



Article An Effective Multi-Scale Feature Network for Detecting Connector Solder Joint Defects

Kaihua Zhang ¹ and Haikuo Shen ^{2,*}

- School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Ministry of Education, Beijing 100044, China; 17116375@bjtu.edu.cn
- ² Key Laboratory of Vehicle Advanced Manufacturing, Measuring and Control Technology, School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Ministry of Education, Beijing 100044, China
- * Correspondence: shenhk@bjtu.edu.cn; Tel.: +86-1774-356-8910

Abstract: With the rapid development of industry, people's requirements for the functionality, stability, and safety of electronic products are becoming higher and higher. As an important medium for power supply and information transmission functions of electronic products, high-quality soldering of cables and connectors ensures that the devices can operate normally. In this paper, we propose a multi-level feature detection network based on multi-level feature maps fusion and feature enhancement for detecting connector solder joints, classifying and locating qualified solder joints, and detecting seven common defective solder joints. This paper proposes a new feature map up-sampling algorithm and introduces a feature enhancement module, which better preserves the semantic information of higher-level feature maps, while at the same time enhancing the fused feature maps and weakening the effect of noise. Through comparison experiments, the mAP of the network proposed in this paper reaches 0.929 and the top-1 accuracy reaches 92%. The detection capability of each type of solder joint is greatly improved compared with the effect of other networks, which can assist engineers in the detection of weld joint quality and thus reduce the workload.

Keywords: connectors; solder joint defect; multi-level feature fusion; feature enhancement

1. Introduction

With the development of technology, electronic products play an important role in people's lives. Within those products, the importance of the connector as a device power supply and the core link for device information transmission cannot be ignored. High-quality solder joints ensure that the connector works properly and stably, whereas defective joints can lead to hidden problems in terms of accuracy, safety, and stability [1-3]. In the industry, the demand for cables is often insufficient to design a high-quality automatic welding machine specifically for the batch production of cables, so the main welding process of wire connectors is still manual. The task of detecting defects in solder joints of connector wires is currently performed primarily by manual inspection. Manual inspection relies on the human eye to observe the characteristics of the welding joints, then manual analysis is conducted empirically to check whether the welding joints meet the requirements of use so that the defective welding joints can be re-welded at any time to ensure that the product is used properly. However, intense workloads easily fatigue welders, which in turn reduces the quality of inspection and welding efficiency. In this case, designing an intelligent inspection method to automatically detect defective solder joints in connectors can significantly reduce the workload of engineers and improve the efficiency of manufacturing high-quality cables.

Theoretically, the use of image inspection can assist workers in analyzing welded joint defects and simplify the human visual inspection, thus improving the efficiency of detection and the quality of welding joints [4,5]. Traditional image detection algorithms



Citation: Zhang, K.; Shen, H. An Effective Multi-Scale Feature Network for Detecting Connector Solder Joint Defects. *Machines* **2022**, *10*, 94. https://doi.org/10.3390/ machines10020094

Academic Editor: Xiang Li

Received: 27 December 2021 Accepted: 22 January 2022 Published: 25 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/). can usually be divided into three parts: pre-processing, feature extraction, and recognition, among which the design of feature extraction and recognition is particularly important, and a large number of algorithms have been proposed by experts in related fields [6-10], collectively known as automatic optical inspection (AOI). AOI is widely used for the detection of solder joint defects on the surface of printed circuit boards (PCBs), in which industrial cameras are used to obtain images and design different algorithms to complete the task of detecting defects in electronic products, leading to a significant improvement in detection. Wang et al. proposed a PCB solder joint detection method based on an automatic threshold segmentation algorithm and an image morphology feature extraction algorithm, which improved the efficiency and accuracy of PCB solder joint detection [7]. Xiao et al. proposed an image acquisition path planning method and solution algorithm for PCB surface defect detection, introduced a negative feedback mechanism to address the problem of the ant colony not being able to get rid of the local optimal solution, and used a position adjustment method to solve the uncertainty of the local position of the image acquisition window, which shortened the path length without changing the time complexity [9]. Fang used Haar-like features and the AdaBoost classifier to detect three defects of multi-angle polarity capacitors and verified the effectiveness of the algorithm through experiments [10]. However, AOI usually requires specific algorithms designed according to the target, which is less universal and the complexity of the algorithm is higher in some scenarios, so it is not very widely used in industrial applications.

With the introduction of machine learning concepts [11], theories such as deep learning [12–26], transfer learning [27], reinforcement learning [28], and federation learning [29,30] have emerged and been applied to many domains, such as natural language processing [31], weather prediction [32], behavior prediction [33], fault diagnosis [34], and target detection [35,36]. Among these theories, deep learning, as a subclass of machine learning, has received extensive attention in defect detection tasks, and various defect detection algorithms based on deep neural networks have performed exceptionally well [13–19,37–40]. Li et al. improved the YOLOv3 algorithm by adding a shallow layer and using the feature map of PCB components extracted from this layer for detection, thus overcoming the small-target detection difficulties. The dataset was also expanded using data federation and data augmentation, which improved the mAP of the network by 16% [17]. Tommaso et al. improved the YOLOv3 target detection network and proposed a YOLOv3-based multi-stage model for detecting panel defects in aerial images acquired by UAVs and predicting the soil coverage area [38]. The model has high generality and is capable of handling thermal or visible images, and the detection accuracy AP@0.5 for panel defects can reach 98%. Zhao et al. proposed a lightweight convolutional neural network to achieve online defect detection on the inner surface of micro tubes [39]. The network uses a shallow segmentation network with a lightweight convolutional neural network for pixel-level crack detection, and its accuracy of defect detection reaches 98.5%, with the product of the average intersection of the segmentation and the concatenation set being 0.834. Zhang et al. proposed a cost-sensitive residual convolutional neural network [40]. The network adds a cost-sensitive adjustment layer to the standard ResNet and assigns a larger weight to a few real defects in the PCB dataset according to the class imbalance. In addition, in the design of the loss function of the network, Zhang et al. proposed optimizing the cost-sensitive residual convolutional neural network by minimizing the weighted cross-entropy loss function. Experiments show that the network can solve the problem of unbalanced class distribution of real and pseudo-defects on the PCB surface. Deep neural network-based target detection algorithms usually use convolution and pooling layers to complete feature extraction and use fully connected layers to classify the extracted features, so that no special algorithms are needed to be designed for different scenarios. Moreover, the use of weight sharing in the feature extraction stage greatly reduces the training difficulty and cost. Therefore, compared with traditional target detection algorithms, deep neural networks can solve more complex target detection tasks, and the algorithm process is simple and straightforward, which can greatly improve the detection accuracy.

Despite the fact that defect detection based on deep neural networks has been extensively explored and many related detection methods have been proposed, the detection of defects in connector solder joints is very rare. The main reasons for this situation are related to the following aspects:

- There are differences in the internal design of different electronic devices, and therefore the type of connector varies greatly. Different connectors have different pin layouts and solder joint locations, and the angles of the solder joint images can vary, making it difficult to design feature extraction algorithms that are adaptable to multiple scenarios.
- 2. Soldering of cables and connectors is different from PCB and is usually done by manual soldering. Manual welding is inevitably subject to large welding errors, such as the temperature of the soldering gun, welding time, the amount of solder used. The uncertainty of manual welding leads to a variety of manifestations of connector solder joints, and the same category of solder defects may also show different forms and textures. Therefore, it is difficult to design feature extraction algorithms and classification algorithms to effectively differentiate between solder defects.
- 3. At present, the manufacturing of small numbers of cables is still done by engineers. They have to complete the welding and manual detection of solder joint defects at the same time, and this manufacturing process hinders the development of automatic defect detection systems. It is believed that engineers can detect solder joint defects in real time during the soldering process and re-weld the defective solder joints. However, in practice, certain solder joint defects require careful observation by experienced engineers, which tends to cause engineer fatigue, leading to them overlooking certain imperceptible defects in the solder joints, which thus reduces the lifetime of the devices.

Therefore, research on intelligent detection methods for connector solder joint defects is essential. Given that deep neural networks are capable of performing defect detection tasks in various industries with their unique advantages, we believe that deep neural networks will continue to perform well in connector solder joint defect detection tasks. In our previous work [19], an improved Faster-RCNN algorithm was used to do detection on five types of solder joints with a mAP of 0.941 and a top-1 detection accuracy of 94%. The experimental results strongly demonstrate the feasibility of using deep neural networks for the connector solder joint defect detection task. However, as the first dataset of connector solder joints was simulated by us for research purposes, we did not have as high soldering skills as professional engineers, so the samples were relatively simple. For the latest dataset, we invited a professional soldering engineer to complete the soldering of the connector and the core wire. A total of eight types of solder joints were realized, and the solder joint conditions were more in line with production reality. When experimenting with the latest dataset, we found that all the metrics were significantly lower. The latest connector solder joint samples were analyzed and it was found that in the new dataset there were more connector pins, larger connector sizes, and correspondingly smaller solder joints in the image in terms of proportion and number of pixels. In other words, in the latest connector solder joints dataset, as the number of solder joints becomes larger but the scale becomes smaller, the features of the solder joints become somewhat blurred. Our previous work only used the feature maps that were down-sampled from the original image with a factor of 32 for the extraction and classification of the proposals, which is less effective for detecting small-scale, blurred solder joints.

To solve the above problems, an effective multi-scale feature network is proposed in this paper. The network uses the fusion of adjacent-stage feature maps to generate feature pyramids and introduces a feature enhancement module to further enhance the fused feature maps. Experimental results show that the network can effectively detect eight types of connector solder joints with higher detection accuracy than other mainstream target detection frameworks and has better detection for small-scale and blurred solder joint defects. The sections in this paper are organized as follows: Section 2 details the overall structure of the proposed multi-scale feature network, the highlight design in the network, the loss function used in the network, and the specifics of the dataset used in the experiments. Section 3 shows the experimental results related to the ablation experiment and the comparison experiment and provides a detailed analysis of the experimental results. Section 4 discusses the experimental results from a macro perspective and points out the focus of future work. Section 5 summarizes the research content of this paper.

The contributions of this paper are as follows:

- 1. There has been little research on the detection of defective solder joints in connectors. In this paper, a network that can effectively detect defective solder joints in connectors is proposed, which uses feature fusion and feature enhancement to extract multi-scale feature maps and improve the detection accuracy of defective solder joints in connectors.
- 2. In this paper, a higher-stage feature map up-sampling algorithm is presented, which can be applied to deep neural networks that need to fuse feature maps of different scales to improve the expressiveness of the fused feature maps. In addition, the feature enhancement module used in this paper can further enhance the expressiveness of the fused feature maps, thus improving the overall detection accuracy of the network.
- 3. This paper provides a detailed analysis of the causes, manifestations, and pitfalls of different connector solder joint defects and provides a basis for analyzing the quality of solder joints when welding connectors in industry.
- 4. There is no open source dataset for connector defective solder joints. The dataset used in this paper will be made available to other scholars who are interested in connector defective solder joint detection.

2. Materials and Methods

2.1. Network Architecture

In this paper, the details of feature pyramid network (FPN) [20] are improved by introducing a learnable feature map up-sampling algorithm as well as a feature enhancement module to fuse multi-level feature maps to enrich and highlight features. The structure of the proposed network is shown in Figure 1.

The multi-scale feature network proposed can be divided into four modules from an overall perspective, namely the feature map sets generation module, the feature enhancement module, the region proposal generation module, and the classification and regression module.

The feature map sets generation module is designed with three branches, in which the bottom-to-up branch uses ResNet-101 to extract multi-scale feature maps [C1, C2, C3, C4, C5] with scales corresponding to [1/2, 1/4, 1/8, 1/16, 1/32] compared to the original image. In all laterally connected branches, except the top branch, 1×1 convolution is used to limit the number of corresponding feature map channels to 256 in order to facilitate the fusion of feature maps in adjacent stages. In the top branch, in addition to 1×1 convolution, the highest stage feature map is further down-sampled using the maximum pooling operation to obtain the virtual higher stage feature map P6. The two are superimposed and further modified by 3×3 convolution to complete the fusion of adjacent order feature maps. In this way, the feature map generation module finally generates a pyramid-like feature map set [P2, P3, P4, P5, P6]. As P6 is simply designed for the fusion of P5 with a down-sampling ratio of 1/64, which cannot characterize the target features well, we discard P6 and only use [P2, P3, P4, P5] as the feature map set to participate in the subsequent process.



Figure 1. The structure of network proposed in this paper, where the green dashed box represents the bottom-up branch, the yellow dashed box represents the laterally connected branch, the blue dashed box represents the top-down branch, and the red solid box represents the final generated feature map set.

The feature enhancement module introduces learnable parameters from both spatial and channel directions to enhance the feature sets [P2, P3, P4, P5] so that the noise generated by different stages of feature fusion is weakened and the useful feature information is emphasized, resulting in a feature map set [F2, F3, F4, F5] with stronger feature representation capability.

The region proposal generation module uses the region proposal network (RPN) to generate proposals. Nine anchors are generated for each pixel position of the feature map at each stage of the feature map group [F2, F3, F4, F5], and the RPN corrects the anchors and maps them to the input image according to the reduced ratio of the feature map at each stage. The non-maximum suppression (NMS) algorithm is used to remove redundant proposals from all generated original proposals, and the intersection of union (IOU) between proposals and labeled boxes is used to determine whether the proposals belong to the foreground or background. In other words, RPN generates proposals at each stage of the feature map for the feature map group [F2, F3, F4, F5] and adds labels to them.

The classification and regression module uses two fully connected layer branches to calculate the category scores and location regression parameters of the proposals, respectively. As generated proposals are of variable size, the fully-connected layer requires the input to be a vector of fixed size. Therefore, RoI Align is used to reset the size of the proposals to 7×7 and subsequently calculate the category score and location regression parameters.

The feature extraction module uses ResNet-101 as the base network. The bottom-up branch generates multi-stage feature maps set [C1, C2, C3, C4, C5] from the input image, the highest-stage feature map C5 generates P6 after the max-pooling layer, and then the top-down branch fuses the current feature maps with the lower-stage feature maps in the adjacent bottom-up branches to generate a new feature map. The lateral connection branch serves to limit the number of feature map channels and allows for easy fusion of feature maps at adjacent levels. The approach borrows from the generated way of FPN feature map sets to obtain [P2, P3, P4, P5, P6]. For the purpose of highlighting the useful features, a feature enhance module is introduced to further modify the generated feature map sets, and finally the feature map sets [F2, F3, F4, F5] are yielded.

The pseudo-code of the training process of the proposed multi-scale feature network in this paper are shown in Algorithm 1.

Algorithm 1 Training process pseudo-code of multi-scale feature network.

Label boxes: t(x, y, w, h)# Label category: v # Input: *D*(*batch*, *channel*, *height*, *width*) # Max epochs: *Epochs* Initialize network parameters and load ResNet-101 pre-trained model For *epoch* in *Epochs*: Extract multi-scale feature map set $C = [C_1, C_2, C_3, C_4, C_5]$ $P_6 = Maxpool(Conv(C_5))$ For *i* in range (5, 1, -1): \rightarrow Get feature map set $F = [F_2, F_3, F_4, F_5]$ $P_{i+1}' = CAFUS(P_{i+1})$ $C'_i = Conv(C_i)$ $P_i = Conv(Concat(P'_{i+1}, C'_i))$ $F_i = FEM(P_i)$ Set anchors on *F*, get initial proposals p(x, y, w, h)Mapping *p* to *D*, get $\hat{p}(\hat{x}, \hat{y}, \hat{w}, \hat{h})$ $\hat{p} = NMS(\hat{p})$ $overlap = Sort(IOU(\hat{p}, t))$ Select 128 foreground proposals and 128 background proposals p'p' = RoIAlign(p') $cls_score = FC_{classification}(p')$ $box_reg = FC_{regression}(p')$ $Loss = CrossEntropy(cls_score, v) + SmoothL1(box_reg, t)$ Update network parameters by backpropagation

2.2. Content-Aware Feature Up-Sampling (CAFUS)

Because neighboring feature maps have different scale levels and cannot be fused, the introduction of up-sampling to amplify higher-level feature maps is a very important part of the network. Feature map up-sampling is an essential image amplification technique, and we aim to retain the maximum amount of semantic information contained in the higher-level feature maps after up-sampling, without introducing too much noise and computational effort. Image up-sampling techniques include interpolation [41] (nearest neighbor interpolation, bilinear interpolation, bicubic interpolation), deconvolution [42], and dynamic filters [26]. The interpolation method computes new pixels only by pixel positions, which does not take advantage of the semantic information of the image, and the perceptual field and computation cost cannot be balanced. For example, the nearestneighbor interpolation and bilinear interpolation methods are relatively small in computation, but the perceptual fields are only 1 imes 1 and 2 imes 2, whereas the perceptual field of bicubic interpolation can reach 3×3 , despite its large computation cost. In addition, the up-sampled images obtained by the interpolation method usually have local distortion, and the noise introduced affects the learning effect of the network for features. The deconvolution method can improve the above problems to some extent by learning the kernel parameters through convolutional networks without considering the pixel positions. However, this method does not consider local semantics and uses the same convolutional kernel for each local region, which still cannot effectively restore local feature information. Moreover, when the convolution kernel is designed to be too large, the computational effort grows exponentially. The dynamic filtering method designs a convolutional kernel for each position of the image, and, as one can imagine, the number of parameters introduced by this method is too large for practical applications.

Content-aware reassembly of features (CARAFE) [43] is a learnable image up-sampling algorithm based on the input content, which divides the feature map up-sampling into two parts, i.e., up-sampling kernel prediction and feature reassembly. Experiments show

that CARAFE has a large perceptual field; the model is sufficiently light enough to retain the feature information of the input better. More importantly, CARAFE will enhance the feature semantics to some extent. The proposed feature map up-sampling method, named content-aware feature up-sampling, improves the CARAFE algorithm, as shown in Figure 2. CAFUS modifies the bilinear interpolated image by predicting the up-sampling kernel parameters based on the input feature map, which makes up for the disadvantages of the interpolation method without introducing too many parameters that would result in excessive training cost.



Figure 2. The structure of CAFUS can be seen above. We choose a prediction kernel size of 5×5 and an up-sampling ratio of 2, so the number of channels of the encoded feature map is 100, which is $2 \times 2 \times 5 \times 5$. As for the choice of the number of compressed channels, we found that increasing the number of compressed channels had no significant improvement on the effectiveness of the algorithm.

CAFUS up-samples the input feature map F(h, w, c) using bilinear interpolation to obtain the up-sampled feature map $F'(\sigma h, \sigma w, c)$, where σ denotes the up-sampling ratio and (h, w, c) denotes the dimensionality of the input feature map. The $n \times n$ neighbourhood of the feature map F at position l is defined as $N_F^l(n)$ for easier understanding. Subsequently, the up-sampling modification kernel prediction module predicts the up-sampling modification kernel parameters based on the elements at each position *l* of the input feature map, as shown in Equations (1)–(4). The whole process is mainly implemented by two convolutional layers: the first convolution layer uses a 3×3 convolution kernel to compress the number of channels in the feature map to 64, and the second convolution layer uses a 3×3 convolution kernel to expand the number of channels in the feature map to ($\sigma \times \sigma \times n \times n$), where *n* represents the scale of the neighborhoods involved in the calculation when the features are modified, and get W_{encode} . The dimension of W_{encode} to $(\sigma \times h, \sigma \times w, n \times n)$ is resized, and the elements at each position l' is normalized, with the resulted W' containing all the up-sampling modification kernels. Finally, the elements at each position in W' are convolved with the $n \times n$ neighborhoods at the corresponding position in F' to finalise the modification of the upsampled image after bilinear interpolation, as shown in Equation (5).

$$W_{compress} = Conv_{3\times3} \left(N_F^{(x,y)}(n), kernel_{3\times3} \right), \ x \in [1,h], \ y \in [1,w],$$
(1)

$$W_{encode} = Conv_{3\times3} \left(N_{W_{compress}}^{(x,y)}(n), kernel_{3\times3} \right), \ x \in [1,h], \ y \in [1,w],$$
(2)

$$W = resize(W_{encode}), \tag{3}$$

$$W'_{(x,y)} = normalize(W_{(x,y)}), \ x \in [1, \sigma \times h], \ y \in [1, \sigma \times w],$$
(4)

$$F_{up} = N_{F'}^{(x,y)}(n) \times W'_{(x,y)}, \ x \in [1, \sigma \times h], \ y \in [1, \sigma \times w],$$

$$(5)$$

In this paper, $\sigma = 2$, n = 5.

2.3. Feature Enhancement Module (FEM)

During the generation of the feature map, as the number of convolution layers increases, the content of the extracted feature expressions deepens layer by layer. In other words, shallow convolution extracts low-level feature representations, such as points, lines, and curves, whereas deep convolution extracts high-level feature representations, such as contours, textures, and chromatic differences. Although the feature extraction module used in this paper uses a 3×3 convolution to correct the fused feature maps when fusing the feature maps of adjacent stages, conflicting representations are inevitable due to the large span of feature levels represented by the five stage feature maps extracted by ResNet-101. In this case, the fused feature maps do not fully contain both low-stage semantic information and high-stage semantic information, but instead lead to weakened feature representation. In order to solve this problem, this paper proposes a feature weighting algorithm to further enhance the feature information, i.e., the feature enhance module. The algorithm principle of the feature enhance module is shown in Figure 3. The feature enhancement module assigns weights to the feature maps from two directions, thus enhancing the useful features.



Figure 3. Details of the feature enhance module. Feature spatial enhance (FSE) fuses multi-channel information using a 1×1 convolution kernel and normalizes it using a sigmoid function to obtain two-dimensional spatial weights. Feature channel enhance (FCE) obtains a 1×256 vector using global average pooling. The vector is then calculated by two fully connected layers and normalized using the sigmoid function to obtain the one-dimensional channel weights. Enhanced features from the two approaches are superimposed together, followed by further fusion of local feature information using 3×3 convolution.

In the spatial direction, the fused feature map contains both target features and background information. The feature map is modified along the two-dimensional spatial direction to enhance the target features. The specific weight parameters are calculated as shown in Equations (6) and (7). In Equation (7), \circledast represents the multiplication of elements in the corresponding position.

$$\hat{Z}_{(x,y)} = \sigma \left(\sum_{i=1}^{C} P_{(x,y)}^{i} \cdot kernel_{(x,y)}^{1 \times 1} \right), \ x \in [1,h], \ y \in [1,w],$$
(6)

$$P_{FSE}^{c} = P^{c} \circledast \hat{Z}, \ c \in [1, C],$$

$$\tag{7}$$

In the channel direction, as each level of the feature map group contains 256 channels, not every channel of the two-dimensional feature map can contain the target feature information well. Therefore, feature map channel weights are introduced to assign a coefficient to each channel, so that the feature maps containing the target features in the feature map sets can be used to greater effect. The feature map channel weight parameters are calculated as shown in Equations (8)–(10). In Equation (4), σ represents a sigmoid function and δ represents an ReLU activation function.

$$Z^{c} = \frac{1}{h \times w} \sum_{x=1}^{h} \sum_{y=1}^{w} P_{(x,y)}, \ c \in [1, C],$$
(8)

$$\hat{Z} = \sigma(\delta(W_2 \times \delta(W_1 \times Z))), \tag{9}$$

$$P_{FCE}^c = \hat{Z}^c \times P^c, \ c \in [1, C],$$

$$\tag{10}$$

Finally, the features modified in both directions are fused together, as shown in Equation (11), and the fused feature map will contain stronger feature information.

$$P_{FEM} = Conv_{3\times3}(P_{FSE} + P_{FCE}), \tag{11}$$

2.4. Loss Function

The training of a deep neural network refers to the iterative optimization of the network parameters, where the key lies in the design of a reasonable function to calculate the network prediction error, i.e., the loss function. In the process of supervised learning, samples and labels are passed into the network at the same time, and the network predicts the results and participates in the calculation of the loss together with the input labels, subsequently updating the parameters layer by layer according to the magnitude of the loss in reverse, and finally obtaining a refined model. Image target detection requires not only determining the category of the target, but also finding the location of the target, so the total loss is mainly composed of two parts: category loss and location loss. The proposed network uses both cross-entropy and smooth-L1 to calculate the loss. Cross-entropy can accurately calculate the category loss of multi-category targets while adaptively controlling the update of parameters during back-propagation according to the loss size to avoid over-fitting of network parameters. The smooth-L1 curve is smooth, the derivatives do not change abruptly, and it has good robustness to outliers. The network proposed in this paper is two-stage, and the losses of the main network and the region proposed network need to be calculated separately, as shown in Equations (12)–(16).

$$L_{RPN_cls} = \frac{1}{N_{batch}} \sum_{i=1}^{N_{batch}} cross_entro(p_i, p_i^*),$$
(12)

$$L_{RPN_bbox} = \frac{\lambda}{N_{boxes}} \sum_{i=1}^{N_{boxes}} p_i^* \times smooth_L1(t_i, t_i^*),$$
(13)

where N_{batch} represents the number of images in one mini-batch at training, N_{boxes} represents the number of anchors generated in each image, and λ represents the balance factor. The value p_i refers to the category score vector of the proposal box, p_i^* represents the label of the proposal box, t_i represents the position parameter of the proposal box, and t_i^* represents the position parameter of the proposal box.

$$L_{RCNN\ cls} = cross_entro(p, u), \tag{14}$$

$$L_{RCNN \ bbox} = smooth_L1(t^u, v), \tag{15}$$

where p represents the category score predicted by the network, u represents the score of the ground truth, t_u represents the coordinate of the ground truth, and v represents coordinate of the predicted box.

$$Loss = L_{RPN \ cls} + L_{RPN \ bbox} + L_{RCNN \ cls} + L_{RCNN \ bbox}, \tag{16}$$

2.5. Description of Connector Solder Joint Dataset

Connector solder joints have many types of defects, and defective joints will affect the accuracy, safety, and stability of the connector during use. We use multi-level feature fusion and feature enhancement to complete the task of detecting qualified solder joints and seven types of defective solder joints; the details of the solder joints to be detected can be seen in Table 1 and Figure 4.

Туре	Description	Negative		
Qualified solder	The surface of the solder joint is polished and smooth, with the solder filling the entire solder cup and no obvious overflow.			
Multi-solder	Solder overflows the solder cup and the overflow exceeds the pin spacing by more than 50%.	The risk of connector short-circuiting is increased and therefore the connector has insufficient safety.		
Less-solder	Solder fails to fill more than 75% of the solder cup.	The solder joints are not strong enough for the connection and are prone to open circuit, thus indicating insufficient stability.		
Connected-solder	Solder spills out of the solder cup and touches the adjacent pin.	The connector is short-circuited and becomes out of function.		
Tip-solder	Obvious tip pulling can be observed on the surface of the solder joints.	Solder peels off easily, which can cause short circuits in the connector and affects the stability of use.		
Rough-solder	Uneven heating of the solder and rough surface of the solder joints.	Insufficient local strength of the solder joint connection, which affects the stability of the connector.		
Bare wire The core wire outside the solder cup is not protected by an insulated rubber bushing.		The exposed core wire is easily corroded or broken, thus resulting in short-circuiting or open circuit, and therefore making the connector less stable.		
Skewed wire The core wire does not fit perfectly into the solder cup and there is a large angular deviation.		The core wire peels off easily, causing the connector to disconnect and affecting the stability of use.		

 Table 1. Detailed description of solder joints to be detected.



Figure 4. Type of the solder joint to be detected: (**a**) qualified solder, (**b**) multi-solder, (**c**) less-solder, (**d**) connected-solder, (**e**) tin-solder, (**f**) rough-solder, (**g**) skewed wire, and (**h**) bare wire.

3. Results

Currently, there is very little research on connector solder joint defect detection. Therefore, a common deep convolutional neural network-based image target detection network was chosen to compare with the multi-scale feature network proposed in this paper. Additionally, in order to verify the effectiveness of the feature map up-sampling algorithm CA-FUS and the feature enhancement module FEM, corresponding ablation experiments were conducted. In terms of hardware, all experiments in this paper use NVidia GTX1050Ti 8G GPU and Inter Core I7-8750h CPU. In terms of software, due to the different libraries used by the open source framework, Faster-RCNN [21], SSD [25], and our previous work [19] were implemented using Tensorflow, whereas YOLO [22], FPN [20], and the proposed multi-scale feature network were implemented using Pytorch. In all experiments, simple names were used to indicate the types of connector solder joint. Where "hg" denotes qualified solder joints, "dx" represents multi-solder joints, "sx" denotes less-solder joints, "xj" denotes tip-solder joints, "lh" refers to connected-solder joints, "cc" denotes rough-solder joints, "lx" denotes bare wire joints, and "wx" denotes skewed wire joints.

Figure 5 shows the variation of the total loss of the network during the training process of multi-scale feature network. The number of iterations was set as 60,000 and the total loss of the network was recorded every 100 iterations, and the recorded results were plotted as curves. It is not difficult to find that at the beginning of the network training, the losses are larger and the magnitude of the loss oscillation is obvious. As the network parameters are progressively optimized, the training loss decreases. After 10,000 steps of training, the loss decreases less and after 40,000 steps of training, the loss stabilizes at around 0.05. It can be seen that the multi-scale feature network optimizes the network parameters well during the training process, the final total loss is very low, and the network accuracy is high.

In a multi-classification target detection task, each category can be plotted on a curve, known as a P-R curve, based on recall and accuracy. The average precision (AP) is the area under that curve, and the mean average precision (mAP) is the average value of the APs for each category. The mAP is one of the important indicators of how good a deep learning model is, and its value ranges between 0 and 1: the higher the value, the better the model. To verify the effectiveness of CAFUS and FEM in improving the detection capability of the network, corresponding ablation experiments were conducted to train FPN, FPN with CAFUS, FPN with FEM, and the proposed multi-scale feature network. The experimental results are presented in Figure 6 and Table 2. Comparing the experimental results, it was found that a 5.4% increase in the mAP of the network from 0.817 to 0.871 could be achieved using CAFUS, compared to the FPN using bilinear interpolation for

feature map up-sampling. Correspondingly, the introduction of FEM increases the mAP of the network by 0.862. A mAP of 0.929 was achieved when using both CAFUS and FEM for the multi-scale feature network, an increase of 11.2% compared to FPN. This shows that the use of CAFUS can guarantee the up-sampling quality of the high-level feature maps during the multi-scale feature map fusion. Additionally, FEM has a better enhancement effect on the feature representation ability of the feature map.







Figure 6. P-R curves of different methods in ablation experiment: (**a**) FPN; (**b**) FPN with CAFUS; (**c**) FPN with FEM; (**d**) multi-scale feature network.

	FPN	FPN with CAFUS	FPN with FEM	MFN ¹
hg	0.851	0.866	0.861	0.908
dx	0.787	0.864	0.858	0.931
SX	0.822	0.855	0.876	0.909
lh	0.938	0.888	0.875	0.968
xj	0.725	0.816	0.801	0.900
cc	0.651	0.801	0.763	0.856
lx	0.909	0.981	0.969	0.998
WX	0.855	0.898	0.890	0.961
Mean AP	0.817	0.871	0.862	0.929

Table 2. APs and mAPs of different methods in the ablation experiment.

¹ Multi-scale feature network.

In order to better demonstrate the advantages of the multi-scale feature network in the detection task of connector solder joint defect, several commonly used detection frameworks that are well recognized in the field of image target detection, namely Faster-RCNN, YOLO, SSD, and FPN, were chosen for comparison of detection results. In our previous work, an improved Faster-RCNN was also added to the comparison experiments. Figure 7 shows the P-R curves of the different frameworks. The APs for each connector solder joint and mAPs of the different frameworks are recorded in Table 3, and the top-1 detection accuracy of the connector solder joints in 92 test images for the different frameworks are recorded in Table 4. The experimental results show that the multi-scale feature network significantly outperforms other frameworks in a number of metrics, with the network achieving a mAP of 0.929, which is improved by 20.3%, 23.1%, 18.7%, 14.3%, and 11.2%, respectively, compared to the other networks. Also, in terms of actual detection accuracy, the multi-scale feature network is able to detect connector solder joint defects better than other frameworks, with a higher top-1 detection accuracy of 92% on average than the other frameworks.

Table 5 lists the test time per image for the target detection frameworks involved in the experiments. Being a two-stage framework, its detection time consumption is higher compared to the one-stage framework, but the introduction of CAFUS and FEM does not make the network more complex compared to FPN, and the detection times are almost identical.

Some results of the connector solder joint detection framework proposed in this paper are shown in Figure 8.

	Faster-RCNN	YOLO	SSD	Improved Faster-RCNN	FPN	MFN ¹
hg	0.740	0.682	0.729	0.847	0.851	0.908
dx	0.643	0.613	0.658	0.780	0.787	0.931
SX	0.714	0.745	0.698	0.805	0.822	0.909
lh	0.886	0.861	0.937	0.902	0.938	0.968
xj	0.659	0.624	0.652	0.685	0.725	0.900
cc	0.532	0.429	0.573	0.479	0.651	0.856
lx	0.903	0.902	0.899	0.909	0.909	0.998
WX	0.733	0.725	0.791	0.880	0.855	0.961
Mean AP	0.726	0.698	0.742	0.786	0.817	0.929

Table 3. APs and mAPs of different methods in the ablation experiment.

¹ Multi-scale feature network.

	Faster-RCNN	YOLO	SSD	Improved Faster-RCNN	FPN	MFN ¹
hg	78%	75%	81%	87%	88%	94%
dx	69%	66%	74%	77%	83%	91%
SX	68%	63%	74%	79%	84%	95%
lh	86%	84%	89%	95%	92%	97%
xj	72%	66%	80%	86%	92%	96%
cc	60%	50%	65%	65%	75%	80%
lx	89%	89%	93%	89%	89%	96%
WX	68%	58%	68%	74%	79%	89%
Mean Top-1	74%	69%	78%	82%	85%	92%

Table 4. Top-1 accuracy for each type of solder joint in different frameworks.

¹ Multi-scale feature network.



Figure 7. P-R curves of different method in ablation experiment: (a) Faster-RCNN; (b) YOLO; (c) SSD; (d) Improved Faster-RCNN; (e) FPN; (f) multi-scale feature network.

	Faster-RCNN	YOLO	SSD	Improved Faster-RCNN	FPN	MFN ¹
Average testing time	0.28	0.22	0.26	0.31	0.45	0.47

Table 5. Average testing time per image in different frameworks.

¹ Multi-scale feature network.



Figure 8. Classification and positioning results of solder joint instances in the connectors.

4. Discussion

In this paper, the FPN was optimized and a more efficient multi-level feature fusion and feature enhancement network was proposed to solve the connector solder joint defect detection problem. In our latest dataset, many types of solder joints contain a small number of pixels and the detection of solder joints is converted to a small target detection problem. Additionally, as most connectors are soldered manually, multiple uncertainties lead to insignificant inter-class differences in connector solder joints, making the detection of connector solder joints difficult. The experimental results in Table 4 show that the proposed framework using top-level feature maps for extraction and classification is grossly inadequate for detecting a variety of small-scale solder joint defects, such as multi-solder, less-solder, tip-solder, rough-solder, and skewed wire, whereas it is better for detecting other large-scale solder joint defects. The main reason for this phenomenon is that the down-sampling ratio of highest-level feature map is too large compared to the input image, so the feature representation capability of the highest-level feature map is weak and the network cannot optimize the parameters based on the existing feature information. Multiscale feature fusion can fuse high-level features and low-level features, which reduces the difficulty of self-learning of the detection network, and therefore greatly improves the detection capability of small-scale targets while ensuring that the accuracy of detection of large-scale targets is not degraded.

The multi-level feature fusion and feature enhancement network proposed in this paper uses content-aware feature up-sampling instead of bilinear interpolation to complete the up-sampling of the high-level feature maps, and it introduces the feature enhancement module to further highlight the useful features. To verify the effectiveness of CAFUS and FEM, relevant ablation experiments were performed, and the experimental results are presented in Table 2. The experimental results show that the accuracy of the network is greatly improved by using both CAFUS and FEM. It can be demonstrated that CAFUS can retain the feature information of the high-level feature maps after sampling to a greater extent, and the enhancement effect of FEM on the feature information is also obvious.

As shown in Table 5, the detection time of the multi-level feature fusion and feature enhancement network proposed in this paper is 0.47 s, which means that two images can be detected per second. This result indicates that although the framework proposed in this paper has advantages in detection capability, it cannot meet the requirement of real-time detection. The main reason for this result is that the structure of the network proposed in this paper is relatively complex, so simplifying the network will be the focus of our work in the future.

5. Conclusions

In this paper, a multi-level feature fusion and feature enhancement network is proposed to solve the connector solder joint detection problem, which achieves the detection and localization of eight solder joint types, resulting in a large improvement in model accuracy and detection accuracy compared to other mainstream frameworks and our previous work. In addition, content-aware feature up-sampling was employed in the network to up-sample the high-level feature maps, and a feature enhancement module was used to further enhance the feature representation of the fused feature maps. Experiments show that CAFUS and FEM improve the effect of the network significantly.

Nevertheless, there is still much room for improvement regarding the research in this paper. First, the network proposed in this paper only achieves high accuracy detection in eight types of common connector solder joints, and the detection effect for other defects still needs to be studied. Second, the network in this paper is relatively complex, which greatly increases the detection time of the network for each image and cannot meet the requirements of real-time detection. Therefore, in future work, the detection of connector solder joints will be further improved by analyzing more types of connector solder joint defects and using more novel algorithms. At the same time, we will introduce ideas such as federated learning to improve network training methods and enhance the effectiveness of the network. Finally, we will also focus more on the complexity of the network to make it meet the requirements of real-time detection.

Author Contributions: Conceptualization, H.S.; methodology, K.Z.; software, K.Z.; validation, K.Z. and H.S.; formal analysis, K.Z.; investigation, K.Z.; data curation, H.S.; writing—original draft preparation, K.Z.; writing—review and editing, K.Z.; project administration, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Ma, D.; Lei, X.; Zhao, H. The Research of Qualification Detection of Cable Joint Solder Joint Based on DCNN. In Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), Dublin, Ireland, 16–18 October 2019.
- 2. Lei, X. Research on Qualification Detection of Cable Joint Solder Joint Based on Deep Learning; Lanzhou University of Technology: Lanzhou, China, 2020.
- Zhaoguo, W.; Weisheng, H.; Haibing, Z.; Bangfei, D.; Qian, W.; Hong, X. Research on live detection technology of cable joint defects based on high-speed light sensing and pressure wave method. J. Phys. Conf. Ser. 2021, 1871, 012015. [CrossRef]
- 4. Hani, A.F.M.; Malik, A.S.; Kamil, R.; Thong, C.-M. A review of SMD-PCB defects and detection algorithms. *SPIE Proc.* 2012, 8350, 60.
- Lee, Y.; Lee, J. 2D Industrial Image Registration Method for the Detection of Defects. J. Korea Multimed. Soc. 2012, 15, 1369–1376. [CrossRef]
- Li, Z.; Yang, Q. System design for PCB defects detection based on AOI technology. In Proceedings of the 2011 4th International Congress on Image and Signal Processing, Shanghai, China, 15–17 October 2011; Volume 4, pp. 1988–1991.
- Wang, S.Y.; Zhao, Y.; Wen, L. PCB welding spot detection with image processing method based on automatic threshold image segmentation algorithm and mathematical morphology. *Circuit World* 2016, 42, 97–103. [CrossRef]
- Mei, S.; Wang, Y.; Wen, G.; Hu, Y. Automated Inspection of Defects in Optical Fiber Connector End Face Using Novel Morphology Approaches. *Sensors* 2018, 18, 1408. [CrossRef]
- Xiao, Z.; Wang, Z.; Liu, D.; Wang, H. A path planning algorithm for PCB surface quality automatic inspection. *J. Intell. Manuf.* 2021, 2021, 1–13. [CrossRef]
- Fang, J.; Shang, L.; Xiong, K.; Gao, G.; Zhang, C. An Automatic Optical Inspection Algorithm of Capacitor Based on Multi-angle Classification and Recognition. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2020; Volume 1646, p. 012031. [CrossRef]
- 11. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
- 12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. Acm.* 2017, 60, 84–90. [CrossRef]
- 13. Huang, W.; Hua, G.; Yu, Z.; Liu, H. Recurrent spatial transformer network for high-accuracy image registration in moving PCB defect detection. *J. Eng.* **2020**, 2020, 438–443. [CrossRef]
- 14. Hao, W.; Wenbin, G.; Xiangrong, X. Solder Joint Recognition Using Mask R-CNN Method. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2019**, *10*, 525–530.
- 15. Dai, W.; Mujeeb, A.; Erdt, M.; Sourin, A. Soldering defect detection in automatic optical inspection. *Adv. Eng. Inform.* **2020**, 43, 101004. [CrossRef]
- 16. Mujeeb, A.; Dai, W.; Erdt, M.; Sourin, A. One class based feature learning approach for defect detection using deep autoencoders. *Adv. Eng. Inform.* **2019**, *42*, 100933. [CrossRef]
- 17. 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]
- Cai, N.; Cen, G.; Wu, J.; Li, F.; Wang, H.; Chen, X. SMT Solder Joint Inspection via a Novel Cascaded Convolutional Neural Network. *IEEE Trans. Compon. Packag. Manuf. Technol.* 2018, *8*, 670–677. [CrossRef]
- 19. Zhang, K.; Shen, H. Solder Joint Defect Detection in the Connectors Using Improved Faster-RCNN Algorithm. *Appl. Sci.* **2021**, *11*, 576. [CrossRef]
- Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Honolulu, HI, USA, 22–25 July 2017; pp. 2117–2125.
- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 1137–1149. [CrossRef]
- 22. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- 23. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 7263–7271.
- 24. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. arXiv 2018, arXiv:1804.02767v1.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, The Netherlands, 2016.
- Jo, Y.; Oh, S.W.; Kang, J.; Kim, S.J. Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 3224–3232.
- Pio, G.; Mignone, P.; Magazzù, G.; Zampieri, G.; Ceci, M.; Angione, C. Integrating genome-scale metabolic modelling and transfer learning for human gene regulatory network reconstruction. *Bioinformatics* 2021, *38*, 487–493. [CrossRef]

- Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* 2019, 518, 529–533. [CrossRef]
- Chhikara, P.; Singh, P.; Tekchandani, R.; Kumar, N.; Guizani, M. Federated Learning Meets Human Emotions: A Decentralized Framework for Human–Computer Interaction for IoT Applications. *IEEE Internet Things J.* 2020, *8*, 6949–6962. [CrossRef]
- Poap, D.; Woniak, M. A hybridization of distributed policy and heuristic augmentation for improving federated learning approach. *Neural Netw.* 2021, 146, 130–140. [CrossRef] [PubMed]
- 31. Deshmukh, P.R.; Phalnikar, R. Information extraction for prognostic stage prediction from breast cancer medical records using NLP and ML. *Med. Biol. Eng. Comput.* **2021**, 59, 1751–1772. [CrossRef] [PubMed]
- 32. Bihlo, A. A generative adversarial network approach to (ensemble) weather prediction. *Neural Netw.* **2021**, *139*, 1–16. [CrossRef] [PubMed]
- Hema, D.D.; Kumar, K.A. Hyperparameter optimization of LSTM based Driver's Aggressive Behavior Prediction Model. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021.
- Qi, G.; Zhu, Z.; Erqinhu, K.; Chen, Y.; Chai, Y.; Sun, J. Fault-diagnosis for reciprocating compressors using big data and machine learning. *Simul. Model. Pract. Theory* 2018, 80, 104–127. [CrossRef]
- Pang, J.; Li, C.; Shi, J.; Xu, Z.; Feng, H. R2-CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images. arXiv 2019, arXiv:1902.06042.
- 36. Wu, X.; Hong, D.; Tian, J.; Chanussot, J.; Li, W.; Tao, R. ORSIm Detector: A Novel Object Detection Framework in Optical Remote Sensing Imagery Using Spatial-Frequency Channel Features. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5146–5158. [CrossRef]
- Yazdani, N.M. Defect Detection in CK45 Steel Structures through C-scan Images Using Deep Learning Method. *Artif. Intell. Adv.* 2021, 3, 8.
- Di Tommaso, A.; Betti, A.; Fontanelli, G.; Michelozzi, B. A Multi-Stage model based on YOLOv3 for defect detection in PV panels based on IR and Visible Imaging by Unmanned Aerial Vehicle. *arXiv* 2021, arXiv:2111.11709.
- 39. Xinyu, Z.; Bin, W. Algorithm for real-time defect detection of micro pipe inner surface. Appl. Opt. 2021, 60, 9167–9179.
- 40. Zhang, H.; Jiang, L.; Li, C. CS-ResNet: Cost-sensitive residual convolutional neural network for PCB cosmetic defect detection. *Expert Syst. Appl.* **2021**, *185*, 115673. [CrossRef]
- 41. Zhong, B.; Lu, Z.; Ji, J. Review on Image Interpolation Techniques. J. Data Acquis. Processing 2016, 31, 1083–1096.
- 42. Dumoulin, V.; Visin, F. A guide to convolution arithmetic for deep learning. arXiv Prepr. 2016, arXiv:1603.07285.
- 43. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. CARAFE: Content-Aware ReAssembly of FEatures. *arXiv* 2019, arXiv:1905.02188.