

# Application of AI intelligent vision detection technology using deep learning algorithm

**Yan Huang**

College of General Education, Chongqing Chemical Industry Vocational College, Chongqing, China

**E-mail:** [hyly606@163.com](mailto:hyly606@163.com)

Received 17 July 2023; accepted 8 November 2023; published online 5 December 2023

DOI <https://doi.org/10.21595/jme.2023.233506>



Copyright © 2023 Yan Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract.** This study aims to design efficient and reliable artificial intelligence vision detection models to improve detection efficiency and accuracy. The study filters defect-free images by image preprocessing and region of interest detection techniques. AlexNet network is enhanced by introducing attention mechanism modules, deep separable convolutions, and more to effectively boost the network's feature extraction capacity. An area convolutional neural network is developed to rapidly identify and locate defects on steel plate surfaces, utilizing an enhanced AlexNet network for feature extraction. Results demonstrated that the algorithm attained an average detection rate of 98 % and can identify defects in a minimal time of only 0.0011 seconds. For the detection of six types of steel plate defects, the average accuracy of the optimized fast regional convolutional neural network reached more than 0.9, especially for the detection of small-size defects with excellent performance. This improved AlexNet network has a great advantage in F1 value. The conclusion of the study shows that the designed artificial intelligence vision detection model has high detection accuracy, speed, and performance stability in steel plate surface defect detection and has a wide range of application prospects.

**Keywords:** deep learning, artificial intelligence, vision detection, steel plates.

## 1. Introduction

With the stable development of the world steel industry, the demand for surface defect detection of steel plates is increasing. The traditional manual visual sampling method cannot meet today's requirements of high efficiency and quality inspection, so a high-precision and high-efficiency inspection method is needed. Deep learning technology (DL) has become a new idea to realize the detection of steel plate surface defects (SPSD). This method with deep learning technology is not only highly accurate and automated but also very reliable. It enables steel companies to avoid the problem of large errors in manual inspection and further improve production efficiency and reduce production costs [1-3]. Deep learning-based SPSPD technology not only improves surface quality but also supports subsequent maintenance and upgrading of steel plate production and equipment. The data shows that it can achieve defects classification and location detection, and the location coordinates and area size of defects can be obtained in real-time and saved in a specified document [4-6]. Deep learning technology can not only provide a reliable basis for the production and maintenance of steel plates but also support technicians in upgrading steel plate production equipment, further improving production efficiency and increasing revenue generation. In addition, the methodological conclusion of the deep learning-based SPSPD technology can also provide a detection reference for other surface defective items. The detection of surface defects in items such as wood and glass is also a popular direction of research today [7-9]. Therefore, the research in this paper is not only relevant for the steel industry but also can provide references for surface defect detection in other industries. The research is divided into four parts, the first part explains the purpose of the research and introduces the research, the second part designs the research strategy and establishes the AI intelligent visual inspection model, the third part tests the model performance through performance tests and simulation experiments, and the fourth part draws the research conclusions.

The novelty of this study is mainly reflected in the following aspects: firstly, the researchers have designed an efficient and reliable AI visual detection model tailored to the specific application scenario of steel plate surface defects, in order to improve detection efficiency and accuracy; Secondly, the AlexNet network was improved and attention mechanism modules and deep separable convolutions were introduced to enhance the network's feature extraction ability; In addition, a steel plate surface defect localization and detection model based on Faster RCNN was adopted, and an improved AlexNet network was used as the feature extraction network; Finally, for the detection of six types of steel plate defects, the optimized Faster RCNN network has an AP value of above 0.9, especially for the detection of small size defects.

## 2. Related works

Yang et al. [10] analyzed the latest applications and research progress of intelligent vision perception in industrial applications. The article highlighted the challenges faced by intelligent vision perception and explored its future trends [10]. Dai et al. [11] proposed a method for automatic inspection of automotive body spot welds using machine vision technology. The method was improved by introducing a lightweight network and complete IoU loss, among three other improvements. The results showed that this method successfully implemented resistance spot weld visual inspection in quantitative results of the spot weld dataset and improved the accuracy and real-time performance of the inspection results. Wu and Li [12] explored the electrical connector defect detection problem and proposed an improved Yolo V3 algorithm. The algorithm was more accurate and faster than traditional methods, with an accuracy of 93.5 %. It was concluded that the algorithm can satisfy electrical connector testing in industry. Ismail and Malik [13] investigated an automated fruit grading inspection system using deep learning techniques and a stacked integration approach, with the underlying hardware using low-cost Raspberry Pi modules to provide a real-time visual inspection, freshness and appearance inspection solution with significant benefits, where the average accuracy of the apple and banana test sets reached 99.2 % and 98.6 %, with 96.7 % and 93.8 % accuracy in real-time testing, respectively. Thus, the system significantly outperformed existing methods. Han et al. [14] developed a structured light vision sensor by a structured light sensor with narrowband filters. The sensor used image pre-processing, etc. for weld identification and feature extraction to obtain weld dimensions and assess weld quality. The results showed that the technology welding industry provided effective technical support and a wide range of application prospects.

Deep learning algorithms are also being used in depth in various fields. Jacob and Darney [15] collected extensive data on IoT deployment categories, security, and privacy challenges, in addition to applying efficient recognition rates to IoT-based image recognition using deep learning techniques. Results showed that deep learning can improve image recognition for IoT systems. This study suggested appropriate criteria for incorporating deep learning methods into IoT research. Ranganathan [16] explored and analyzed deep learning-based applications in analyzing preprocessing steps in signal, image, and text classification. Thus, deep learning algorithms and preprocessing steps are of great value for the development of human society. Kaushal et al. [17] investigated the systematic bias present in deep learning applications, especially related to the geographical distribution of patient cohorts. A search using PubMed identified 74 studies that met the requirements. The results suggested that more attention should be paid to the geographic distribution of the patient cohort when training the algorithm to avoid any cognitive and technical bias introduced to the algorithm. Wang et al. [18] designed a diagnostic method using artificial intelligence technology that can be very helpful for disease prevention and control by analyzing CT images of patients with coronary pneumonia with results proving an accuracy of 79.3 % to 89.5 %.

The Dai W. team studied the application of deep learning in the visual inspection of resistance spot welding. Research was conducted on the widely used solder joints in the automotive industry, using machine vision for fully automated inspection to improve automotive performance. The

existing YOLOv3 model was improved by introducing the lightweight network MobileNetV3 and optimizing the feature pyramid network architecture to enhance small object detection capabilities. In addition, the CIoU loss function was also applied to improve the convergence speed and regression accuracy of the network, and novel data augmentation methods were utilized in the model training process. Experiments showed that the proposed method had good detection performance and can be applied to visual inspection in the automotive manufacturing industry [11]. The Pelletier M. G. team designed a machine vision-based detection and removal system for plastic contamination in cotton processing. The system utilized inexpensive color cameras to detect plastic during cotton processing and remove it. In order to reduce costs, manpower, and technical barriers, an automatic calibration algorithm that transforms detection equipment into a “plug and play” format was designed. The automatic calibration system can dynamically track the color of cotton and use frequency statistics to avoid the misidentification of plastic images. The new method can greatly save installation costs and facilitate the promotion and application in the cotton processing industry [19]. The Fang R. team proposed an improved model-independent meta-learning network-based automatic detection method for milling surface roughness. Compared to traditional visual detection methods, this method can better adapt to complex lighting conditions and different sample distributions in industrial environments. Using a small number of labeled samples, the proposed method can adaptively learn surface roughness features and perform fast and accurate recognition. This eliminated the need for retraining the detection model when learning new categories, reducing data requirements and computational costs. This method had good application prospects and improved the performance of existing machine vision detection in general industrial environments, providing new technical support for online detection of surface roughness [20].

The aim of this study is to enhance the efficiency and accuracy of visual detection of surface defects on steel plates by improving the AlexNet network through the introduction of attention mechanism modules and the use of deep separable convolutions. Similar to the research conducted by the Dai W. team, this study strives to improve industrial testing efficiency and accuracy. However, this study demonstrates higher efficiency, accuracy, and stability in detecting small-sized surface defects, as reflected in the average inspection rate, inspection time, and inspection results. This study shares similarities with the visual detection method proposed by the Dai W. team [11], as both emphasize the importance of detection efficiency and accuracy. Nonetheless, this research has significant advantages in F1 values and FPS, particularly in terms of better detection performance for small defects, which distinguishes it from the Dai W. team’s approach. Compared to the research methods of the Pelletier M. G. team, this study focuses more on improving detection efficiency and accuracy, while the Pelletier M. G. team mainly focuses on reducing plastic pollution during cotton processing through effective solutions. Although the Pelletier M. G. team’s automatic calibration system can prevent misidentification, this research method still has significant improvements in accuracy and speed for specific fields such as steel surface defect detection. This study is also consistent with the goal of the Fang R. team, which aims to enhance efficiency and accuracy in visual detection through the improvement of existing methods. However, this research breakthrough is in detecting surface defects on steel plates, while their research focuses on milling surface roughness. The research method presented in this study does not require a large number of annotated samples or retraining the detection model when learning new categories; thus, it significantly reduces data requirements and computational costs, demonstrating its superiority. Moreover, this method has improved the existing AlexNet network by introducing attention mechanism modules and deep separable convolutions, opening up a new path for improving the network’s feature extraction ability. These contributions show that this study has significant potential to improve the efficiency and accuracy of visual detection, making it relevant in a broad range of applications in the future.

Overall, this research method has shown significant advantages and innovation in the visual detection of surface defects on steel plates. This study demonstrates significant leadership and practical application value in improving detection accuracy, speed, and feature extraction

capabilities. This study has made significant contributions to improving visual inspection efficiency and accuracy, demonstrating high detection accuracy, speed, and performance stability for detecting surface defects on steel plates, and has broad application prospects.

### 3. AI vision inspection model design for steel surface defects

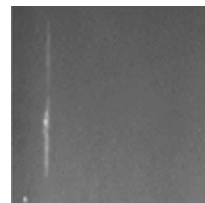
In SPSPD detection, image pre-processing is first performed, and the Region of Interest Detection (ROI) detection technique is used to filter defect-free images to improve the detection efficiency. Then, the AlexNet network is improved by introducing the attention mechanism module, depth-separable convolution, and batch normalization layer to improve the network feature extraction ability and generalization ability. Next, the Faster RCNN-based steel plate surface defect localization detection model is designed. The feature extraction network utilized is an enhanced AlexNet network, while the IOU K-means clustering algorithm optimizes Anchor generation in the RPN network. Moreover, to enhance its capability to detect small defects, ROI Align is employed to replace ROI Pooling.

#### 3.1. Defect image pre-processing

Before visual inspection of SPSPD can be performed, the visual image needs to be pre-processed first. The main technique used in the study is ROI detection. This technique aims to achieve ROI detection of steel plate surface images in order to filter out defective images from defect-free images and avoid unnecessary image processing. An example of the defect image is shown in Fig. 1.



a) Example of defect free images



b) Example of defect images

**Fig. 1.** Example of defect image

The algorithm needs to have a high detection rate as well as a fast detection speed considering the fast production of steel plates. An algorithm is proposed by the projection gray mean, which performs ROI detection by calculating the maximum and minimum values (MMV) of the horizontal and vertical projection gray mean and obtaining the difference. In addition, the algorithm sets a preset threshold to compare with the difference value to determine whether there is a defect in this image, and if there is, the next image preprocessing step is performed. In the absence of defects, the algorithm clears the image to free up the memory and re-inputs the image to improve the detection efficiency. The specific flow is shown in Fig. 2.

As can be seen, first, the MMV of the horizontal and vertical projection gray mean values of the image are obtained, and then the difference between them is calculated and compared with the preset threshold value. If the difference value is greater than the threshold, the image is defective and the next image pre-processing step is required. Conversely, if the difference is less than or equal to the threshold, the image is not defective. In this case, the image needs to be cleared to free the memory, and the image is re-imported for detection. The horizontal projection gray mean value is shown in Eq. (1):

$$f_x = N^{-1} \sum_{y=1}^N f_{xy}, \quad (1)$$

where,  $x$  is row number,  $y$  is column number, and  $N$  represents the total number of pixels in each row. The mean vertical projection grayscale value is shown in Eq. (2):

$$f_y = M^{-1} \sum_{x=1}^M f_{xy}, \quad (2)$$

where,  $M$  is the total number of pixels in each column. The difference between MMV of the horizontal projection grayscale mean is shown in Eq. (3):

$$f_{xsub} = f_{xmax} - f_{xmin}, \quad (3)$$

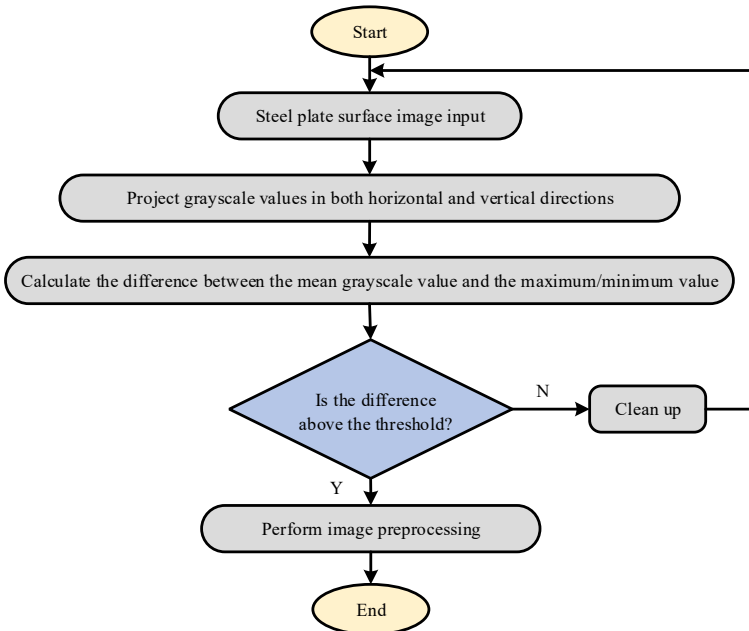
where,  $f_{xmax}$  and  $f_{xmin}$  means the MMV of the horizontal projection of the average gray value, respectively, and the difference between the MMV of the vertical projection of the average gray value is shown in Eq. (4):

$$f_{ysub} = f_{ymax} - f_{ymin}, \quad (4)$$

where,  $f_{ymax}$  and  $f_{ymin}$  represent the MMV of the mean vertical projection gray value, respectively. The conditions for determining the presence of defective images are shown in Eq. (5):

$$\begin{cases} f_{xsub} > T_1, \\ f_{ysub} > T_2, \end{cases} \quad (5)$$

where,  $T_1$  and  $T_2$  represent the detection thresholds for horizontal and vertical directions, respectively. In defect image processing, denoising is required to improve defect localization detection. For this problem, the study uses the Improved Adaptive Median Filtering Algorithm (IAMF) algorithm as the main aluminum filtering algorithm, which is able to remove noise well while retaining the edge and detail information of the defects, and its flow is shown below.



**Fig. 2.** ROI detection algorithm

The processing steps of the algorithm include boundary filling, grayscale judgment, filter window size adjustment, and value extraction output. When using the IAMF algorithm to filter the defective image, the original image needs to be boundary-filled first, and then the window size is gradually adjusted according to the principle of the algorithm and the grayscale values of the corresponding pixel points are output. If the final output grayscale value still cannot remove the noise, calculating the average grayscale value of all pixel points is needed within the maximum filtering window except the noise point and output it as the grayscale value of the current pixel. If all pixels within the maximum filter window are noisy points, the weighted average of the minimum and maximum grayscale values within the current filter window needs to be taken as the grayscale value output for the current pixel. Finally, the filtered image can be obtained by removing the filled boundary. the IAMF algorithm can better remove all kinds of noise, including pretzel noise and Gaussian noise, etc., thus making the defect image processing more reliable.

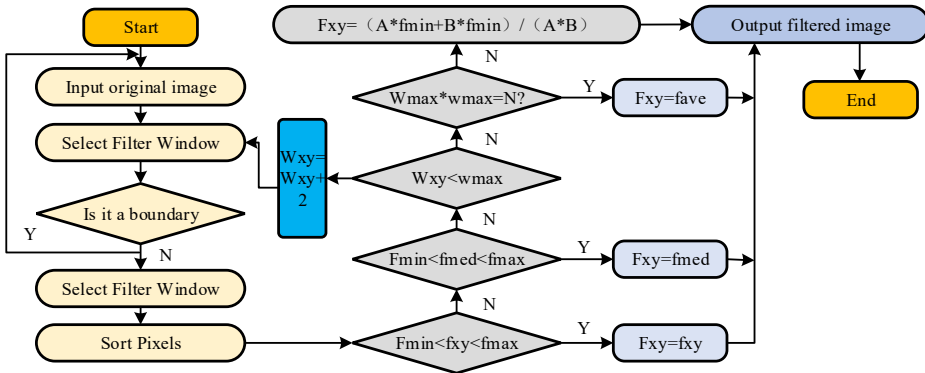


Fig. 3. IAMF algorithm

### 3.2. AlexNet network defect classification model design

The classification of SPSD is important in the steel production process. The performance of the traditional AlexNet structure in identifying steel surface defects is not satisfactory, so improvements are made to the traditional AlexNet network to improve steel surface defect classification. The main improvements include the attention mechanism module, depth-separable convolution, batch normalization layer, residual network structure, and the replacement of the activation function. First is the attention mechanism module. It is introduced to filter the important features in the feature map. The attention mechanism (Squeeze-and-Excitation (SE) module is derived from SENet. The design idea of the SE module is to learn the feature weights of each channel through the loss function, so that the feature map weights of task-related channels are significant, and thus improve the model classification performance. In the traditional AlexNet structure, the SE module is added before pooling layer 1, so that the input feature map can retain the important features better. The improved structure is shown below.

Second, the depth-separable convolution is introduced. The introduction of it can make the model have a better classification effect. Depthwise Convolution is divided into Depthwise Convolution (DW) and Pointwise Convolution (PW), and convolution kernel size of DW is  $D_k \times D_k$  and number of  $k$ . The convolution kernel size of PW convolution is  $1 \times 1 \times k$  and the number of  $N$ . Compared with the traditional convolution, the computation and the number of parameters of Depthwise Convolution are smaller. The depth-separable convolution is computationally smaller and the number of parameters is smaller, so it has higher computational efficiency. In the modified network, the  $3 \times 3$  convolutional layers in the original AlexNet structure are replaced with deeply separable convolutions. Assume that the input feature map size is shown in Eq. (6):

$$S_a = D_F \times D_F \times k. \quad (6)$$

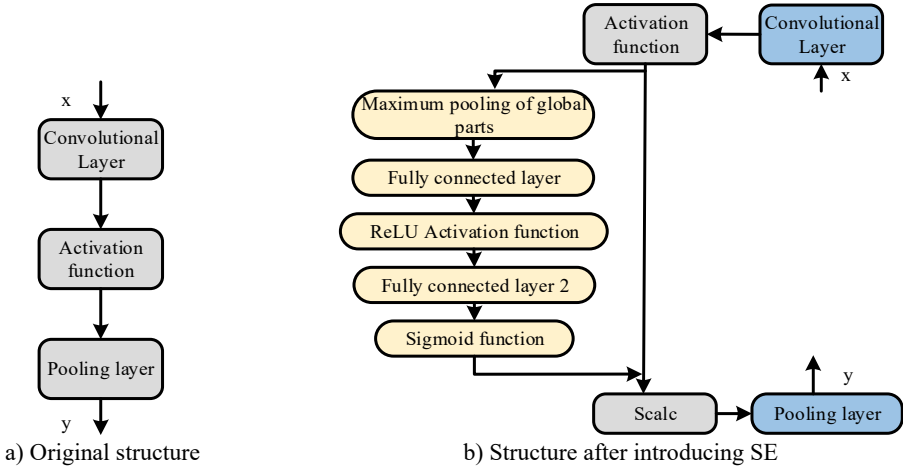


Fig. 4. Structure after introducing SE

The output feature map dimensions are shown in Eq. (7):

$$S_b = D_F \times D_F \times N. \quad (7)$$

Convolution kernel size is shown in Eq. (8):

$$S_c = D_k \times D_k \times k. \quad (8)$$

Then the total calculation is shown in Eq. (9):

$$Z_a = D_F \times D_F \times D_k \times D_k \times k \times N. \quad (9)$$

Then the total number of parameters is shown in Eq. (10):

$$Z_b = D_k \times D_k \times k \times N. \quad (10)$$

The total computation of the depth-separable convolution is shown in Eq. (11):

$$Z_b = D_F \times D_F \times D_k \times D_k \times k + D_F \times D_F \times k \times N. \quad (11)$$

Eq. (12) demonstrates the ratio of the total computation required for depth-separable convolution to standard convolution:

$$\frac{D_F \times D_F \times D_k \times D_k \times k + D_F \times D_F \times k \times N}{D_F \times D_F \times D_k \times D_k \times k \times N} = \frac{1}{N} + \frac{1}{D_k^2}. \quad (12)$$

Next, the batch normalization layer is introduced, which is typically positioned after the activation function to ensure that the activation function input is at the requisite position where the nonlinear function is more sensitive to the input and the network gets a larger gradient. The introduction of the batch normalization layer can reduce model overfitting. Furthermore, the residual network structure is presented, a deep learning architecture designed to effectively address the issues of gradient disappearance and explosion. This model directly transfers the previous layer's output to the subsequent layers and adds it with the data obtained through convolution of the residual structure to serve as the input of the lower layer network. Lastly, the activation function is replaced. ReLU activation in the original AlexNet structure causes the neuron to fail to update the parameters when the input is negative. Therefore, the ReLU function

is replaced by the Leaky ReLU function with leakage correction as the activation function, and the function output has a very small slope when the input is negative, which can ensure the response of the function input to negative information. In summary, by introducing improvements such as attention mechanism module, deep separable convolution, batch normalization, residual network structure, and activation function replacement, the improved AlexNet network can be better applied to SPSD classification. In particular, the introduction of an attention mechanism and deep separable convolution can help extract more key features and thus improve the classification performance. The batch normalization layer makes the data distribution more uniform, which helps to reduce the overfitting phenomenon. The residual network structure is achieved by introducing short-join connections to increase the depth and width of the model network while effectively avoiding gradient disappearance and gradient explosion problems. The replacement of the activation function avoids the degradation of learning ability due to the failure to update some neuron parameters. The improved AlexNet network structure package is shown in Fig. 5.

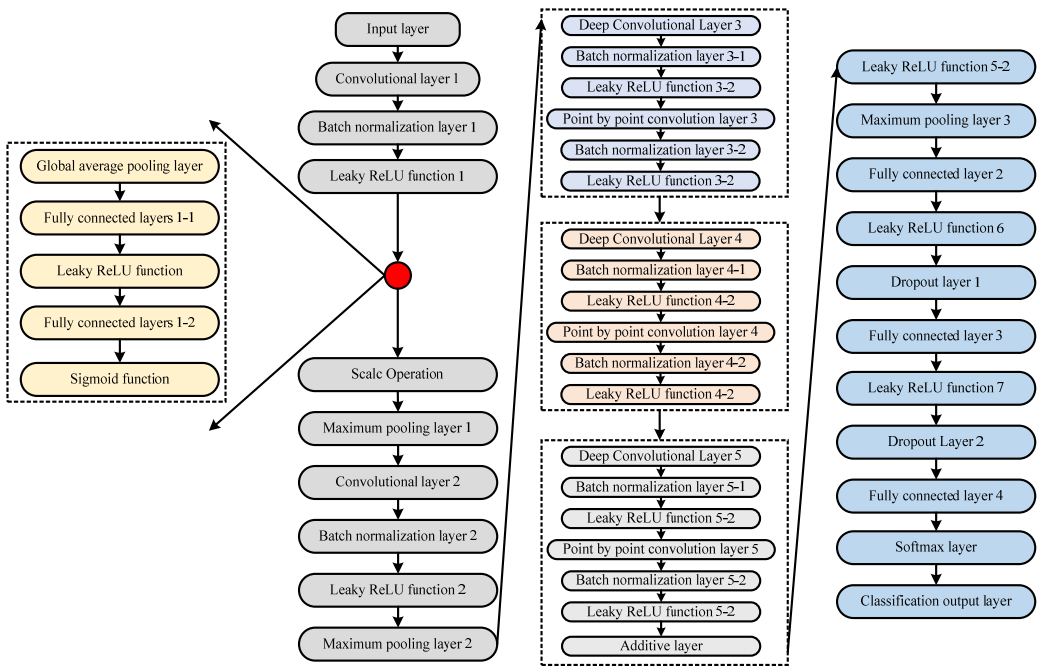


Fig. 5. Improving the AlexNet network structure

The improved AlexNet network structure includes the convolutional, pooling, and fully connected layers from the original AlexNet network, as well as the new SE module, deep separable convolution, batch normalization layer, residual network structure, and Leaky ReLU activation function. These improvements enable the network to achieve better performance in steel plate surface defect classification tasks with enhanced feature extraction and generalization capabilities while keeping the number of parameters and computational effort low.

When modifying the standard AlexNet for surface defect classification of steel plates, the study first considers the network's parameter selection. Network parameters can be adjusted and optimized through experiments to achieve optimal performance. Regarding the learning rate, it is an essential parameter in optimization algorithms. A smaller learning rate may slow down convergence speed, while a larger one may lead to unstable training processes. The adaptive learning rate optimization algorithm, such as Adam, can be chosen to be used, which enables the automatic adjustment of the learning rate. It is suggested that the initial learning rate be set to



0.001. In terms of batch size, the number of samples employed in each training iteration is referred to. Better utilization of hardware acceleration is typically achieved with larger batch sizes, but performance limitations may arise. Different batch sizes, such as 32, 64, 128, etc., can be experimented with. Concerning activation function selection, the study adopts the method provided by Leaky ReLU to improve ReLU and selects the activation function with the best performance through cross-validation.

In terms of network architecture, research has introduced an attention mechanism module - by introducing the Sequence-and-Excitation (SE) module, important features in the feature map can be filtered and the weight of task related channels can be increased, thereby improving the classification performance of the model. In the improved structure, the SE module is added to the position before pooling layer 1 to better preserve important features in the input feature map. Simultaneously introducing deep separable convolution - deep separable convolution can improve the model's feature extraction ability, thereby improving classification performance. In the improved network, replace the original AlexNet structure with  $3 \times$  Replace 3 convolutional layers with deep separable convolutions. And introducing batch normalization layer (BN layer) - batch normalization can reduce model overfitting and make the distribution of input data in each layer more uniform. The BN layer is usually used before activating the function. Studying the use of residual network structure - residual network structure can effectively solve the problems of gradient vanishing and gradient explosion. By introducing a short circuit connection, the output of the previous layer is directly transmitted to the subsequent layers, and added to the convolutional data obtained by passing through the residual structure as the input of the lower layer network. And replace the activation function - the ReLU activation function used in the original AlexNet structure may cause neurons to be unable to update parameters when inputting negative numbers. In the improvement method, Leaky ReLU is used to replace ReLU, ensuring that the function can respond to negative information when inputting negative values. In summary, this study has improved the AlexNet network by introducing attention mechanism modules, deep separable convolutions, batch normalization layers, residual network structure, and replacement of activation functions. This has enabled the improved AlexNet network to have stronger feature extraction and generalization capabilities while maintaining a low parameter and computational complexity, thus achieving better performance in steel plate surface defect classification tasks. In terms of network architecture, the research has introduced an attention mechanism module known as the Sequence-and-Excitation (SE) module. This module filters important features in the feature map and increases the weight of task-related channels, thereby enhancing the classification performance of the model. In the enhanced structure, the SE module is incorporated before pooling layer 1 to better preserve crucial features in the input feature map.

Simultaneously, deep separable convolution has been introduced to improve the model's feature extraction ability, thereby enhancing classification performance. In the improved network, the original AlexNet structure is replaced with a  $3 \times 3$  deep separable convolution for each of the three convolutional layers. Furthermore, a batch normalization layer (BN layer) has been introduced. Batch normalization reduces model overfitting and promotes a more uniform distribution of input data at each layer. The BN layer is typically utilized before activating the function. The study also explores the use of a residual network structure, which effectively addresses the issues of gradient vanishing and gradient explosion. By introducing short-circuit connections, the output of the previous layer is directly transmitted to subsequent layers and added to the convolutional data obtained through the residual structure as the input of the lower layer network. Additionally, the activation function has been replaced. The ReLU activation function used in the original AlexNet structure may hinder parameter updates when negative numbers are input. In the improvement method, Leaky ReLU is employed as a replacement to ensure that the function can appropriately respond to negative information. In summary, this study has enhanced the AlexNet network by introducing attention mechanism modules, deep separable convolutions, batch normalization layers, residual network structure, and replacement of activation functions. These enhancements equip the improved AlexNet network with stronger feature extraction and

generalization capabilities while maintaining low parameter and computational complexity. Consequently, the network achieves superior performance in steel plate surface defect classification tasks.

### 3.3. Improved RCNN defect location detection model design

The aim of this study is to solve SPSD using a Faster Region-based Convolutional Neural Network (Faster RCNN) method for locating and detecting defects on steel plate surfaces. In achieving this goal, the differences between different target detection algorithms are discussed and optimization work for Faster RCNN is carried out to improve the accuracy in locating and detecting SPSD. SPSD detection is a key process to ensure steel quality, improve product value, and reduce production costs are crucial objectives. Target detection algorithms are categorized into two main groups: faster YOLO and SSD algorithms and the more accurate Faster RCNN algorithm. Through the statistics of the steel plate surface defect data set, it is found that a large proportion of small size and slender type defects characterize the steel plate. Therefore, to improve defect detection capability, a Faster RCNN algorithm with higher accuracy is selected for optimization in this study, and the overall structure is shown in Fig. 6.

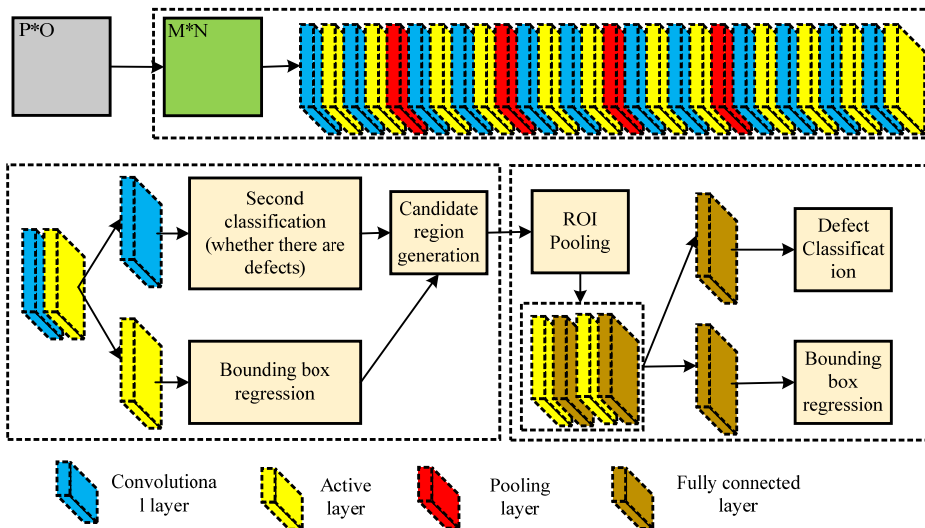


Fig. 6. Faster RCNN algorithm structure

First, the feature extraction part of Faster RCNN is improved for Faster RCNN. The traditional Faster RCNN extracts feature by VGG16 whose output feature map has a low resolution, which may decrease localization accuracy of small-sized defects. Therefore, this study adopts the improved AlexNet network to extract features, aiming to ensure the high resolution of the feature maps and improve the accuracy of global semantic information. Secondly, in order to overcome the effect of subjectivity generated by setting the Anchor manually, the IOU K-means clustering-based algorithm can optimize the Anchor generation method in the RPN network. By calculating the average intersection ratio between each real defective border and the k Anchors obtained by clustering, the appropriate k value is determined, and the number and size of Anchors are adjusted accordingly. The average cross-merge ratio can be expressed in the form of Eq. (13):

$$S = (Avg, IOU). \tag{13}$$

The average cross-merge ratio can be calculated using Eq. (14):

$$\max_{avg} = \frac{\sum_{i=1}^k \sum_{j=1}^{m_k} IOU(\theta, \psi)}{m}, \quad (14)$$

where,  $\theta$  denotes the real edges of defects under labeling, while  $\psi$  denotes the clustered edges,  $m$  denotes total number of defect training sets, and  $k$  denotes the amount of clustering. In addition,  $IOU$  can be calculated using Eq. (15):

$$IOU = \frac{area(C) \cap area(GT)}{area(C) \cup area(GT)}. \quad (15)$$

Finally, in order to avoid the errors brought by the ROI Pooling algorithm to the detection results and to improve small defects detection, ROI Align can replace ROI Pooling. It retains the floating point results and calculates the eigenvalues of the sampling points by bilinear interpolation, which can better maintain the accuracy of the defect feature size. The ROI Align algorithm implementation process is shown in Fig. 7.

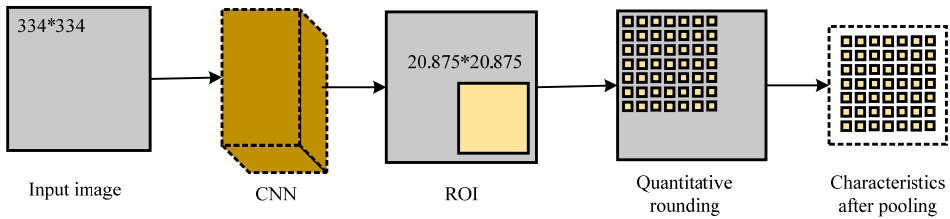


Fig. 7. ROI Align algorithm

After optimizing the Faster RCNN network, a more suitable feature extraction network and a more reasonable Anchor generation method for steel plate surface defect features, as well as a more accurate pooling algorithm, can be obtained. All these optimizations provide strong support for the detection task. During the training of the model, multiple iterations of training can be performed using methods such as cross-validation to evaluate the optimization effect of the Faster RCNN network and improve defect detection performance. This ensures that a more robust and robust network model is obtained.

#### 4. Visual inspection model effect analysis

In the visual inspection model effect analysis, the detection effect and defect classification performance of the ROI detection algorithm were tested and different algorithms' detection performance was compared. The performance of the traditional AlexNet network was also evaluated by the confusion matrix, and the performance of the traditional Faster RCNN network, optimized Faster RCNN network, SSD network, and YOLOv2 network was compared for steel plate surface defect detection.

##### 4.1. Analysis of the detection effect

The study first analyzed the detection effect of ROI detection for the model, and the detection rate and detection time were mainly used as the main test indexes in the analysis process, and the specific results are shown below.

First, from the perspective of detection rate, the data showed the difference between different defect types. From a low of 88 % to a high of 100 %, there was a more significant change in the detection rate of the algorithm in detecting different defects. Among them, the detection rate for defect type Fr was the lowest, with only 88 % of the defects detected correctly, but the detection rates for the rest of the defect types reached 100 %. This indicated that the algorithm's detection

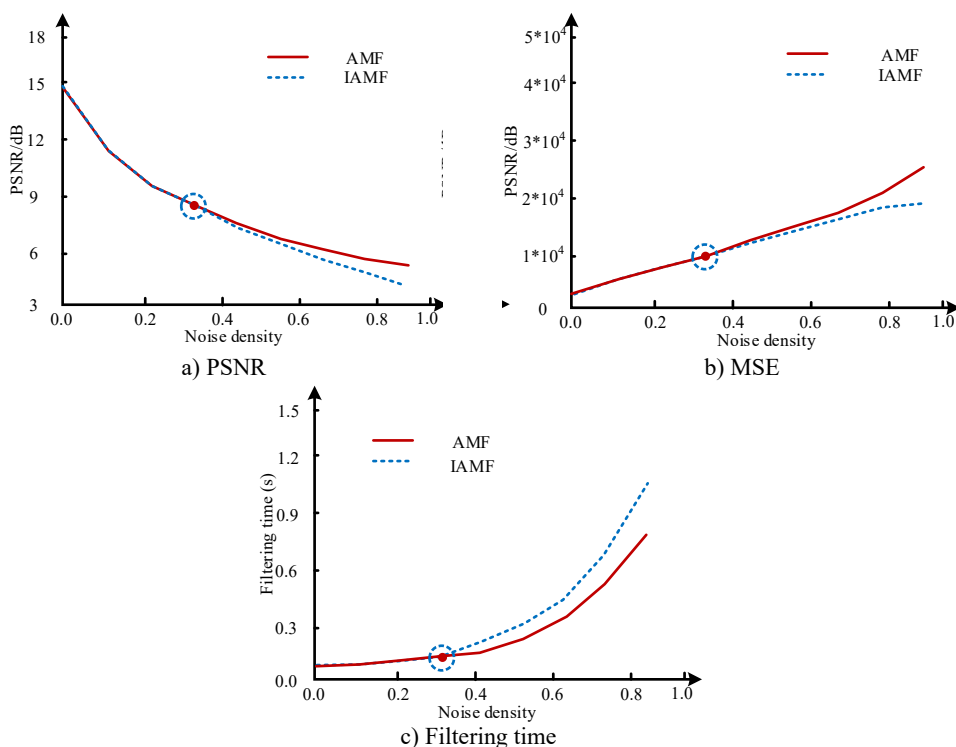
effect could vary with different defect types. However, overall, the average detection rate of the algorithm was 98 %, and the overall test result was more satisfactory. Second, from the perspective of detection time, it can be found that the detection time was relatively short, with the shortest detection time being only 0.0011 seconds. Although the detection time under the Sc defect type reached 0.0083 seconds, this was not too obvious from the detection time of other defect types. This indicated that the algorithm had a fast detection speed and efficient processing capability, which can be effectively applied to practical application scenarios.

**Table 1.** Detection effect analysis

| Defect type | Number | Detected quantity | Detection rate (%) | Detection time (s) |
|-------------|--------|-------------------|--------------------|--------------------|
| Fr          | 50     | 44                | 88                 | 0.0014             |
| Cr          | 50     | 50                | 100                | 0.0069             |
| In          | 50     | 48                | 96                 | 0.0011             |
| Pa          | 50     | 50                | 100                | 0.0055             |
| Ps          | 50     | 50                | 100                | 0.0017             |
| Rs          | 50     | 50                | 100                | 0.0035             |
| Sc          | 50     | 49                | 98                 | 0.0083             |

#### 4.2. Defect classification effect analysis

For the performance analysis of steel plate defect classification, different algorithms for the image filtering task were used to compare two different algorithms' performance. For classification, traditional AlexNet network performance was evaluated by confusion matrix of the test set. The performance of the traditional AlexNet network was measured again after improving the network. The results of the peak signal-to-noise ratio analysis are shown in Fig. 8.



**Fig. 8.** Peak signal-to-noise ratio analysis

As can be seen from Fig. 8, with a noise density of 0.1, the peak signal-to-noise ratio (SNR)

of the AMF algorithm was 14.448, while the peak SNR of the IAMF algorithm was 14.521. The difference between the two was not significant, but the IAMF algorithm was closer to the SNR level of the original image. In mean square error (MSE), the AMF algorithm's error value was 2335.581, while the error value of the IAMF algorithm was 1197.881. The error value of the IAMF algorithm was not only lower, but also about 1/2 of the error value of the AMF algorithm, which means that the IAMF algorithm has stronger noise suppression ability than the AMF algorithm. As the noise density increased, performance gap between two algorithms became more obvious. At a noise density of 0.7, the PSNR of the IAMF algorithm was 6.096, while the PSNR of the AMF algorithm was only 5.781. In MSE, error value of the IAMF algorithm was 15971.058, compared to the error value of the AMF algorithm of 17197.951. The lower error value indicates that the IAMF algorithm was more suitable for processing images with high noise density defects. The confusion matrix is shown in Fig. 9.

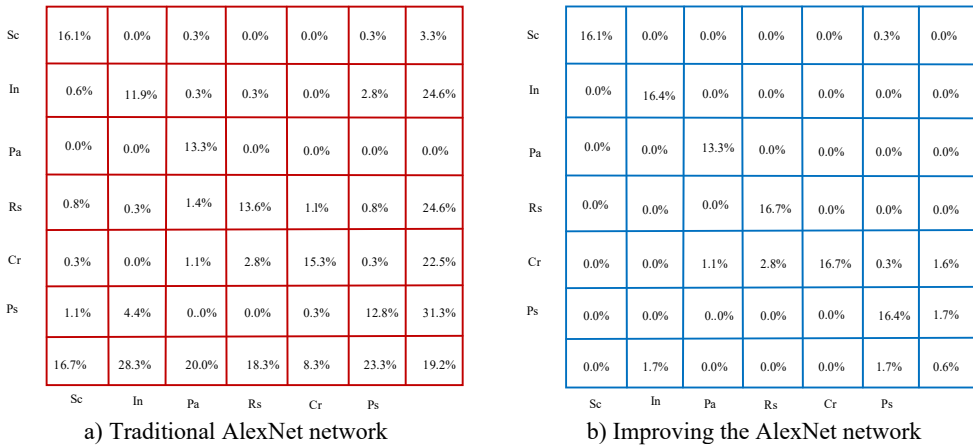


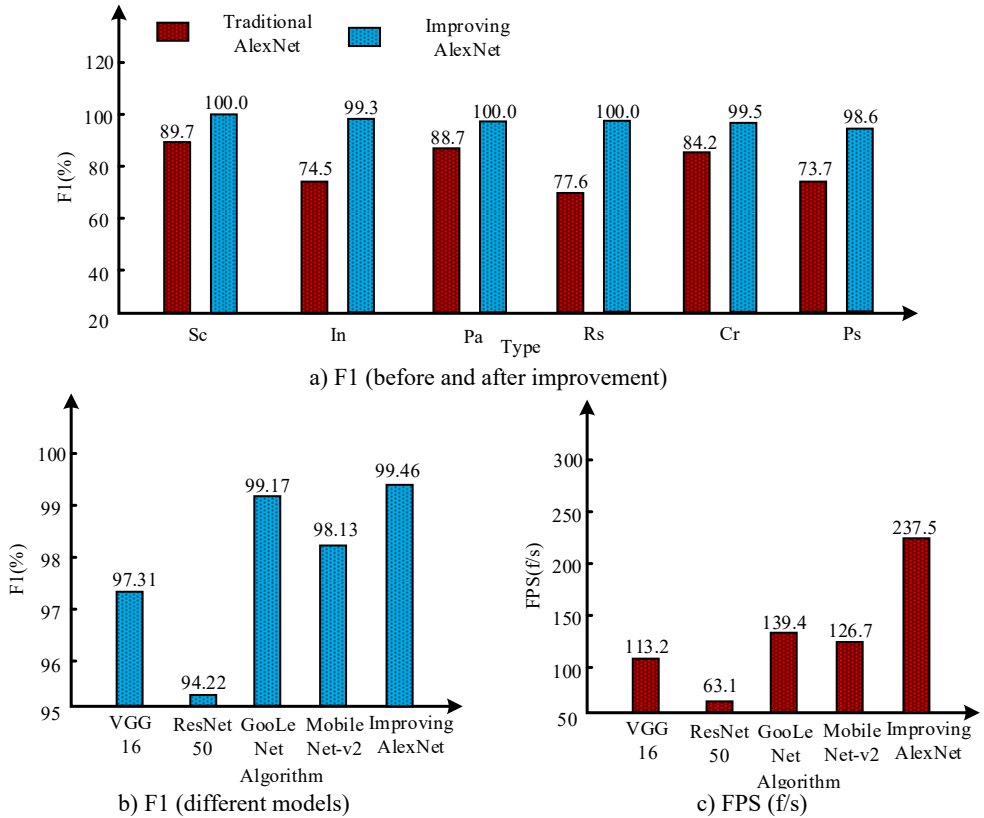
Fig. 9. Peak signal-to-noise ratio analysis

In Fig. 9, F1 values of the traditional AlexNet network for the classification of steel plate defects were relatively low, especially since the misclassification rate of pockmarks and inclusions was high. The specific data results obtained are shown in Fig. 9.

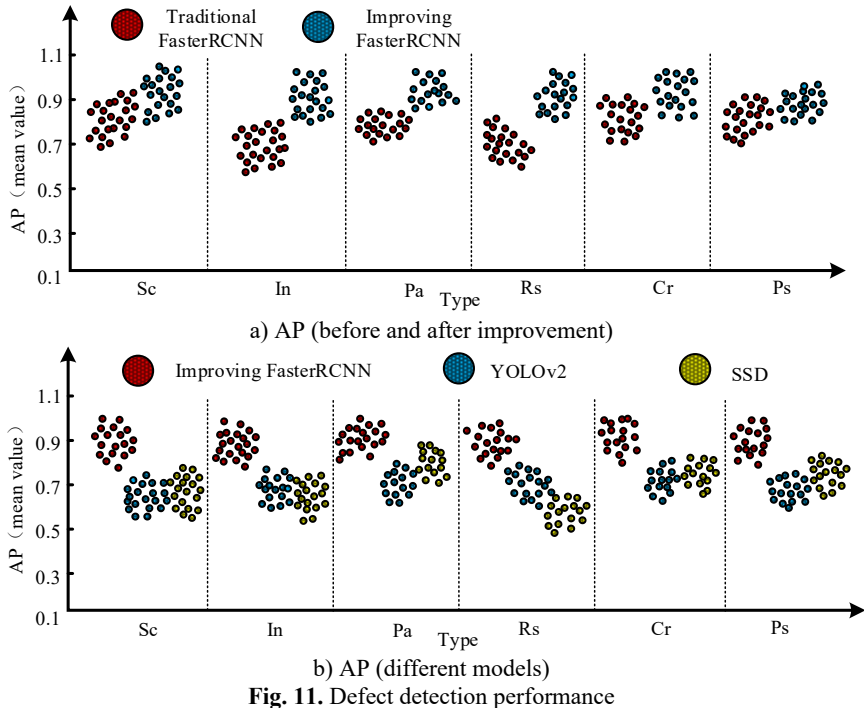
As can be seen in Fig. 10, the improved AlexNet network achieves a great improvement in F1 values for each defect type, especially for the recognition of scratches, patches, and cracks with 100 % accuracy. In addition, the improved network also succeeded in the classification of pockmarks, reaching an F1 value of 98.6 %. In addition, after comparison, it was found that the improved AlexNet network had great advantages in both F1 value and FPS. Compared with Goog Le Net with the highest F1 value among the four networks, the F1 value of the improved AlexNet network was improved by 0.29 %, and the FPS of the improved AlexNet network reached 237.5 f/s, which was the highest among the classical CNN classification networks. This indicated that the improved AlexNet network was better in classification performance compared to the conventional CNN network.

### 4.3. Defect location detection effect analysis

For the field of image defect detection, the Average Precision (AP) metric is one of the important performance evaluation parameters. The defect detection effectiveness of different algorithms can be evaluated by comparing their AP scores to classify and select the appropriate algorithm. This experiment compared the performance of the conventional Faster RCNN network with the optimized Faster RCNN network and the SSD network and YOLOv2 network for the detection of SPSD, as shown below.



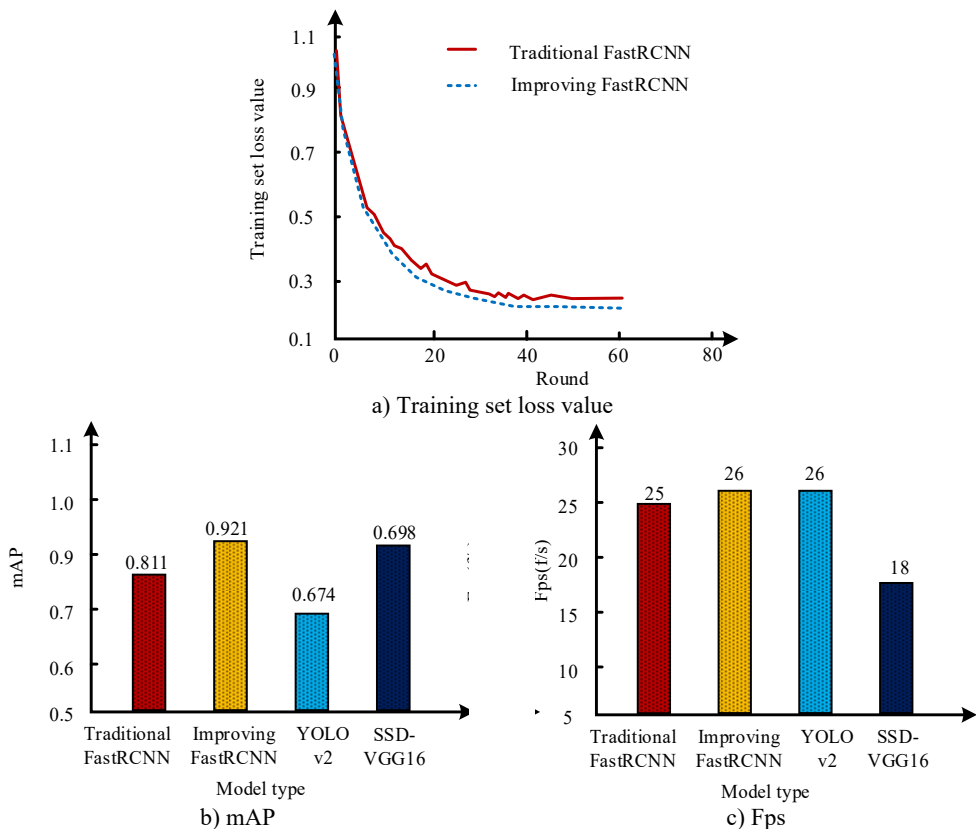
**Fig. 10.** Improving AlexNet performance analysis



**Fig. 11.** Defect detection performance

From Fig. 11, the traditional Faster RCNN network had an AP value below 0.9 for all kinds of defects detection, while the optimized Faster RCNN network had an AP value above 0.9 for all six defects detection, especially for small size defects detection. The APs of the SSD network were slightly higher than those of the YOLOv2 network, but still lower than those of the optimized Faster RCNN network. Therefore, the optimized Faster RCNN network was one of the best-performing algorithms for surface defect detection with high accuracy and robustness, which was ideal for practical applications. A comparison of network convergence status and metrics is shown in Fig. 12.

In Fig. 12, the optimized Faster RCNN network performed excellently. Compared with the conventional network, the mAP value of the optimized Faster RCNN network was improved by 0.110 and the FPS was not reduced. Compared with the SSD network and YOLOv2 network, the detection accuracy of the optimized Faster RCNN network was higher. In addition, the optimized Faster RCNN network performed better in terms of loss values and converged better. Therefore, the model designed in the study had more performance advantages, had higher detection accuracy and performance stability, and had wide application prospects.



**Fig. 12.** Comparison of network convergence status and indicators

## 5. Conclusions

This study mainly addresses how to detect and classify surface defects on steel plates. Firstly, the defects are localized by image pre-processing and using the ROI detection technique; secondly, a more powerful defect classification model is designed using an improved AlexNet network, and the defects are detected using the Faster RCNN network. The data results showed that in terms of detection rate, the detection rate of different defect types ranged from the lowest 88 % to the

highest 100 %, with an average detection rate of 98 %; in terms of detection time, the shortest was 0.0011 seconds and the longest was 0.0083 seconds. The improved AlexNet network had significant advantages in terms of F1 value and FPS, and its F1 value reached 237.5 f/s, which was the highest among the classical CNN classification networks. In addition, the optimized Faster RCNN network had higher detection accuracy compared with the SSD network and YOLOv2 network. The proposed method has high practicality and value in practical applications, the improved AlexNet network is better in classification performance and the optimized Faster RCNN network is one of the best-performing algorithms in surface defect detection with wide application prospects.

## Acknowledgements

The authors have not disclosed any funding.

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- [1] R. Hao, B. Lu, Y. Cheng, X. Li, and B. Huang, "A steel surface defect inspection approach towards smart industrial monitoring," *Journal of Intelligent Manufacturing*, Vol. 32, No. 7, pp. 1833–1843, Oct. 2021, <https://doi.org/10.1007/s10845-020-01670-2>
- [2] H. Chen, Y. Pang, Q. Hu, and K. Liu, "Solar cell surface defect inspection based on multispectral convolutional neural network," *Journal of Intelligent Manufacturing*, Vol. 31, No. 2, pp. 453–468, Feb. 2020, <https://doi.org/10.1007/s10845-018-1458-z>
- [3] J. P. Yun, W. C. Shin, G. Koo, M. S. Kim, C. Lee, and S. J. Lee, "Automated defect inspection system for metal surfaces based on deep learning and data augmentation," *Journal of Manufacturing Systems*, Vol. 55, pp. 317–324, Apr. 2020, <https://doi.org/10.1016/j.jmsy.2020.03.009>
- [4] X. Pan and T. Y. Yang, "3D vision-based out-of-plane displacement quantification for steel plate structures using structure-from-motion, deep learning, and point-cloud processing," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 38, No. 5, pp. 547–561, Mar. 2023, <https://doi.org/10.1111/mice.12906>
- [5] B. Taşar, "Comparative analysis of machine learning algorithms for steel plate fault detection," *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, Vol. 10, No. 3, pp. 1578–1588, 2022.
- [6] C. Y. Park, J. W. Kim, B. Kim, and J. Lee, "Prediction for manufacturing factors in a steel plate rolling smart factory using data clustering-based machine learning," *IEEE Access*, Vol. 8, pp. 60890–60905, 2020, <https://doi.org/10.1109/access.2020.2983188>
- [7] Y. Guo, Z. Mustafaoglu, and D. Koundal, "Spam detection using bidirectional transformers and machine learning classifier algorithms," *Journal of Computational and Cognitive Engineering*, Vol. 2, No. 1, pp. 5–9, Apr. 2022, <https://doi.org/10.47852/bonviewjccce2202192>
- [8] J. Zan, "Research on robot path perception and optimization technology based on whale optimization algorithm," *Journal of Computational and Cognitive Engineering*, Vol. 1, No. 4, pp. 201–208, Mar. 2022, <https://doi.org/10.47852/bonviewjccce597820205514>
- [9] S. Feng, K. Ji, L. Zhang, X. Ma, and G. Kuang, "SAR target classification based on integration of ASC parts model and deep learning algorithm," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, pp. 10213–10225, 2021, <https://doi.org/10.1109/jstars.2021.3116979>
- [10] J. Yang, C. Wang, B. Jiang, H. Song, and Q. Meng, "Visual perception enabled industry intelligence: state of the art, challenges and prospects," *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 3, pp. 2204–2219, Mar. 2021, <https://doi.org/10.1109/tii.2020.2998818>



- [11] W. Dai et al., “Deep learning assisted vision inspection of resistance spot welds,” *Journal of Manufacturing Processes*, Vol. 62, pp. 262–274, Feb. 2021, <https://doi.org/10.1016/j.jmapro.2020.12.015>
- [12] W. Wu and Q. Li, “Machine vision inspection of electrical connectors based on improved Yolo v3,” *IEEE Access*, Vol. 8, pp. 166184–166196, 2020, <https://doi.org/10.1109/access.2020.3022405>
- [13] N. Ismail and O. A. Malik, “Real-time visual inspection system for grading fruits using computer vision and deep learning techniques,” *Information Processing in Agriculture*, Vol. 9, No. 1, pp. 24–37, Mar. 2022, <https://doi.org/10.1016/j.inpa.2021.01.005>
- [14] Y. Han, J. Fan, and X. Yang, “A structured light vision sensor for on-line weld bead measurement and weld quality inspection,” *The International Journal of Advanced Manufacturing Technology*, Vol. 106, No. 5-6, pp. 2065–2078, Jan. 2020, <https://doi.org/10.1007/s00170-019-04450-2>
- [15] I. J. Jacob and P. E. Darney, “Design of Deep Learning Algorithm for IoT Application by Image based Recognition,” *September 2021*, Vol. 3, No. 3, pp. 276–290, Aug. 2021, <https://doi.org/10.36548/jismac.2021.3.008>
- [16] G. Ranganathan, “A study to find facts behind preprocessing on deep learning algorithms,” *Journal of Innovative Image Processing*, Vol. 3, No. 1, pp. 66–74, Apr. 2021, <https://doi.org/10.36548/jiip.2021.1.006>
- [17] A. Kaushal, R. Altman, and C. Langlotz, “Geographic distribution of US cohorts used to train deep learning algorithms,” *JAMA*, Vol. 324, No. 12, pp. 1212–1213, Sep. 2020, <https://doi.org/10.1001/jama.2020.12067>
- [18] S. Wang et al., “A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19),” *European Radiology*, Vol. 31, No. 8, pp. 6096–6104, Feb. 2021, <https://doi.org/10.1007/s00330-021-07715-1>
- [19] M. G. Pelletier, J. D. Wanjura, G. A. Holt, and N. Kothari, “Cotton Gin Stand Machine-Vision Inspection and Removal System for Plastic Contamination: Auto-Calibration Design,” *AgriEngineering*, Vol. 5, No. 3, pp. 1243–1258, Jul. 2023, <https://doi.org/10.3390/agriengineering5030079>
- [20] R. Fang, H. Yi, A. Shu, X. Lv, and S. Wang, “Illumination-robust milling surface roughness machine vision inspection based on MAML++ network,” *Optical Engineering*, Vol. 61, No. 12, p. 124105, Dec. 2022, <https://doi.org/10.1117/1.oe.61.12.124105>



**Yan Huang**, Master of Engineering, Lecturer. Graduated from the Sichuan University of Science and Engineering in 2010. Worked in Chongqing Chemical Industry Vocational College. Her research interests include artificial intelligence, pattern recognition, intelligent information processing.