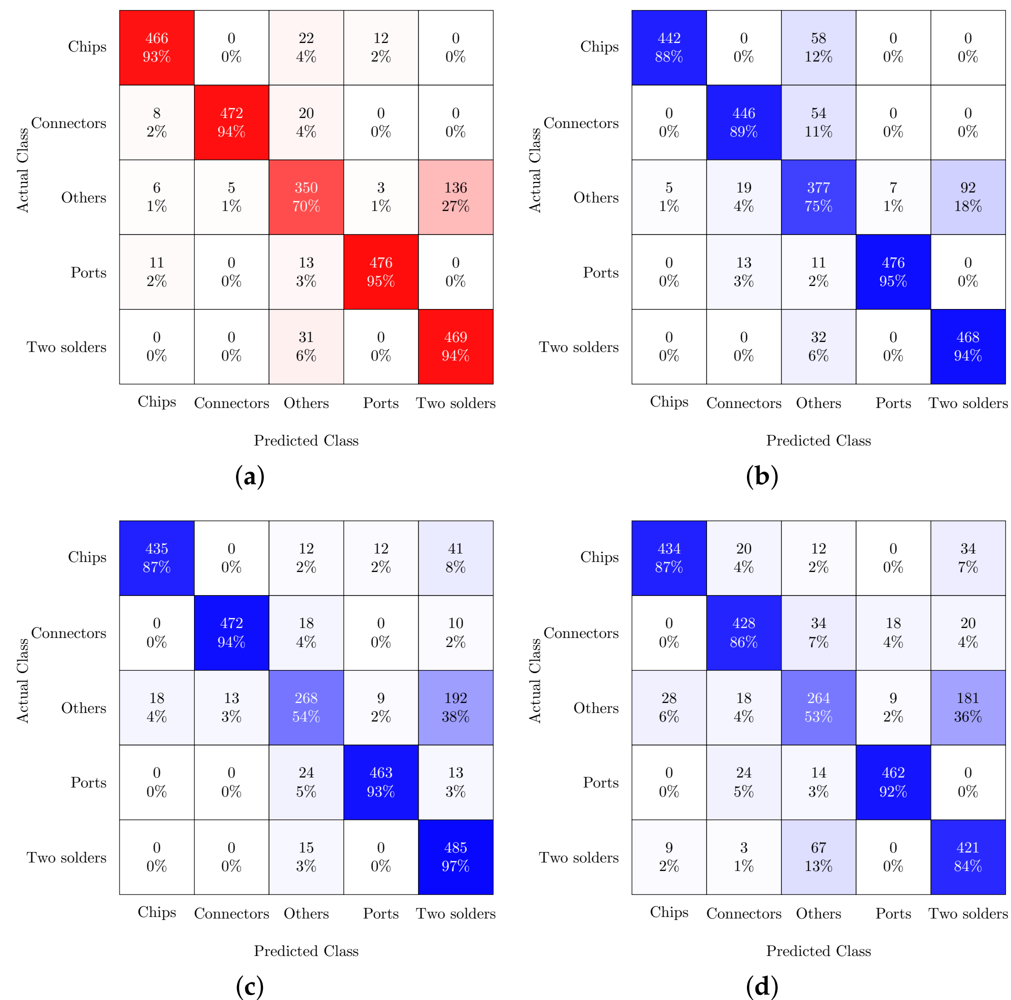

Figure 7. Confusion matrices for the X-ray dataset using different DL networks. (**a**) WaferCaps; (**b**) ResNet-50; (**c**) Inception-v3; (**d**) MLP.

Table 6 and Figure 8 illustrate the metrics and confusion matrices of four classifiers using fused images as inputs. Based on the results, WaferCaps classifies with an overall accuracy of 92.4%, which is the highest among all classifiers. The comparison of the results of Tables 4–6 shows that the WaferCaps classification method performs better on all three optical, X-ray, and fused datasets.

Table 6. Measures for evaluating test data for various deep learning models based on a fused dataset.

Model	Metric	Chips	Connectors	Others	Ports	Two Solders
WaferCaps	Recall	0.938	0.926	0.876	0.922	0.958
	Precision	0.971	0.957	0.82	0.968	0.916
	F1-score	0.954	0.941	0.847	0.945	0.936
ResNet-50	Recall	0.914	0.858	0.868	0.904	0.934
	Precision	0.912	0.947	0.781	0.944	0.914
	F1-score	0.913	0.9	0.822	0.923	0.924
Inception-v3	Recall	0.914	0.89	0.736	0.924	0.95
	Precision	0.849	0.916	0.872	0.887	0.892
	F1-score	0.881	0.903	0.798	0.905	0.92
MLP	Recall	0.834	0.854	0.816	0.926	0.912
	Precision	0.95	0.91	0.756	0.937	0.819
	F1-score	0.888	0.88	0.785	0.932	0.863

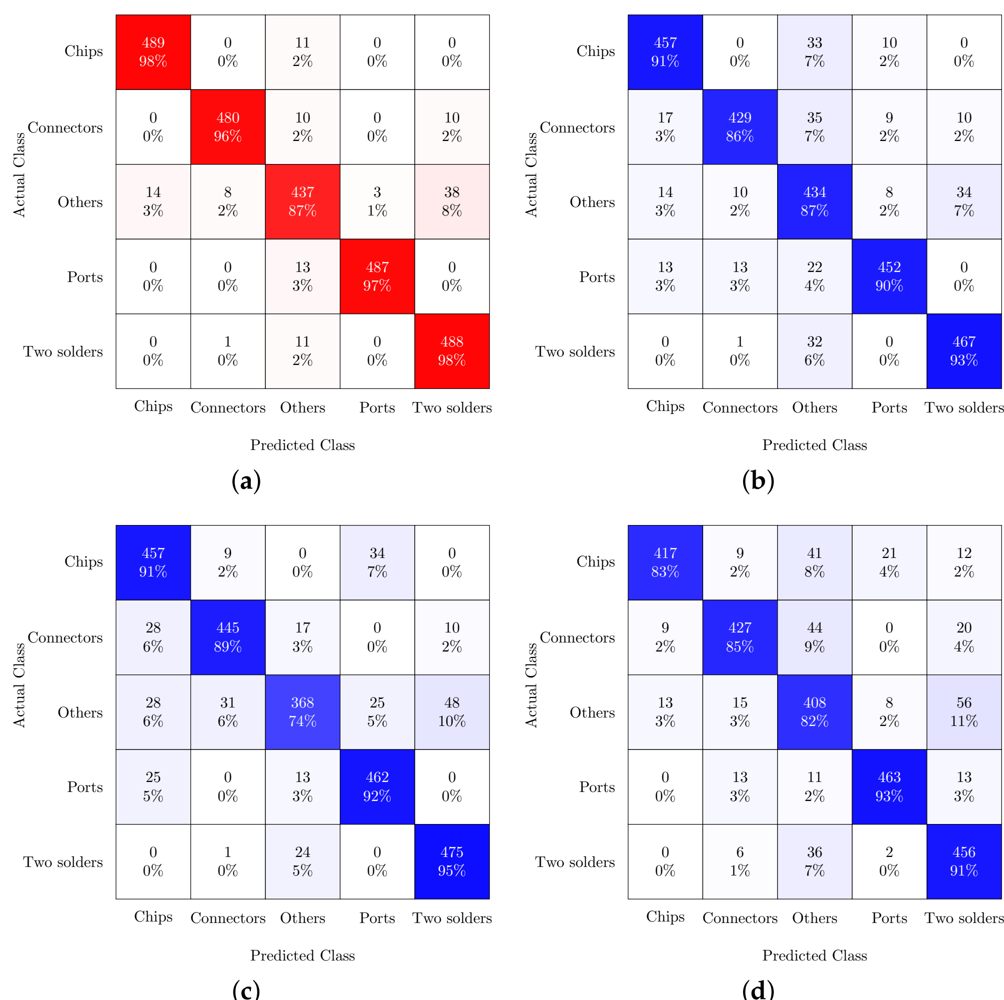

Figure 8. Confusion matrices for the fused dataset using different DL networks. (**a**) WaferCaps; (**b**) ResNet-50; (**c**) Inception-v3; (**d**) MLP.

However, using only one of these classifiers does not provide better results. Studying the results shows that the recall (sensitivity) of the WaferCaps classifier for optical images of two-solder components is higher than for X-ray and fused images. Also, in the port components classification, the results on X-ray images show superior sensitivity. Therefore, it can be expected that the decision fusion of these three classifiers can achieve much better outcomes as shown in Figure 9 and Table 7.

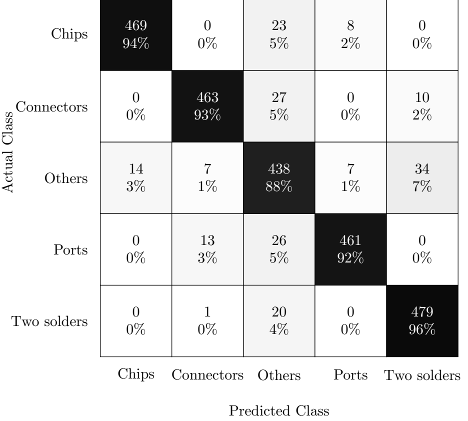

Figure 9. Confusion matrix for the decision fusion approach using WaferCaps.

Table 7. Metrics for evaluating the test data of the decision fusion approach.

Model	Metric	Chips	Connectors	Others	Ports	Two Solders
WaferCaps (Decision Fusion)	Recall	0.978	0.96	0.874	0.974	0.976
	Precision	0.972	0.982	0.907	0.994	0.91
	F1-score	0.975	0.971	0.89	0.984	0.942

5. Conclusions

This study aimed to develop a deep learning and data fusion solution that could classify PCB components into five categories (chips, connectors, ports, two-solders and other components). To accomplish this objective, we proposed a strategy encompassing image capture, data integration, and a deep learning classifier, which merges three WaferCaps-based classifiers via parallel decision fusion.

We evaluated the effectiveness of our suggested approach by juxtaposing it with the performance of leading-edge deep learning systems including WaferCaps, ResNet-50, Inception-V3, and MLP. The results demonstrated that the comprehensive performance of the decision fusion system surpassed each of the other models.

Author Contributions: Conceptualization, S.D. and A.M.; methodology, S.D. and A.A.R.M.A.E.; software, D.S. and S.D.; validation, A.M. and S.D.; writing—original draft preparation, D.S. and A.A.R.M.A.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data supporting the findings of this study are unavailable due to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abu Ebayyeh, A.M.; Mousavi, A. A Review and Analysis of Automatic Optical Inspection and Quality Monitoring Methods in Electronics Industry. *IEEE Access* **2020**, *8*, 183192–183271. [CrossRef]
2. Glučina, M.; Anđelić, N.; Lorencin, I.; Car, Z. Detection and Classification of Printed Circuit Boards Using YOLO Algorithm. *Electronics* **2023**, *12*, 667. [CrossRef]
3. Li, J.; Li, W.; Chen, Y.; Gu, J. A PCB Electronic Components Detection Network Design Based on Effective Receptive Field Size and Anchor Size Matching. *Comput. Intell. Neurosci.* **2021**, *2021*, 6682710. [CrossRef]
4. Jessurun, N.T.; Dizon-Paradis, O.P.; Tehranipoor, M.; Asadizanjani, N. SHADE: Automated Refinement of PCB Component Estimates Using Detected Shadows. In Proceedings of the 2020 IEEE Physical Assurance and Inspection of Electronics (PAINE), Virtual, 15–16 December 2020; pp. 1–6. [CrossRef]
5. Li, Y.T.; Kuo, P.; Guo, J.I. Automatic Industry PCB Board DIP Process Defect Detection System Based on Deep Ensemble Self-Adaption Method. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2021**, *11*, 312–323. [CrossRef]
6. Liu, X.; Hu, J.; Wang, H.; Zhang, Z.; Lu, X.; Sheng, C.; Song, S.; Nie, J. Gaussian-IoU loss: Better learning for bounding box regression on PCB component detection. *Expert Syst. Appl.* **2022**, *190*, 116178. [CrossRef]
7. Teramoto, A.; Murakoshi, T.; Tsuzaka, M.; Fujita, H. Development of an automated X-ray inspection method for microsolder bumps. In Proceedings of the 2005 International Symposium on Electronics Materials and Packaging, Tokyo, Japan, 11–14 December 2005; pp. 21–26. [CrossRef]
8. Wang, Z.; Ziou, D.; Armenakis, C.; Li, D.; Li, Q. A comparative analysis of image fusion methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1391–1402. [CrossRef]
9. Abu Ebayyeh, A.M.; Danishvar, S.; Mousavi, A. An Improved Capsule Network (WaferCaps) for Wafer Bin Map Classification Based on DCGAN Data Upsampling. *IEEE Trans. Semicond. Manuf.* **2022**, *35*, 50–59. [CrossRef]
10. Wu, C.H.; Wang, D.Z.; Ip, A.; Wang, D.W.; Chan, C.Y.; Wang, H.F. A particle swarm optimization approach for components placement inspection on printed circuit boards. *J. Intell. Manuf.* **2008**, *20*, 551. [CrossRef]
11. Li, J.; Gu, J.; Huang, Z.; Wen, J. Application Research of Improved YOLO V3 Algorithm in PCB Electronic Component Detection. *Appl. Sci.* **2019**, *9*, 3750. [CrossRef]
12. Schmidt, K.; Thielen, N.; Voigt, C.; Seidel, R.; Franke, J.; Milde, Y.; Bönig, J.; Beitinger, G. Enhanced X-Ray Inspection of Solder Joints in SMT Electronics Production using Convolutional Neural Networks. In Proceedings of the 2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME), Pitesti, Romania, 21–24 October 2020; pp. 26–31. [CrossRef]

13. Zhao, W.; Gurudu, S.R.; Taheri, S.; Ghosh, S.; Mallaiyan Sathiaseelan, M.A.; Asadizanjani, N. PCB Component Detection Using Computer Vision for Hardware Assurance. *Big Data Cogn. Comput.* **2022**, *6*, 39. [CrossRef]
14. Tao, H. Erasing-inpainting-based data augmentation using denoising diffusion probabilistic models with limited samples for generalized surface defect inspection. *Mech. Syst. Signal Process.* **2024**, *208*, 111082. [CrossRef]
15. Tao, H. A gated multi-hierarchical feature fusion network for recognizing steel plate surface defects. *Multimed. Syst.* **2023**, *29*, 1347–1360. [CrossRef]
16. Zhou, J.; Ren, K.; Wan, M.; Cheng, B.; Gu, G.; Chen, Q. An infrared and visible image fusion method based on VGG-19 network. *Optik* **2021**, *248*, 168084. [CrossRef]
17. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [CrossRef]
18. Li, H.; Wu, X.J.; Kittler, J. RFN-Nest: An end-to-end residual fusion network for infrared and visible images. *Inf. Fusion* **2021**, *73*, 72–86. [CrossRef]
19. Li, H.; Wu, X.J.; Kittler, J. Infrared and Visible Image Fusion using a Deep Learning Framework. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 2705–2710. [CrossRef]
20. Daneshvar, S.; Ghassemian, H. MRI and PET image fusion by combining IHS and retina-inspired models. *Inf. Fusion* **2010**, *11*, 114–123. [CrossRef]
21. Punjabi, A.; Schmid, J.; Katsaggelos, A. Examining the Benefits of Capsule Neural Networks. *arXiv* **2020**, arXiv:2001.10964.
22. Deng, F.; Pu, S.; Chen, X.; Shi, Y.; Yuan, T.; Pu, S. Hyperspectral Image Classification with Capsule Network Using Limited Training Samples. *Sensors* **2018**, *18*, 3153. [CrossRef]
23. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic Routing between Capsules. In Proceedings of the NIPS'17, 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3859–3869.
24. Wang, H.; Shao, K.; Huo, X. An improved CapsNet applied to recognition of 3D vertebral images. *Appl. Intell.* **2020**, *50*, 3276–3290. [CrossRef]
25. Patrick, M.; Adekoya, A.; Mighty, A.; Edward, B. Capsule Networks—A survey. *J. King Saud Univ.—Comput. Inf. Sci.* **2019**, *34*, 1295–1310. [CrossRef]
26. Abu Ebayyeh, A.M.; Mousavi, A.; Danishvar, S.; Blaser, S.; Gresch, T.; Landry, O.; Müller, A. Waveguide quality inspection in quantum cascade lasers: A capsule neural network approach. *Expert Syst. Appl.* **2022**, *210*, 118421. [CrossRef]
27. Xing, L.; Shao, S.; Ma, Y.; Wang, Y.; Liu, W.; Liu, B. Learning to Cooperate: Decision Fusion Method for Few-Shot Remote-Sensing Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6507505. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.