ARTICLE

Industrial Fusion Cascade Detection of Solder Joint

Chunyuan Li[1,2,3], Peng Zhang[1,2,3], Shuangming Wang[4], Lie Liu[4] and Mingquan Shi[2,*]

[1]Chongqing School, University of Chinese Academy of Sciences, Chongqing University, Chongqing, 400044, China

[2]Automatic Reasoning and Cognition Research Center, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, 400714, China

[3]College of Artificial Intelligence, Chongqing School, University of Chinese Academy of Sciences, Chongqing, 400714, China

[4]Design Department, Liuzhou Fuzhen Bodywork Industrial Co., Ltd., Liuzhou, 545006, China

*Corresponding Author: Mingquan Shi. Email: shimq@cigit.ac.cn

ABSTRACT

With the remarkable advancements in machine vision research and its ever-expanding applications, scholars have increasingly focused on harnessing various vision methodologies within the industrial realm. Specifically, detecting vehicle floor welding points poses unique challenges, including high operational costs and limited portability in practical settings. To address these challenges, this paper innovatively integrates template matching and the Faster RCNN algorithm, presenting an industrial fusion cascaded solder joint detection algorithm that seamlessly blends template matching with deep learning techniques. This algorithm meticulously weights and fuses the optimized features of both methodologies, enhancing the overall detection capabilities. Furthermore, it introduces an optimized multi-scale and multi-template matching approach, leveraging a diverse array of templates and image pyramid algorithms to bolster the accuracy and resilience of object detection. By integrating deep learning algorithms with this multi-scale and multi-template matching strategy, the cascaded target matching algorithm effectively accurately identifies solder joint types and positions. A comprehensive welding point dataset, labeled by experts specifically for vehicle detection, was constructed based on images from authentic industrial environments to validate the algorithm's performance. Experiments demonstrate the algorithm's compelling performance in industrial scenarios, outperforming the single-template matching algorithm by 21.3%, the multi-scale and multi-template matching algorithm by 3.4%, the Faster RCNN algorithm by 19.7%, and the YOLOv9 algorithm by 17.3% in terms of solder joint detection accuracy. This optimized algorithm exhibits remarkable robustness and portability, ideally suited for detecting solder joints across diverse vehicle workpieces. Notably, this study's dataset and feature fusion approach can be a valuable resource for other algorithms seeking to enhance their solder joint detection capabilities. This work thus not only presents a novel and effective solution for industrial solder joint detection but lays the groundwork for future advancements in this critical area.

KEYWORDS

Cascade object detection; deep learning; feature fusion; multi-scale and multi-template matching; solder joint dataset

1 Introduction

With the advancement of computer vision and artificial intelligence (AI) technology, computer vision systems can rapidly acquire, process, and analyze large volumes of image data. They easily integrate with design, process, and control systems. These systems have been widely adopted in commercial and industrial sectors [1]. In particular, computer vision detection technology has been extensively studied and applied to identify various defects and workpiece types on industrial assembly lines [2,3].

The industrial production environment of vehicle floor parts is complex and noisy, often relying on manual detection and with less data storage. This results in insufficient data samples and uneven distribution of samples, which seriously hinders the detection process. Different from standard database images, industrial images are plagued by noise, crowding, blur, low quality, and unstable pixel values, making it challenging to meet the detection requirements. Template matching algorithms are simple to implement and effective for small target detection with high real-time performance. They are sensitive to image interference and lack robust feature extraction, which limits their detection accuracy.

On the other hand, Faster Region-based Convolutional Neural Networks (Faster RCNN) can achieve high accuracy and end-to-end detection and are suitable for complex scenes [4,5]. However, it requires a large amount of data for training data and needs help with detection accuracy when dealing with small samples or small targets. In some resource-constrained application scenarios, deep learning-based object detection systems may require more effort to deploy. Therefore, a new target detection method is needed in industrial production to solve these problems.

This paper focuses on optimizing the solder joint features and the detection algorithm framework based on specific scenarios and practical requirements. An industrial fusion cascade detection algorithm for solder joints has been proposed. Fig. 1 illustrates the principle block diagram of the solder joint algorithm, which is based on industrial fusion cascade detection. It comprises five parts: image input, dataset creation, fusion feature module, fusion cascade detection module, and results output. The dashed box on the left represents the feature-weighted fusion module. The dashed box on the right represents the fusion cascade detection module. The proposed algorithm integrates features from both algorithms and addresses their shortcomings through a fused detection process. Template matching serves as the basis for detecting qualified solder joints. Faster RCNN supplements the detection of other solder joint types, making the approach more suitable for solder joint detection in various workpieces. The fused cascade object detection algorithm is also more adaptable in different object detection scenarios.
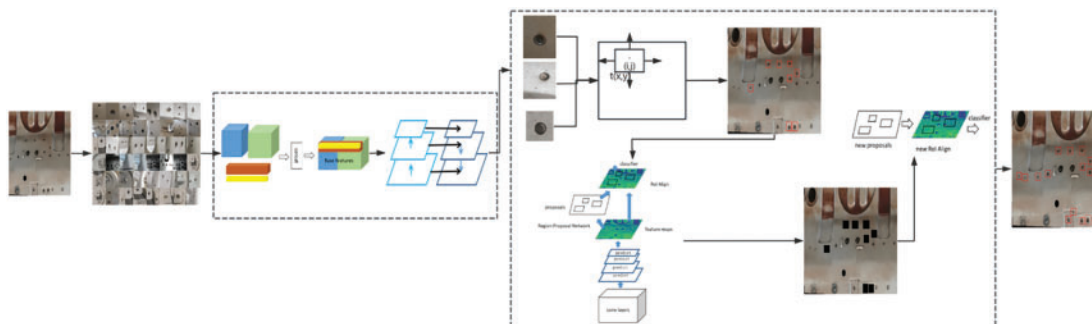

Figure 1: Principle block diagram of the solder joint algorithm based on industrial fusion cascade detection

The main contributions of this work are as follows:

(1) Collected industrial site images and created a dataset labeled by experts for vehicle solder joints.

(2) Improved the fusion method for solder joint features and proposed a cascaded detection algorithm based on multi-scale and multi-template matching combined with optimized Faster RCNN.

(3) Conducted experimental tests on real-world datasets.

The second section reviews the latest work on object detection and detection algorithms for vehicle solder joints. The third section presents the industrial cascade solder joint fusion detection algorithm model. The fourth section conducts experimental analysis, and the fifth section concludes with a summary of the findings.

2 Related Work

Object detection algorithms are widely studied, and models are rapidly evolving. Cai et al. [6] proposed a multi-stage object detection architecture called Cascade Region-based Convolutional Neural Network (Cascade R-CNN), which addresses issues such as overfitting during training and quality mismatch during inference. Ge et al. [7] cleverly integrated and advanced in the fields of object detection—such as decoupling heads, data augmentation, anchor-free methods, and label classification—with You Only Look Once (YOLO) to create a new object detection algorithm You Only Look Once, extreme (YOLOX). Wang et al. [8] tracked the problem of information loss during data transmission, proposed the PGI and GELAN architectures, and developed the YOLOv9 detection model, which is suitable for both lightweight and large models. Zuo et al. [9] constructed a large dataset of 5000 images guided by experts and used the latest YOLOv3 to detect solder joint defect types in greater detail. Instead of using original images, Cardellicchio et al. [10] employed Lidar scanning to reconstruct the three-dimensional (3D) artifacts, extracting feature classification output from the point cloud. They then used deep learning networks for detection, achieving an accuracy of over 99.7%. Recently, attention mechanisms have been introduced into object detection, with additional attention modules generating features that enhance detection accuracy [11,12]. Ge et al. formulated attention response maps as extra objective functions, combining them with the original detection loss to train detectors in an end-to-end manner [13]. While these studies provide valuable insights into detection accuracy and effectiveness, they have yet to achieve optimal results in detecting small targets.

Specifically, it is crucial to detect the shape and presence of components on the vehicle manufacturing assembly lines and accurately identify the various types of solder joints. While some researchers focus on improving detection speed and accuracy, others explore new techniques and application scenarios. Han et al. [14] conducted the weld inspection using an improved Hough transform detection algorithm. Borovkov et al. [15] demonstrated the applicability of the ultrasonic echo method for quality control of welded joints produced by spot friction stir welding. Lin et al. [16] utilized Support Vector Machines (SVM), Backpropagation Neural Networks (BPNN), and AdaBoost classifiers for solder joint quality inspection. Xie et al. [17] established a white body solder joints database and proposed a detection method based on the lightweight YOLOv7.

Currently, various target detection algorithms focus on enhancing detection speed and accuracy, each with its strengths and weaknesses in terms of precision and efficiency. However, differences in sample databases, detection standards, heavy workload, and interference factors such as workpiece deformation and occlusion result in poor detection accuracy. Therefore, there is a need for a solder joint detection method that can adapt to different types of solder joints while minimizing the impact of

image interference. Template matching algorithm, Faster RCNN algorithm, and YOLOv9 [5,8,18] have been applied to detect vehicle floor solder joints. However, the detection results need to be sufficiently accurate for practical application, and the detection accuracy for various types of solder joints still needs to be improved, as shown in the fourth section.

Analyzing the results and detection models reveals that feature description and algorithm construction are the main challenges. The use of general detection algorithms for vehicle solder joint detection has resulted in insufficient solder joint data and a lack of understanding of the particular detection objects, leading to incomplete feature representation. The limited feature extraction ability of simple detection network structures further contributes to poor detection performance. These shortcomings prevent the algorithms from meeting the requirements for detecting various types of solder joints [5,19]. Therefore, this paper aims to optimize and enhance solder joint features and the detection network by combining original image features with deep learning features and integrating template matching and Faster RCNN detection processes.

3 Measurement

Template matching is good at identifying and locating specific small targets in images, while Faster R-CNN is particular for handling multi-scale and multi-target detection tasks. Integrating features across different scales is crucial in enhancing performance [20].

3.1 Template Matching

In the actual welding process, the shape and color of solder joints are affected by various factors. They limit the detection accuracy of traditional single template matching, resulting in suboptimal defect detection. Image features—color, texture, shape, and others—are crucial elements extracted from an image that describes its content. Familiar image features include RGB color values, grayscale values, morphology, closure, texture features, and spatial relationships. Unlike traditional template matching, which utilizes a single feature, this paper employed ablation experimental results and statistical analysis to compare several types of features. The selected features included representative attributes such as color and morphological and geometric characteristics to identify qualified solder joints. Based on the analysis of the experimental results and using a weighted averages method, the image features are defined as follows:

$$T_t(x, y) = r_1 \times r + r_2 \times G(x, y) + r_3 \times E(m, n), \tag{1}$$

Referencing the above equation, the features of the solder joint are described as follows:

$$r = \sqrt{(x - a)^2 + (y - b)^2}, \tag{2}$$

$$G(x, y) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2}, \tag{3}$$

$$\begin{cases} g_x(x, y) = I(x, y + 1) - I(x, y - 1) \\ g_y(x, y) = I(x - 1, y) - I(x + 1, y) \end{cases}, \tag{4}$$

$$E(M, N) = \left[\sum_{m=1}^{M} \sum_{n=1}^{N} t(i + m, j + n) \cdot f(i, j) \right], \tag{5}$$

where $T_t(x, y)$ represents the qualified solder joint features obtained from the template matching algorithm at coordinates (x,y), $r_1, r_2,$ and r_3 represent weighted average parameters, with $r_1 + r_2 + r_3 = 1$. x,y are the horizontal and vertical coordinates of a point, (a,b) represents the circle's center with radius r, which is the ideal radius value of the solder joint. $G(x,y)$ represents the gradient strength at point

(x,y), g_x and g_y represent the horizontal and the vertical gradient intensity, respectively, I represents the pixel value of the image, $E(M,N)$ represents the output values for an image of the length M and width N, $t(i + m, j + n)$ represents the pixel value relative to the upper-right corner (i,j) of the image with length M and width N. $f(i,j)$ denotes the selected image processing filter.

It is recognized that the characteristics such as shape, grayscale value, and occlusion of solder joints at different positions on the vehicle floor may vary, and the effectiveness of different welding guns can also differ. To address this problem, multi-template matching detection for the same solder joints was implemented using multiple templates. During the image enhancement process, the template image may undergo changes in scale, leading to challenges in detection accuracy with traditional template matching, especially regarding scale and rotation transformations. To overcome this, the multi-scale template detection was achieved using the image pyramid method. Fig. 2 shows the flowchart of the multi-scale and multi-template matching algorithm.
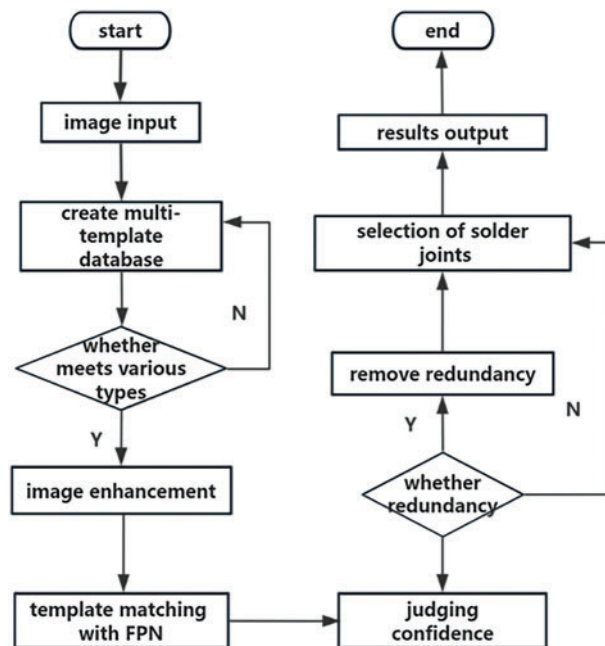

Figure 2: Flowchart of multi-scale and multi-template matching

The Laplace pyramid is employed to transform the image scale, and a hierarchical search strategy is used to enhance the information available in the detection target. Experiments have demonstrated that this method effectively matches solder joints of varying sizes [21]. The multi-template matching algorithm allows for selecting various features from multiple templates, which are then extracted, concatenated, and fused to form a single template vector [22]. Combining coarse and fine matching algorithms allows the method to quickly identify the approximate region of matching points, significantly reducing the overall matching frequency. For multiple template detection, selecting the appropriate confidence level for each template is necessary, and then removing the redundancy. These rates are then used as thresholds to determine and filter matching results [23].

Template matching was used to detect qualified solder joints on the vehicle floor by adjusting the matching threshold, utilizing the standard correlation matching method template matching

Correlation Coefficient normed (TM-CCOEFF-NORMED) with a similarity of

$$R\left(x, y\right) = \frac{\sum_{x',y'}\left(T\left(x', y'\right) \cdot I(x + x', y + y')\right)}{\sqrt{\sum_{x',y'}T'(x', y')^2 \cdot \sum_{x',y'}I'(x + x', y + y')^2}}, \tag{6}$$

where $T\left(x', y'\right)$ represents the feature value (pixel value) of the template image at (x', y'). $I(x + x', y + y')$ represents the feature value (pixel value) of the image to be detected at $(x + x', y + y')$. (x', y') denotes the coordinate of the upper-left corner of the template, (x,y) denotes the coordinate of the upper-left corner of the image to be detected, $(x + x', y + y')$ represents the position of the template within the image to be detected.

3.2 Optimizing Based on Faster RCNN

Faster R-CNN offers powerful feature extraction capabilities, enabling the extraction of deep image features through its network [5,18]. The positions are manually judged and labeled based on the characteristics of qualified, unqualified, and unwelded solder joints. A solder joint database is then constructed, and the samples are trained using the Faster RCNN network to obtain various solder joint features. The training and testing of the Faster RCNN network were conducted to extract these features. The feature value of the sliding window centered on the image coordinates(x, y), as output by the Faster RCNN network, is denoted as $T_d(x, y)$.

Compared with the original Faster RCNN algorithm, the optimized version replaces Visual Geometry Group16 (VGG16) with the Residual Network50 (ResNet50) [24] and feature pyramid network (FPN), which has more vital feature extraction abilities. By incorporating multi-scale dilated convolution modules, the algorithm enhances the contextual features of the feature pyramid, expands small target information, and facilitates the fusion of high-level and low-level image information, thereby improving detection accuracy. ResNet50 is the backbone network for feature extraction in this optimized algorithm, providing deep feature representations. The FPN structure extracts multi-scale features from the deep-learning feature maps generated by ResNet50 [24,25]. These features are then fused with the original image feature information. These fused feature maps are subsequently used for object detection, and candidate target regions are quickly generated based on the multi-scale feature maps. Subsequently, the region of interest (RoI) Align layer is used to unify the sizes of different regions, followed by processing with a classifier and bounding box regressor. Finally, the algorithm outputs the positions and categories of the detected solder joints. RoI Align is specifically used to optimize the region mismatch problem caused by double quantizations during the RoI Pooling.

The loss function serves as the guiding mechanism for adjusting the weights in the classifier, thereby improving the performance of the detector model. The loss function of Faster RCNN includes classification and regression loss [5,26], and it's defined as

$$L\left(\{p_i\}, \{t_i\}\right) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*), \tag{7}$$

where i represents the number of indices for the anchor point. p_i represents the probability that anchor i is the target result. p_i^* represents the true value of p_i, when detected as the target, the true value is 1, on the contrary, it is 0. $t_i = \left(t_x, t_y, t_w, t_h\right)$ represents the positional parameter of the candidate box. t_i^* represents the positional parameter of the candidate box. L_{cls} represents the logarithmic loss function for whether it is the target. L_{reg} represents the Intersection over Union (IoU) regression loss function. λ represents the balanced weight that balances L_{reg} and L_{cls}. N_{cls} and N_{reg} represent normalization

parameters. In detection, utilize L_{reg} and L_{cls} complementarity to optimize the detector. Where

$$L_{cls}(p_i, p_i^*) = -log\left[p_i p_i^* + (1 - p_i)(1 - p_i^*)\right], \tag{8}$$

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*), \tag{9}$$

$$R = smooth_{L1}(x) = \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases}, \tag{10}$$

3.3 Industrial Fusion Cascade Detection

Combining the original image features (obtained through template matching) with deep learning features (extracted from Faster R-CNN) enhances the model's expressive ability. The fusion of features improves the accuracy of detecting small and deformed targets. The final fused features are obtained by assigning different weights to various features and performing a weighted summation in a specific proportion [27,28]. The structure of the feature fusion process is shown in Fig. 3.
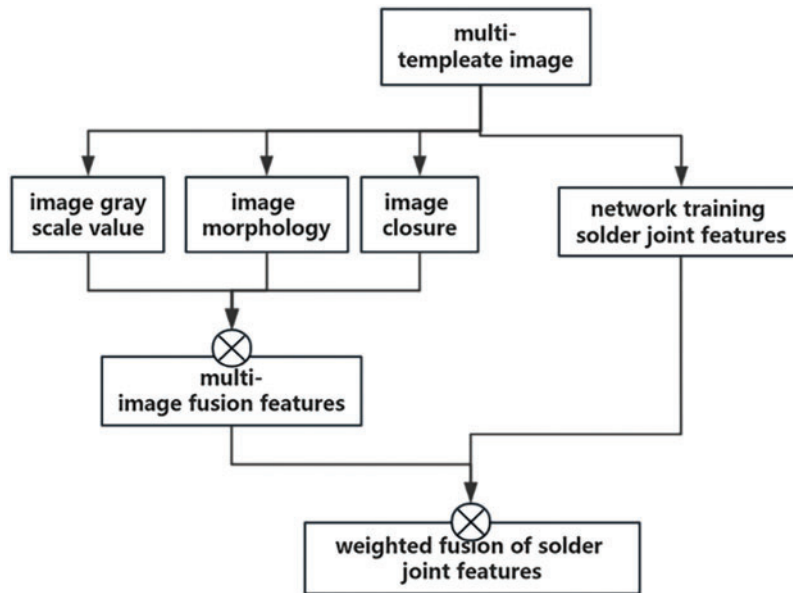

Figure 3: Principle diagram of weighted features fusion

The fused feature values of the image at the coordinate points (x, y) is

$$T(x, y) = \varphi T_t(x, y) + (1 - \varphi) T_d(x, y), \tag{11}$$

where $T(x, y)$ represents the qualified solder joint features on (x, y) after fusion. φ represents the one-dimensional parameter of the template matching algorithm during fusion and $\varphi \in [0, 1]$. The process is shown in the feature fusion module in Fig. 1. After adaptive weighted feature processing fusion, the image and deep learning feature values are fused into one feature value.

The total mean square error is

$$\sigma^2 = \varphi^2 \sigma_t^2 + (1 - \varphi)^2 \sigma_d^2, \tag{12}$$

where σ^2 is the multivariate quadratic function of each weight factor.

The FPN [23,29] was employed to refine and modify the network structure for object detection. The process included rejecting the original solder joint features and the fusion of the feature maps obtained from deep learning through horizontal connections and upsampling operations. The result was a more diverse feature pyramid structure encompassing feature information at varying scales. This structure proved highly effective for multi-scale object detection [30,31]. The feature fusion of the two algorithms improves the detection accuracy of specific targets. It enhances the system's robustness, as shown in the fusion feature module in Fig. 1. Additionally, the structure of the fusion cascade detector, which employs different algorithms and confidence levels for detection, is shown in Fig. 1 within the intermediate detection module.

Firstly, the multi-scale and multi-template matching algorithm was used to detect qualified solder joints on vehicle floor images [32]. Their similarity scores and corresponding positions were retained for results that did not meet the reliability criteria of the template-matching configuration. For solder joint areas that were undetected template matching, the deep learning-based Faster RCNN w_f was used for further supplementary detection, identifying qualified, unqualified, and unwelded solder joints. The similarity scores and positions of qualified solder joints that did not satisfy the Faster RCNN criteria were also retained. To address cases where neither the template matching confidence nor the Faster RCNN confidence threshold was met, cross-fusion verification w_f' was performed on their similarity scores to supplement the detection results. The detector used different anchors and confidence levels, which helped avoid potential overfitting and underfitting issues. Fig. 4 shows the neural network architecture diagram. This cascaded object detection module combined the multi-scale and multi-template matching detector with an optimized Faster RCNN detector, which conducted two stages of detection at different confidence levels. Each detector received the results from the previous stage and used them as input for the next subsequent detection.

$$f_d = R(x, y) \circ w_f \circ w_f', \tag{13}$$

Detection redundancy can occur due to using different template-matching techniques and varying confidence levels in Faster RCNN. Non-Maximum Suppression(NMS) was employed to address this. The NMS algorithm eliminates redundancy by searching for the targets with local maximum values, achieving the optimal result [26,27]. After removing applying NMS, the actual coordinate information and the corresponding positions for each detected template were obtained.

$$W(i, j) = \sum_{k=i, l=j-1}^{i+1, j+1} f(k, l), \tag{14}$$

where $W(i,j)$ represents the weight corresponding to the sample template at point (i,j).

Assuming the comparison result $D(k)$ of template matching on (i,j). $W_k^T(i, j)$ represents the standard template on (i,j), and T represents the matching threshold. Then, the matching comparison result on (i,j) is

$$D(k) = \begin{cases} 0 & abs(W(i, j) - W_k^T(i, j)) \leq T \\ 1 & abs(W(i, j) - W_k^T(i, j)) > T \end{cases}, \tag{15}$$

In order to solve the possibility of missing detection results, the confidence scores of the two detection algorithms were weighted and fused to form the complementary detection confidence. This process enabled the supplementary detection of potentially qualified solder joints. By using self-learning adaptation methods to adjust the confidence levels of the optimized new object detection, a

fusion of the two algorithms was achieved, enhancing detection accuracy [33]. The new supplementary testing confidence level is

$$T_n = \mu T_t + (1 - \mu)T_d \quad T_n < T_t \, and \, T_n < T_d, \tag{16}$$

where T_n represents the supplementary detection confidence, T_t represents the average confidence level of each template in the multi-scale and multi-template template matching, and T_d represents the confidence level of qualified solder joints in the optimized Faster RCNN. The weighted confidence parameter is μ, and $0 < \mu < 1$.
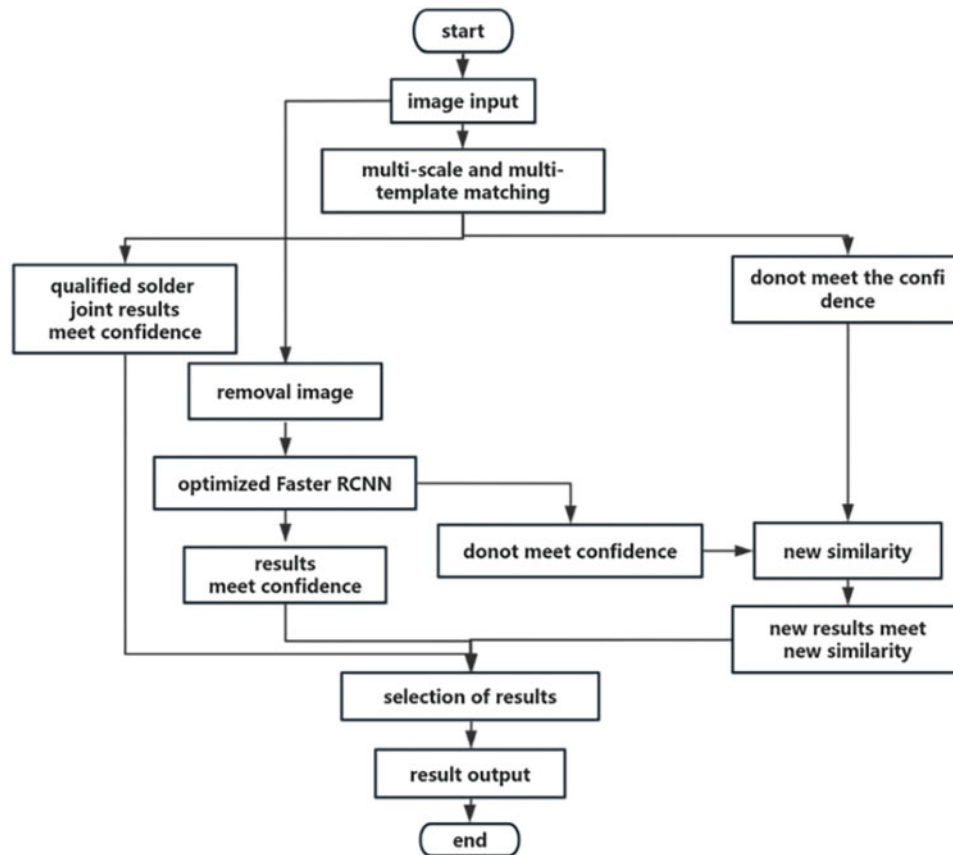
Figure 4: Flowchart of cascade object detection

4 Experiments and Analysis

4.1 Dataset Construction and Experimental Environment

In order to verify the performance of the algorithm proposed in this paper, a vehicle solder joint dataset was developed. The performance of several algorithms mentioned in this paper was compared and analyzed. Photos of the vehicle floor on the workpiece assembly line were captured using cameras, transferred to a computer via an image acquisition card, and processed using Python in combination with the OpenCV image processing library. Visual Studio software was utilized to achieve solder joint detection. The dataset consists of professional labels provided by automotive manufacturing experts, with fewer than one thousand annotations. The dataset includes manually judged and annotated

positions of qualified solder joints and various types of unqualified and unwelded solder joints. Data augmentation techniques, such as rotation, flipping, scaling, blurring, stitching, and others, were applied to enhance the robustness of the training model and increase the number of images. The processed images created a comprehensive image set of vehicle floor solder joints. For specific types of false positives, improvement was achieved by adding negative samples to the training dataset, resulting in a dataset of 1000 images. The image dataset was divided into training and testing sets with 8:2.

4.2 Experimental Results

The images were preprocessed for the twenty sets of vehicle floor photos taken. This paper then employed single-template matching, multi-scale and multi-template matching, Faster RCNN detection, YOLOv9, and an optimized cascading object detection algorithm to detect solder joints. The average results from each method were analyzed and compared. Figs. 5–9 show these detection results with fused features on three types of solder joints from images of the same vehicle floor. The red box represents the qualified solder joint detection result, the blue box represents the unqualified solder joint detection result, and the purple box represents the unwelded solder joint detection results.
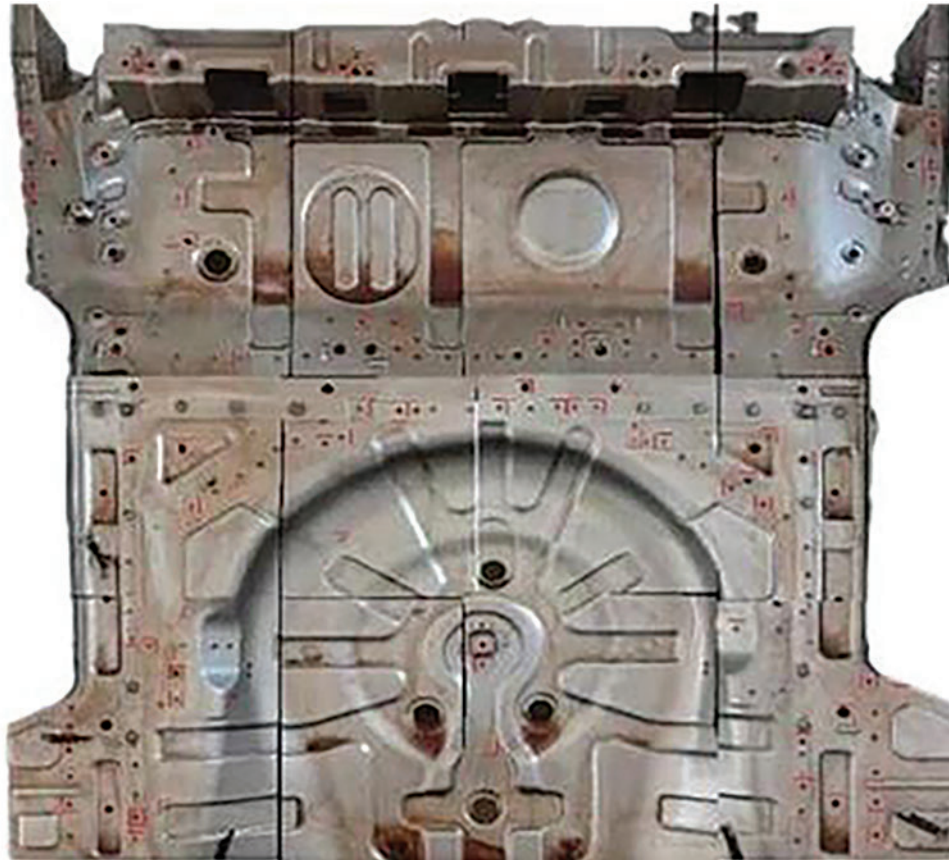

Figure 5: Result of single-template matching detection

Figure 6: Result of multi-scale and multi-template matching detection

Fig. 5 shows the detection results of qualified solder joints using original single-template matching algorithms. While many qualified solder joints are detected, there are cases of repeated, false, and missed detections.

Fig. 6 shows the detection results of qualified solder joints using the multi-scale and multi-template matching algorithms. Compared to the single template matching algorithms, it detects more qualified solder joints and eliminates many duplicate detections. However, it has poor detection performance for other types of solder joints, with instances of both missed and false detections.

Fig. 7 shows the detection results for qualified, unqualified, and unwelded solder joints using the optimized Faster RCNN algorithm. While deep learning algorithms can successfully detect unqualified and unwelded solder joints, the small sample size leads to unsatisfactory detection results for qualified solder joints, with instances of redundant detection still present.
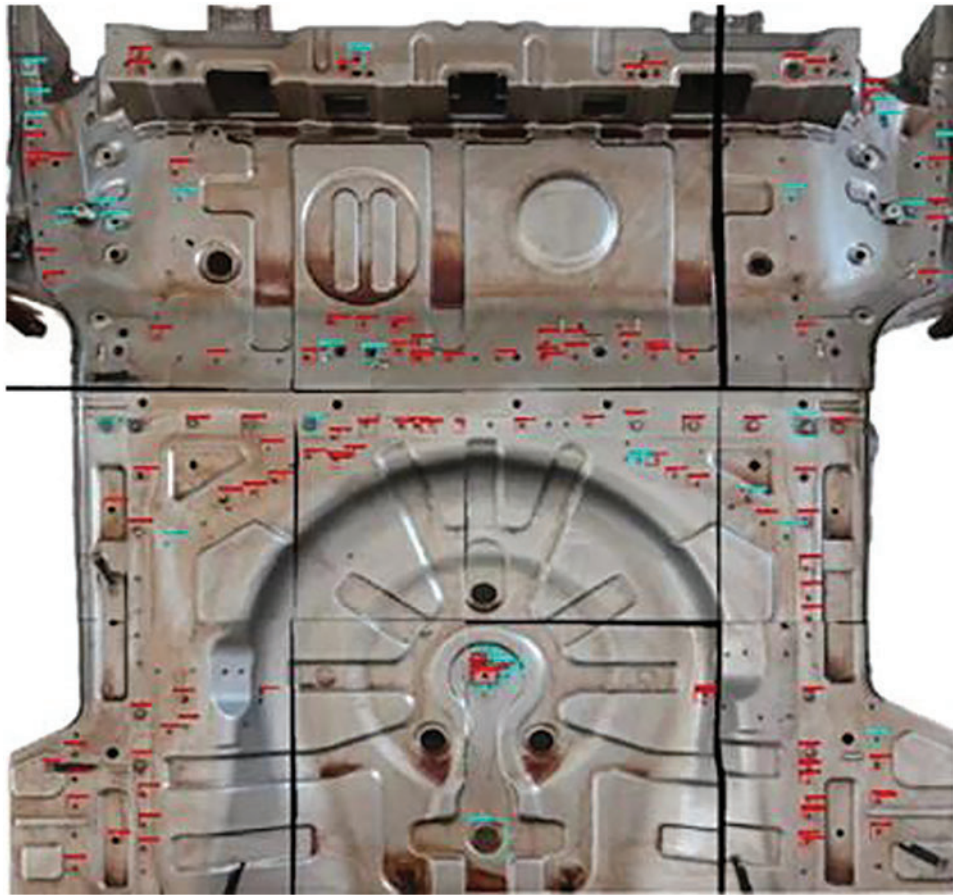
Figure 7: Result of optimized Faster RCNN detection

Fig. 8 shows the detection results for qualified, unqualified, and unwelded solder joints using the YOLOv9 algorithm [8]. Compared to Faster RCNN, YOLOv9 demonstrates better accuracy and precision in detecting qualified solder joints. However, the detection results for other types of solder joints need to be improved. Due to the small dataset, the detection accuracy threshold must be lowered to achieve the results above. However, experiments showed that this adjustment led to an increase in false detections.

Fig. 9 shows the detection results for qualified, unqualified, and unwelded solder joints using the cascade object detection algorithm. This method can detect various solder joints with high accuracy and low false detection rates. However, there are still instances of missed detections and false positives.

4.3 Experimental Analysis

In practical industrial scenarios where accuracy and precision were prioritized over efficiency, the performance of various algorithms was compared and analyzed. Table 1 shows the average accuracy, precision, recall, and F1-measure (F1) of twenty sets of flooring results. Each row represents different algorithm results, and each column represents algorithm performance categories. For the optimized cascaded object detection algorithm, the average precision, accuracy, and recall all exceed 91%. Compared to several other algorithms, the precision improved by 16.7%, 5.8%, 23.8%, and 17.7%,

respectively. The accuracy increased by 21.3%, 3.4%, 19.7%, and 17.3%, respectively. The recall rate improved by 21.5%, 6.8%, 20.5%, and 21.2%, respectively.
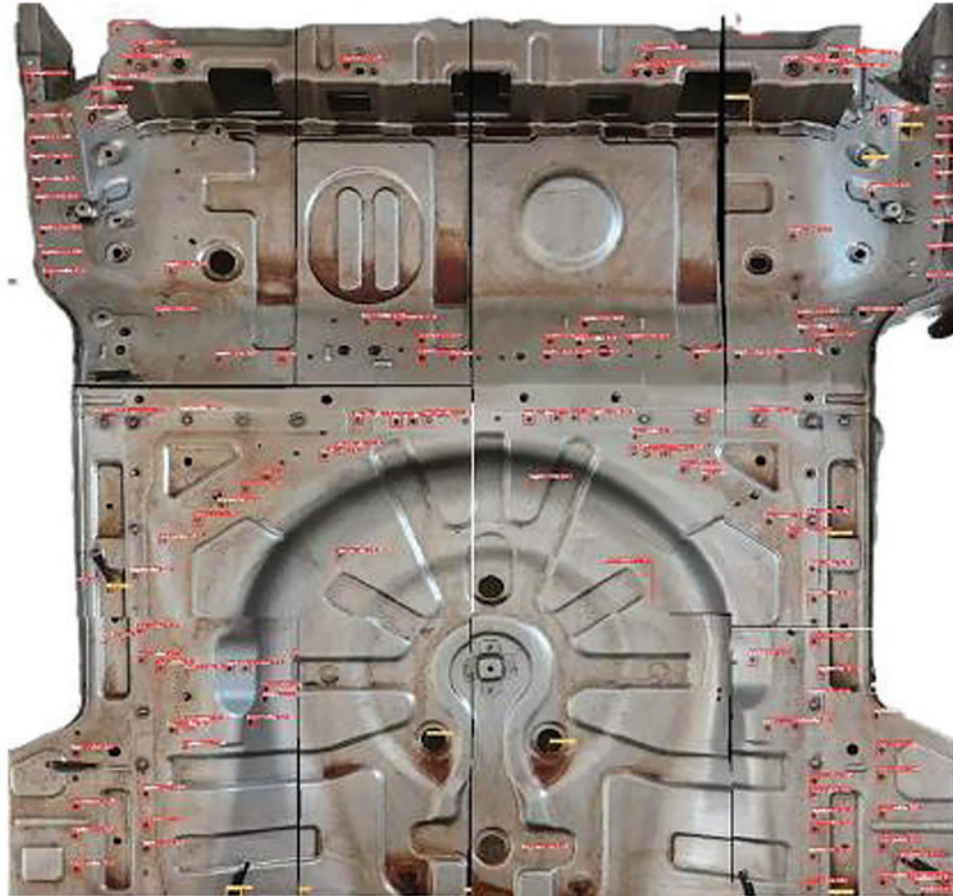

Figure 8: Result of YOLOv9 detection

Table 2 shows the actual testing results of a particular floor. Each row represents different algorithm results, and each column represents the detection results. Although the cascaded object detection algorithm takes longer, it offers higher detection accuracy. For industrial detection scenarios where efficiency is not the primary concern, the proposed algorithm proves more effective and has more excellent practical value. By segmenting the entire vehicle floor image, the detection results of the segmented image are optimized compared to non-segmented images.

When comparing the results of several algorithms, it was found that due to the small sample size of the database, the Faster RCNN and the YOLOv9 performed poorly in detection with high rates of false positives and re-detections. Qualified solder joints often needed to be more qualified. However, these algorithms detected some actual unqualified and unwelded solder joints. The original single template matching algorithm was better detected at qualified solder joints. It effectively eliminated redundant re-detections but produced many false positives and had poor detection accuracy for unqualified and unwelded solder joints. The multi-scale and multi-template matching algorithm was more effective at removing re-detection and false positives. It improved the detection accuracy of qualified solder joints, but its detection effect on other types of solder joints could have been better.

Figure 9: Result of cascade object detection

Table 1: Comparison of average evaluation indicators of various algorithms

Algorithms	AP[1]	AA[1]	AR[1]	F1
TM[1]	74.6%	71.2%	72.6%	72.9%
MMM[1]	85.5%	89.1%	87.3%	87.3%
Faster RCNN[1]	67.5%	72.8%	73.6%	70.4%
YOLOv9	73.6%	75.2%	72.9%	73.2%
COD[1]	91.3%	92.5%	94.1%	92.7%

Note: [1]AP stands for average precision, AA stands for average accuracy, and AR stands for average recall. TM stands for template matching, MMM stands for multi-scale and multi-template matching, Faster RCNN stands for Faster Region-based Convolutional Neural Networks and COD stands for Cascade object detection.

Both template matching algorithms achieved higher precision, accuracy, and recall rate than the Faster RCNN algorithm and higher precision than YOLOv9. However, their effectiveness in detecting unqualified and unwelded solder joints could have been better than that of the Faster RCNN and YOLOv9. Their detection accuracy for qualified solder joints at edges, in depressions, or with deformations was also lower.

Table 2: Comparison of detection results of various algorithms

Algorithms	Total number	Number of qualified	Number of unqualified	Number of unwelded
TM[1]	118	88	0	0
MMM[1]	124	104	0	0
Faster RCNN[1]	110	74	2	1
YOLOv9	146	113	0	0
COD[1]	120	108	1	2

Note: [1]TM stands for template matching, MMM stands for multi-scale and multi-template matching, Faster RCNN stands for Faster Region-based Convolutional Neural Networks and COD stands for Cascade object detection.

Applying the cascaded object detection algorithm is convenient, easy to use, and highly precise. Fusing the features obtained from two algorithms can more accurately represent solder joint information. Therefore, it outperforms single-feature approaches in detecting non-compliant and unwelded solder joints. Moreover, the algorithm demonstrates strong detection performance for solder joints in unique positions and shapes on workpieces and for solder joints deformed by industrial manufacturing processes.

However, current detection algorithms still experience issues with false positives and missed detections. After analysis and comparison, these problems are mainly caused by floor deformation during manufacturing, light spots of shadows, and other environmental influences. Some solder joints are located close to or near the edges of the workpiece and contain less information, which can lead to specific false positives and missed results.

Moreover, despite optimizations to the detection network structure to maximize the detection efficiency, the neural network remains relatively simple, and the amount of collected solder joint samples is limited. In particular, the number of unqualified and unwelded solder joints is even smaller. The imbalance in sample distribution leads to overfitting during the learning process, resulting in inaccurate detection results. Additionally, the current training strategy leaves room for improvement in detection accuracy. Further exploration of the sample sets, detection networks, and image-processing techniques is expected to enhance detection accuracy and enable practical engineering applications.

5 Conclusion

A comprehensive exploration of a feature fusion algorithm that integrates template matching with Faster RCNN is undertaken. This innovative approach harnesses the fused features to detect solder joints on vehicle floors with remarkable precision. The cascaded object detection algorithm seamlessly blends multi-scale and multi-template matching with Faster RCNN, effectively augmenting the detection of residual and diverse solder joint types. By aggregating the confidence scores from both methodologies, this paper forges a new detection confidence metric, significantly enhancing the accuracy of identifying qualified solder joints.

Employing the image pyramid technique, this paper fuses both algorithms' salient features, enabling comprehensive detection of qualified solder joints across the entire area. Subsequently, Faster RCNN, armed with dual confidence scores, supplements the detection within the remaining regions. Experimental evaluations underscore the superiority of this algorithm, boasting higher accuracy and resilience against occlusion, deformation, illumination variations, and scale changes compared to

existing methods. It outperforms three other detection algorithms in accuracy and precision, satisfying the rigorous demands of natural scene workpiece solder joint detection and demonstrating both feasibility and portability.

While this algorithm exhibits promising results, its accuracy, precision, and recall are inevitably influenced by external factors such as the detection environment, image quality, and inherent algorithm limitations. Minor deviations in manufacturing processes and the intricacies of small targets within solder joints can adversely affect test outcomes. Furthermore, challenges posed by 3D workpiece image conversion, including information loss, distortion, and occlusion, compound the detection task. Standardizing the features of unqualified and unwelded solder joints remains intricate, limiting feature expression and ultimately impacting detection accuracy. The imbalance in database samples and the algorithm's limitations in identifying small-scale solder joints hinder its ability to discern them amidst other targets.

A holistic approach to image and data processing is imperative to overcome these hurdles. Deeper analysis and processing of captured images, coupled with comprehensive training and optimization of the detection system, will enhance its adaptability to dynamic industrial production environments and varying targets. Refining the neural network architecture, augmenting the sample dataset, optimizing the detection network, and adopting optimal model parameters and training strategies will propel the algorithm's accuracy, precision, and efficiency even further. This evolution will facilitate real-time target detection and result output, positioning this algorithm as a viable solution for industrial production inspection.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Chunyuan Li, Mingquan Shi; data collection: Shuangming Wang, Lei Liu; analysis and interpretation of results: Chunyuan Li, Peng Zhang; draft manuscript preparation: Chunyuan Li. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] H. Ma, X. Wang, J. Bian, and Y. Ji, "Review of 3D machine vision development and its industrial application," (in Chinese), *Micro/nano Electron. Intell. Manuf.*, vol. 4, no. 4, pp. 50–61, 2022. doi: 10.19816/j.cnki.10-1594/tn.2022.04.050.

[2] S. Han, J. Pool, J. Tran, and W. J. Dally, "Learning both weights and connections for efficient neural network," *Neural Evol. Comput.*, vol. 5, pp. 1135–1143, 2015. doi: 10.48550/arXiv.1506.02626.

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 2980–2988, 2017. doi: 10.1109/TPAMI.2018.2844175.

[4] R. Girshick, "Fast R-CNN," *Comput. Sci.*, pp. 1440–1448, 2015. doi: 10.1109/ICCV.2015.169.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017. doi: 10.1109/TPAMI.2016.2577031.

[6] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 6154–6162. doi: 10.1109/CVPR.2018.00644.

[7] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021. doi: 10.48550/arXiv.2107.08430.

[8] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024. doi: 10.48550/arXiv.2402.13616.

[9] Y. Zuo, J. Wang, and J. Song, "Application of YOLO object detection network in weld surface defect detection," in *2021 IEEE 11th Annual Int. Conf. CYBER Technol. Autom., Control, Intell. Syst. (CYBER)*, Jiaxing, China, 2021, pp. 704–710. doi: 10.1109/CYBER53097.2021.9588269.

[10] A. Cardellicchio *et al.*, "Automatic quality control of aluminium parts welds based on 3D data and artificial intelligence," *J. Intell. Manuf.*, vol. 35, no. 4, pp. 1629–1648, 2024. doi: 10.1007/s10845-023-02124-1.

[11] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, "Relation networks for objectdetection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3588–3597. doi: 10.48550/arXiv.1711.11575.

[12] A. Vaswani *et al.*, "Attention is all you need," 2017. doi: 10.48550/arXiv.1706.03762.

[13] C. Ge, Y. Song, C. Ma, Y. Qi, and P. Luo, "Rethinking attentive object detection via neural attention learning," *IEEE Trans. Image Process.*, vol. 33, pp. 1726–1739, 2024. doi: 10.1109/TIP.2023.3251693.

[14] X. Han, J. Duan, and S. Dong, "Research on automotive weld inspection systembased on machine vision," *J. Changchun Univ. Sci. Technol. (Nat. Sci. Ed.)*, vol. 41, no. 5, pp. 75–79, 2018. doi: 10.3969/j.issn.1672-9870.2018.05.018.

[15] A. I. Borovkov, V. E. Prokhorovich, V. A. Bychenok, I. V. Berkutov, and I. E. Alifanova, "Ultrasonic inspection technique for welded joints obtained by spot friction stir welding," *Russ. J. Nondestruct. Test.*, vol. 59, no. 2, pp. 149–160, 2023. doi: 10.1134/S1061830923700262.

[16] C. Lin, X. Lv, Y. Cao, and J. Zhang, "Research on detection of automobile components solder joint quality,computer measurement & control," *Comput. Meas. Control,*, vol. 22, no. 1, pp. 31–33, 2014.

[17] N. Xie, J. Tan, L. Chen, and Y. Huang, "Welding spot positioning method for body-in-white based on lightweight YOLOv7," *Veh. Power Technol.*, vol. 4, pp. 33–38, 2023. doi: 10.19678/j.issn.1000-3428.0056446.

[18] X. Chen and A. Gupta, "An implementation of faster RCNN with study for region sampling," 2017. doi: 10.48550/arXiv.1702.02138.

[19] R. Liu, G. Yan, H. He, and Y. An, "Small targets detection for transmission tower based on SRGAN and faster RCNN," *Recent Adv. Electr. Electron. Eng.*, vol. 8, pp. 812–825, 2021. doi: 10.2174/2352096514666211026143543.

[20] B. Narani and A. Shenbagavalli, "Image registration by feature based information fusion," *Found. Comput. Sci. (FCS)*, vol. 2, pp. 12–15, 2012.

[21] Z. Peng, Q. Zhang, Y. Wei, and Q. Zhang, "Image matching based on multi-features fusion," *High Power Laser Part. Beams*, vol. 16, no. 3, pp. 281–285, 2004.

[22] L. Yu and R. Wang, "Object detection and recognition based on multiscale deformable template," *J. Comput. Res. Dev.*, vol. 39, no. 10, pp. 1325–1330, 2002. doi: 10.1007/978-3-540-24671-8_6.

[23] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *Comput. Vis. Pattern Recognit.*, pp. 1–10, 2017. doi: 10.1109/CVPR.2017.106.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE*, vol. 1, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[25] L. Beyer, X. Zhai, A. Royer, L. Markeeva, R. Anil and A. Kolesnikov, "Knowledge distillation: A good teacher is patient and consistent," *IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 10915–10924, 2021. doi: 10.1109/CVPR52688.2022.01065.

[26] Z. Wang, R. Song, P. Duan, and X. Li, "EFNet: Enhancement-fusion network for semantic segmentation," *Pattern Recognit.: J. Pattern Recognit. Soc.*, vol. 118, no. 1, pp. 108–134, 2021. doi: 10.1016/j.patcog.2021.108023.

[27] J. Yang, J. Y. Yang, D. Zhang, and J. Lu, "Feature fusion: Parallel strategy *vs.* serial strategy," *Pattern Recognit.*, vol. 36, no. 6, pp. 1369–1381, 2003. doi: 10.1016/S0031-3203(02)00262-5.

[28] H. Yang, Y. Fang, L. Liu, H. Ju, and K. Kang, "Improved YOLOv5 based on feature fusion and attention mechanism and its application in continuous casting slab detection," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–16, 2023. doi: 10.1109/TIM.2023.3284021.

[29] W. Xue, M. He, and Y. Zhang, "A novel decoupled feature pyramid networks for multi-target ship detection," *Sensors*, vol. 23, no. 16, pp. 31–49, 2023. doi: 10.3390/s23167027.

[30] H. Zheng, C. Pang, and R. Lan, "Cross-layer feature attention module for multi-scale object detection," in *Int. Symp. Artif. Intell. Robot.*, Singapore, Springer, 2022. doi: 10.1007/978-981-19-7943-9_17.

[31] R. Tian, H. Shi, B. Guo, and L. Zhu, "Multi-scale object detection for high-speed railway clearance intrusion," *Appl. Intell.: Int. J. Artif. Intell., Neural Netw., Complex Problem-Solving Technol.*, vol. 52, no. 4, pp. 3511–3526, 2022. doi: 10.3390/s23167027.

[32] C. Dong, K. Zhang, and Z. Xie, "An improved cascade RCNN detection method for key components and defects of transmission lines," *IET Gener., Transm. Dis.*, vol. 17, no. 19, pp. 4277–4292, 2023. doi: 10.1049/gtd2.12948.

[33] J. Li, K. Li, and X. Niu, "Self-adaptive perception model for action segment detection," in *Intelligent Computing*, Saga, Japan, 2022, vol. 283, pp. 834–845. doi: 10.1007/978-3-030-80119-9.