Journal of Computers Vol. 35 No. 2, April 2024, pp. 135-150
doi: 10.53106/199115992024043502009

Application of Improved Convolutional Neural Network in Defect Identification of Exhaust Pipe Welds

Qingfang Liu[1*], Xiaoning Bo[1,2], Jinrong Xu[1], Jin Wang[1], Honglan Li[1]

[1] Department of Electrical Engineering, Taiyuan Institute of Technology, Taiyuan 030008, China
[2] School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

liuqf@tit.edu.cn, boxn@tit.edu.cn, 275919672@qq.com,
wangj@tit.edu.cn, lihonglan0207@126.com

Received 27 March 2024; Revised 4 April 2024; Accepted 15 April 2024

Abstract. This article focuses on the identification of welding defects in engine exhaust pipe welds. Firstly, a binocular vision system is built, and the models and parameters of the cameras and lenses involved in the entire system are explained in detail. At the same time, the cameras are calibrated; Then, in response to the problems of large volume, low efficiency, and lack of attention mechanism in the current neural network model, the network model was improved by adding MP structure, CA attention mechanism, and other methods to improve the recognition efficiency of the model. Finally, the reliability of the proposed method was verified through simulation experiments, and the overall recognition efficiency was improved to 97.28%.

Keywords: deep learning, welding seam, binocular vision, YOLOv7

1 Introduction

The engine exhaust pipe is one of the important components of the engine system, and its operation directly affects the reliability, durability, power, and exhaust emissions of the engine. In addition, the exhaust temperature of the engine is relatively high, and the exhaust pipe works at high temperatures for a long time. The exhaust pipe also needs to withstand the load generated by its own weight under the dynamic load of the car, as well as the vibration caused by the car's bumps. Currently, the materials used for engine exhaust pipes are mainly concentrated in cast steel, cast iron, and stainless steel, all of which have good heat resistance characteristics. However, in harsh working environments such as long-term vibration and high-temperature operation, exhaust pipes still face the risk of fatigue failure, which mainly manifests as excessive stress concentration and fatigue cracking.

After the operation of a fuel engine, combustion gases are generated internally, and the high-temperature and high-pressure gases test the performance of the engine. The sealing performance of the engine exhaust pipe is the most basic technical requirement of the engine. Therefore, it is necessary to inspect the weld seams of the engine exhaust pipe, and the quality of the weld seam determines the quality of the engine [1].

During the welding process of exhaust pipes, various unpredictable welding defects may occur in a few welds due to natural environmental interference or unstable welding processes. Research has found that there is a significant difference in the size of welding defects in exhaust pipes, and the detection difficulty, missed detection rate, and false detection rate of small-sized targets mainly consisting of circular defects are high; Cracks, pores, lack of penetration, and lack of fusion are the main defects [2]. Currently, there are the following defects in the detection of weld defects, which are the starting points for the formation of the research plan in this article.

1) Some low tech enterprises still rely mainly on manual testing. Although electronic microscopes, non-destructive testing equipment, and other equipment are used in the testing process, the determination and identification of defects still rely on manual experience, which is not sufficient to support large-scale production conditions.

2) In the scenario of using artificial intelligence to identify defects, there is a situation where the dataset is not complete enough, resulting in insufficient training of the model and ultimately leading to an error rate in defect recognition.

3) The recognition model has a large volume and high requirements for hardware equipment such as vision in enterprises, especially GPUs, which leads to an increase in production costs.

4) The recognition model takes a long time to recognize, and the recognition speed cannot match continuous

* Corresponding Author

and rapid production scenarios.

Therefore, in response to the current issues in identifying exhaust pipe welds, the work done in this article regarding the detection of welding defects in vehicle exhaust pipes is as follows:

1) Analyzed the characteristics of exhaust pipe welds and common welding defects in smaller welds, preparing for the establishment of subsequent datasets;

2) Build a binocular vision system that can complete visual acquisition of welds;

3) Build an improved image recognition model and improve defect recognition ability by adding CA attention mechanism while reducing model size.

In order to clarify the above work, the chapter structure of this article is as follows: Chapter 2 mainly introduces the relevant research results of relevant scholars; Chapter 3 establishes a visual inspection system and camera calibration work; Chapter 4 mainly introduces the improvement process of neural network models; Chapter 5 is the simulation experiment results section; Chapter 6 is the conclusion.

2 Related Work

Many scholars have made many research achievements in the rapid identification of welds, improving recognition accuracy, and improving recognition models. Only by fully analyzing and referencing existing achievements can we find the direction of this study on the basis of existing achievements.

Lechao Xiong proposed a DR image automatic recognition method for the high error rate of weld defects in big data analysis. After processing the image grayscale, the defect position was determined using the Sigma criterion, and the defect area could be determined based on the pixel points of the defect position. The recognition accuracy can be improved by 12% [3].

Fangqi Zhao, whose research focuses on oil and gas pipelines, has achieved quantitative identification and classification of weld defects and proposed a defect recognition method based on local binary patterns and support vector machines, with a classification accuracy of 95% [4].

Yong Zhang, based on the BM feature point matching algorithm and pixel scanning method, obtained a three-dimensional information dataset of continuous feature points on the contour of the weld seam for the cabin grid weld seam. He proposed a strategy based on binocular vision to obtain three-dimensional information of the weld seam path, achieving recognition error control of the weld seam track within 2 millimeters [5].

Shuqiang Wang proposed an image recognition method for circular pipe intersecting line welds, improved image calibration and segmentation methods, and achieved an average processing time of 106ms and an average error of 0 in the processing of the weld model 26 mm, maximum error 0 49 mm [6].

Xiaohu Hang, also using vision and deep learning, studied weld seam recognition methods in this article. Convolutional series networks, self attention networks, and Transformer models were used for weld seam recognition, providing a new solution for the problem of weld seam recognition [7].

Wenkai Xiao established a similarity evaluation method for false images in natural gas pipeline weld seam inspection, calculated the feature similarity of weld seam images, screened duplicate images, established a feature database, and conducted similarity judgment. The experimental results show that under a certain threshold setting, the algorithm can achieve a recognition result of 92.3% for duplicate images [8].

The above methods mainly focus on improving the recognition methods. Some scholars have identified two-dimensional defects obtained from non-destructive testing. Based on two-dimensional image recognition, the efficiency of image processing has been improved, and the recognition speed of weld defects has also been increased.

Zeyu Yu improved the deep learning algorithm for identifying ultrasonic inspection results, obtained sufficient inspection image datasets through data augmentation, and then used an improved memory network model to complete defect recognition of pipeline welds [9].

Ling Long from the Chinese Academy of Sciences aims to improve the efficiency and accuracy of X-ray film evaluation. This paper proposes a real-time X-ray weld defect detection method based on a lightweight YOLO network, designs a cascaded defect detection model, and uses a lightweight designed weld positioning network to locate the weld seam area where defects are concentrated, improving the model's lightweighting level [10].

Hongbin Ma established an intelligent rail flaw detector to collect a dataset of steel rails in high-altitude and cold regions, focusing on the two-dimensional images obtained from ultrasonic image recognition technology in rail flaw detection on the Qinghai Tibet Railway, especially for common types of damage such as rail head nuclear damage and rail surface fish scale damage. The dataset was organized, processed, and model trained using the

YOLOv5 method to accurately identify and locate rail head nuclear damage and rail surface fish scale damage [11].

3 Establishment of Binocular Vision System

The selection of image acquisition equipment is determined based on the working environment and detection object of this article. During operation, the welding gun is 0-20 centimeters away from the workpiece, and the two selected camera models are HT-UB300C industrial cameras. According to the calculation of relevant parameters, the depth of field of the camera is about 3.96m, which meets the corresponding visual range requirements in the experiment. At the same time, the lens resolution is high and the distortion is small, which can obtain high-quality welding image. The camera and computer hardware parameters are shown in Table 1. Subsequent image processing is implemented using MATLAB R2019a software programming, and camera calibration is achieved using the Stereo Camera Calibrator toolbox in the software.

Table 1. Models and parameters of cameras and computer hardware

Parameter	Parameter values
Computing system	Windous 10
Computer processor	Inter Core i5-4200H
CUP main frequency	2.8 GHz
Computer memory	8 GB
System type	64
Industrial camera model	HT-UB300C
Industrial camera sensors	CMOS
Pixel size of industrial cameras	3.2×3.2μm
Industrial camera resolution	1280×960

3.1 Establishment of Binocular Vision Model

The model of binocular vision is shown in Fig. 1. Using a binocular camera, the depth information of a point in vision can be calculated by its imaging position on another imaging plane [12].

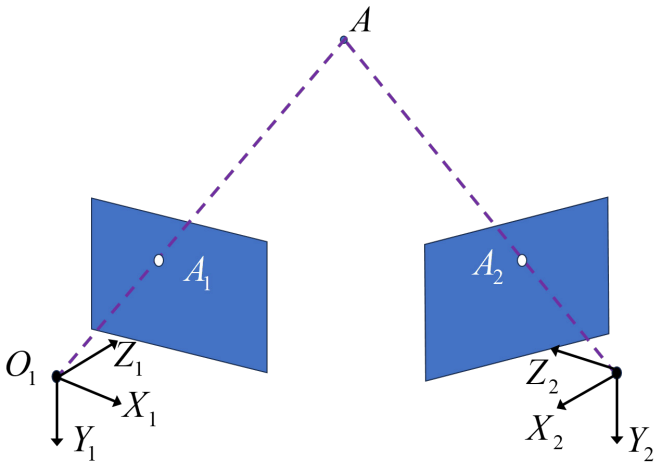

Fig. 1. Binocular visual structure model

The intelligent welding task requires converting pixel coordinates in the image into real-world coordinates, so it is very important to establish a system coordinate system and solve the conversion relationship between different coordinate systems. In the camera imaging model, it mainly includes the mutual conversion of four coordinate systems.

The world coordinate system can artificially specify its origin coordinates and the direction of the coordinate axis, used to describe the specific position of an object in three-dimensional space. Generally, the world coordinate system is set to coincide with the camera coordinate system.

The camera coordinate system is a three-dimensional coordinate system with the camera's optical center as the origin and the Z-axis coinciding with the camera's optical axis. The camera's optical axis is perpendicular to the image plane, and the distance between the camera coordinate system origin and the image plane is the camera focal length. The imaging plane of a camera is actually composed of many pixels, and the total number and size of pixels in different models of cameras also vary. Generally, the upper left corner of the imaging plane is selected as the origin of the pixel coordinate system, with the horizontal axis horizontally to the right and the vertical axis vertically downward. The coordinate value (x, y) represents the row and column coordinates of the pixel point in the image plane, and its unit is pixels.

The physical coordinate system of the image is coplanar with the pixel coordinate system, and its horizontal and vertical axes are parallel and in the same direction. The origin of this coordinate is the intersection point between the camera's optical axis and the image plane, measured in physical units of millimeters [13]. The coordinate system coordinate transformation model is shown in Fig. 2.

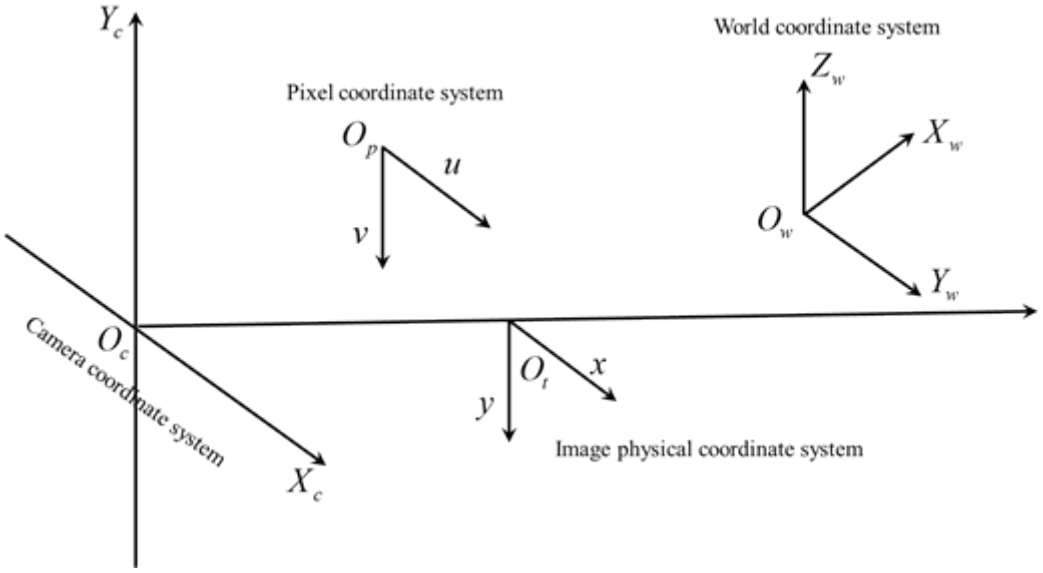

Fig. 2. Coordinate relationship model

According to the binocular model, the coordinate systems of a point in space in the world coordinate system and the left and right camera coordinate systems are x_0, x_z, x_y respectively. The conversion relationship between the camera coordinate system and the world coordinate system is as follows:

$$\begin{cases} x_z = R_z x_w + A_z \\ x_y = R_y x_w + A_y \end{cases}. \tag{1}$$

Among them, R_z and R_y represent the rotation matrices of the left and right cameras, while A_z and A_y represent the translation matrices. From the above equation, it can be further concluded that:

$$x_z = R_z R_y^{-1} x_y + A_z R_y^{-1} A_y. \tag{2}$$

The relative rotation and translation between cameras are represented as R_{zy} and A_{zy}, then:

$$\begin{cases} R_{xy} = R_x R_y^{-1} \\ A_{xy} = A_x - R_y^{-1} A_y \end{cases}. \tag{3}$$

The position conversion relationship between cameras is represented as:

$$x_z = R_{zy} x_y + A_{yx}. \tag{4}$$

The conversion relationship between the camera coordinate system and the image physical coordinate system is:

$$\frac{BC}{O_t A} = \frac{CO_c}{O_1 O_c} = \frac{QB}{AQ_t} = \frac{X_c}{x} = \frac{Y_c}{y} = \frac{Z_c}{f}. \tag{5}$$

The transformation relationship between the camera coordinate system and the image physical coordinate system can be expressed as:

$$x_t = f \frac{x_c}{z_c}. \tag{6}$$

$$y_t = f \frac{y_c}{z_c}. \tag{7}$$

$$z_c \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \tag{8}$$

f represents focal length, (x_t, y_t, z_t) represents points in the physical coordinate system of the image, and (x_c, y_c, z_c) represents points in the camera coordinate system. However, due to the use of physical units of millimeters in the image physical coordinate system, the unit length in the pixel coordinate system is pixels, and the pixel size is unknown. Therefore, assuming that the physical lengths of the unit pixel on the horizontal and vertical axes are l_x and l_y, respectively, the conversion relationship between the two coordinate systems can be expressed as follows:

$$\begin{cases} u = \dfrac{x}{d_x} + u_0 \\ v = \dfrac{y}{d_y} + v_0 \end{cases}. \tag{9}$$

The transformation of pixel coordinates and image coordinate matrices is represented as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{10}$$

The conversion from pixel coordinate system to world coordinate system is shown in Fig. 3:

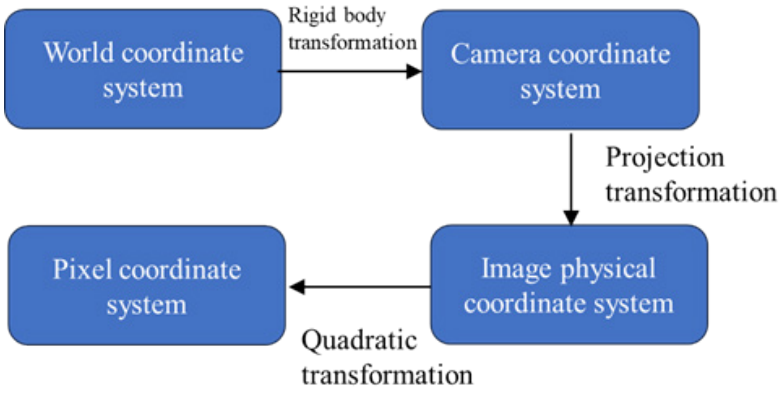

Fig. 3. Coordinate transformation model

After the above process, the conversion between different coordinate systems is achieved and serves as the basis for the visual model.

3.2 Camera Calibration

The camera calibration process collects calibration board images from different poses, and based on the Zhengyou Zhang [14] calibration method, calibration calculations are carried out using the calibration toolbox provided by MATLAB software. In the selection of calibration boards, the more widely used black and white chessboard calibration board was not used, but the circular dot array calibration board was selected. The material of the calibration board determines the accuracy of the corner points in the surface image of the calibration board, which in turn affects the calibration accuracy of the camera parameters of the entire binocular system. Calibration plates usually contain materials such as ceramics, aluminum substrates, and film plates. Among them, ceramic calibration plates can reduce the reflectivity by more than 70% compared to general metal and glass calibration plates, and have high durability and accuracy. So this system uses a calibration board made of ceramic materials as the camera calibration object.

The specific camera calibration process is as follows:

1) Create a calibration board description file, and select the origin array calibration board for camera calibration. The calibration board is shown in Fig. 4.

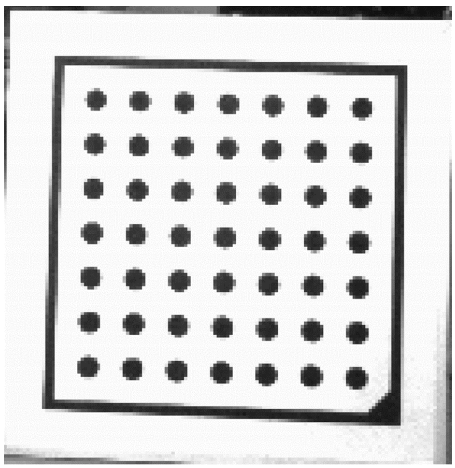

Fig. 4. Camera calibration board

To ensure the quality and accuracy of the images, it is generally recommended to make the calibration board account for about one-third of the size of the image field of view. The grayscale value of the imaging should be greater than 128, and different images should have significant changes in the orientation and position of the calibration board. In this calibration experiment [15], 12 images were collected from the left and right cameras respectively.

3) The calibration results are shown in Table 2.

Table 2. Camera calibration results

	Left camera	Right camera
Rotation matrix		$\begin{bmatrix} 0.9895 & -0.0031 & -0.0289 \\ 0.0021 & 1.2128 & -0.0049 \\ 0.0371 & 0.0052 & 0.8749 \end{bmatrix}$
Translation matrix		$\begin{bmatrix} -83.0103 & 0.3002 & 3.2981 \end{bmatrix}$
Internal parameter matrix	$\begin{bmatrix} 3198.3897 & 0 & 1495.7981 \\ 0 & 3289.9027 & 1109.0983 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3192.5098 & 0 & 1623.3094 \\ 0 & 3306.2813 & 1123.0726 \\ 0 & 0 & 1 \end{bmatrix}$
Distortion coefficient	$\begin{bmatrix} -0.06012 & 0.1009 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -0.0512 & 0.0401 & 0 & 0 \end{bmatrix}$

After completing the camera parameter calibration, in order to ensure that the calibrated camera inner and outer parameter matrices can meet the accuracy requirements of the welding scene, it is necessary to verify the calibration accuracy. Two corner points with known diagonal distances from the calibration board are selected and compared with the distance calculated by the camera calibration parameters from the image projection point. The results obtained are shown in Fig. 5. The average absolute error value of this distance is 2.98mm. By analyzing the placement angle of the calibration board, it is advisable to avoid measuring calibration boards placed at angles greater than 45 degrees during the testing process.

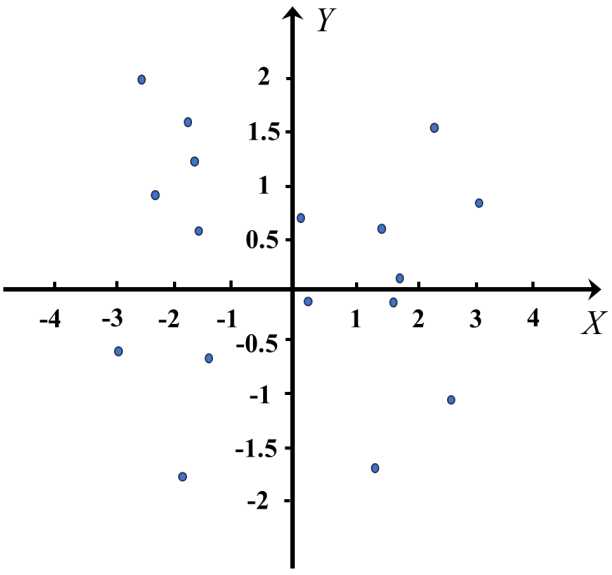

Fig. 5. Binocular visual structure model

After the above process, the camera modeling and calibration work have been completed, enabling the camera to have the foundation for accurate image acquisition.

4 Establishment of Binocular Vision System

When selecting a camera, in order to improve the robustness of the system, a larger model of camera field of view will be chosen, which will result in the target image only occupying a small part of the camera field of view in the images obtained by the binocular camera, while the rest of the images are invalid data. Therefore, in order to effectively reduce the amount of data processing, an improved YOLOv7 [16] model is adopted to obtain the ROI of exhaust pipe welds and improve the algorithm's focus.

YOLOv7 is a high-speed and high-precision object detection model suitable for real-time scene object detection tasks. It includes three basic network models that perform well in terms of speed and accuracy, and can be used as object detectors on embedded systems or mobile devices. Its detection logic is basically similar to the previous YOLO algorithm series. The YOLOv7 network model consists of four parts: input, backbone network, neck, and head. The input part is preprocessed with data augmentation, and the image is sent to the backbone network to extract features. Then, the neck module is fused to obtain features of different sizes. Finally, the fused features are fed into the detection head and the results are output after detection. The YOLOv7 network model mainly consists of convolution, MX module, MCB module, and SPPCSP module. The SPPCSP module adds multiple parallel MaxPool operations after convolution, increasing the receptive field and adapting the algorithm to different resolution images. It also adopts a residual structure like method to combine the conventional processed features with the pooled features, improving the speed and accuracy of the algorithm. The main function of the MX module is downsampling, which combines the features that have been maximally pooled with the features after conventional convolution to form a super downsampling effect [17].

In order to achieve miniaturization of the YOLOv7 model, reduce the false detection rate, missed detection rate, and improve the computational speed of the model, targeted improvements need to be made to the model. The detailed improvement process is as follows:

The improved model is divided into three parts in terms of prediction methods.

1) Feature extraction. The input image is feature extracted in the backbone network, and an MP structure is added. The MP layer consists of Maxpool and convolutional blocks, forming a dual path to enhance the network's feature fusion ability.

2) Due to the gradient vanishing effect caused by increasing depth in deep neural networks, predicting the object situation corresponding to the prior box.

3) This section consists of SPPCSPC module and RepConv module. Among them, the SPPCSPC module adds a residual edge on top of the SPP module and stacks it with the features output after maximum pooling. The RepConv structure can introduce special residual structures to assist in training, reducing the complexity of the network while ensuring its predictive performance. The network structure is shown in Fig. 6.

At the same time, a CA attention mechanism is added to the YOLOv7 model to perform global average pooling of feature maps along the vertical and horizontal directions [18], so that the input features from both directions are aggregated into independent directional feature maps. Encode them separately into two attention maps, each capturing the long-range correlation of the input feature map along a spatial direction. Apply the two attention maps to the input feature map through multiplication to emphasize the representation of interest. This module does not require additional parameters as feature maps to derive attention weights for weld seam recognition. The CA attention mechanism is based on some well-known neuroscience theories, by proposing an optimized energy function to facilitate the identification of the importance of each neuron. The advantage of CA attention mechanism is that it selects most operators based on the solution of the defined energy function, avoiding the energy consumption in structural adjustment, evaluating the importance of each neuron, and aiming to better achieve attention. Therefore, after the above process, the improved module structure of the model was obtained. The attention mechanism switches the output dimension of the width direction feature map and concatenates it with the height direction feature map. After concatenation, the convolution of A is used to compress the number of intermediate feature map channels to. Then, batch normalization is used to normalize the feature map in batches and perform feature mapping with nonlinear operations. After mapping, A convolution is used to decompose the feature map into two independent tensors along the spatial dimension. The output dimension of the width direction feature map is restored to the input state. The CA attention mechanism model is shown in Fig. 7.

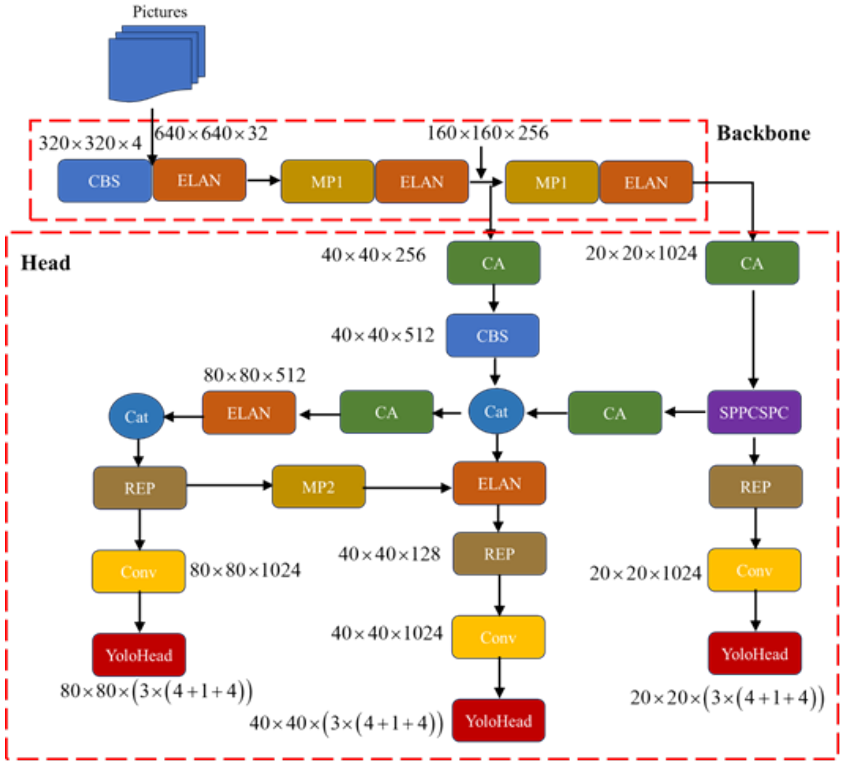

Fig. 6. Overall framework of the model

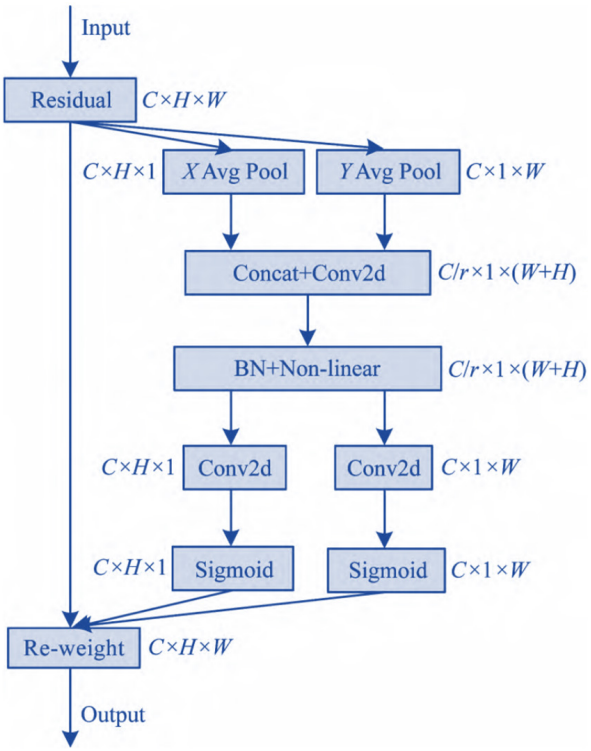

Fig. 7. CA attention mechanism network structure diagram

The various modules added to the model are shown in Fig. 8.

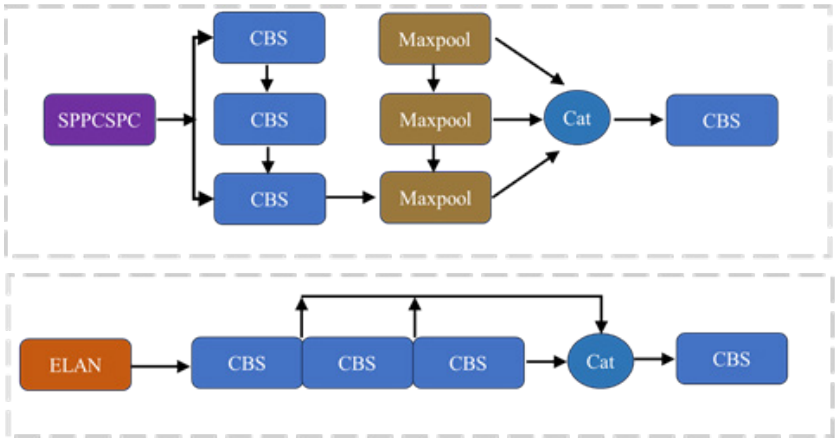

Fig. 8. Overall framework of the model

4.1 Model Description

Firstly, the input feature maps are pooled in the X and Y directions. Then, in order to capture images of remote spatial interactions with precise positional information, the global average pooling is further decomposed. The decomposition formula is as follows:

$$\begin{cases} Z_c^h(h) = \sum_{0 \le j \le W} x_c(h,i) \\ Z_c^w(w) = \sum_{0 \le j \le H} x_c(j,w) \end{cases}. \tag{11}$$

In the formula, x_c represents the c-th channel of input feature X, and h and w represent the height and width of the training model. Z^h and Z^w are attention feature maps. Then, the attention feature map is concatenated and activated using the following formula:

$$t = \alpha\left(B_z\left|Z^h, Z^w\right|\right). \tag{12}$$

α is the activation function, and B_z is the dimensionality reduction operation matrix of the convolutional kernel.

4.2 Model Description

The loss functions in the YOLOv7 network model include angle loss, distance loss, shape loss, and IoU loss, which are multiplied by different weight coefficients [19]. Add an IoU prediction component to the confidence score to measure the accuracy of localization. During the training process, the proposed method incorporates IoU prediction into the logistic regression of confidence scores and adds a parallel channel prediction IoU. Each grid unit predicts 3 prediction boxes, and the number of basic parameters for each prediction box increases from 5 to 6. By adding the number of categories, the output parameter quantity for each prediction box is obtained. When making predictions, multiply the predicted IoU by the classification probability to obtain the final detection confidence [20]. This confidence level effectively avoids the risk of being excluded due to low scores in

accurate weld seam target detection boxes, improves the network's ability to characterize and extract weld seam defect features, and optimizes model parameters.

1) The angle loss is represented as follows:

$$\omega = 1 - 2\sin^2\left(\arctan\left(\frac{h_z}{l_z}\right) - 45^\circ\right). \tag{13}$$

In the formula, ω is the angle loss, h_z is the height difference between the center points of the real box and the predicted box, and l_z is the distance between the center points of the real box and the predicted box.

2) The distance loss formula is expressed as follows:

$$\Delta S = \sum\left(1 - e^{1-\lambda\left(p_x + p_y\right)}\right)$$
$$p_x = \left(\frac{x_z - y_z}{c_w}\right)^2, \; p_y = \left(\frac{x_z - y_z}{c_h}\right)^2. \tag{14}$$

In the formula, ΔS represents distance loss, λ represents angle loss relationship, x_z represents center horizontal axis, and y_z represents center vertical axis.

The formula for shape loss is expressed as follows:

$$\Pi = \left(1 - e^{-w_a}\right)^\theta + \left(1 - e^{-w_b}\right)^\theta. \tag{15}$$

$$w_a = \frac{\left|w - w^g\right|}{\max\left(w, w^g\right)}, \; w_b = \frac{\left|h - h^g\right|}{\max\left(h, h^g\right)}. \tag{16}$$

In the formula, Π represents shape loss, w, h, w^g, and w^h represent the width and height of the predicted box and the true box, respectively, and the value range of parameter θ is [2, 6]. The expression for the final loss function is expressed as:

$$L_{ost} = 1 - IoU + \left(\Delta S + \Pi\right)/2. \tag{17}$$

In the formula, IoU is the intersection and union ratio between the predicted box and the real box. The loss function takes into account the direction between the real box and the predicted box, and introduces the vector angle between the real box and the predicted box, making the convergence speed faster.

5 Defect Detection Experiment

Before collecting images, it is necessary to classify and summarize typical weld defects, including porosity, dense porosity, cracks, lack of fusion, and incomplete penetration. Each typical image is shown in Fig. 5.

This study used the proposed system to collect 3298 defect samples of exhaust pipe weld seam defect removal images. Random cropping, horizontal flipping, vertical flipping, random translation, Gaussian blur and other data augmentation methods were used in any combination to expand the dataset. The annotated and enhanced dataset is divided into training, testing, and validation sets in an 8:1:1 ratio.

In the data augmentation stage, a mixed data augmentation method is used to enhance the dataset data, which utilizes linear interpolation to construct additional training samples and labels [21]. The formula for handling labels is shown in equations 18 and 19:

$$\tilde{x} = \lambda x_i + (1+\lambda) x_j . \tag{18}$$

$$\tilde{y} = \lambda y_i + (1-\lambda) y_j . \tag{19}$$

In the formula, (x_i, y_i) and (x_j, y_j) are the training sample pairs in the original dataset, λ is a parameter that follows the β distribution, and $\tilde{x}$ is the training sample enhanced by mixed data; $\tilde{y}$ represents the data result of weld defects after data augmentation processing with different fusion ratios. The enhancement effect is shown in Fig. 9.

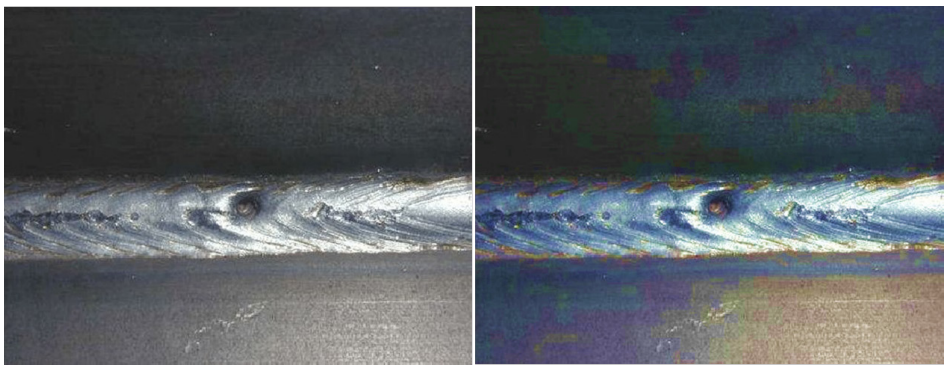

Fig. 9. The effect image after data augmentation

After the above process, a dataset of weld defects was formed in this article, and various types of weld defects in the dataset are shown in Fig. 10.

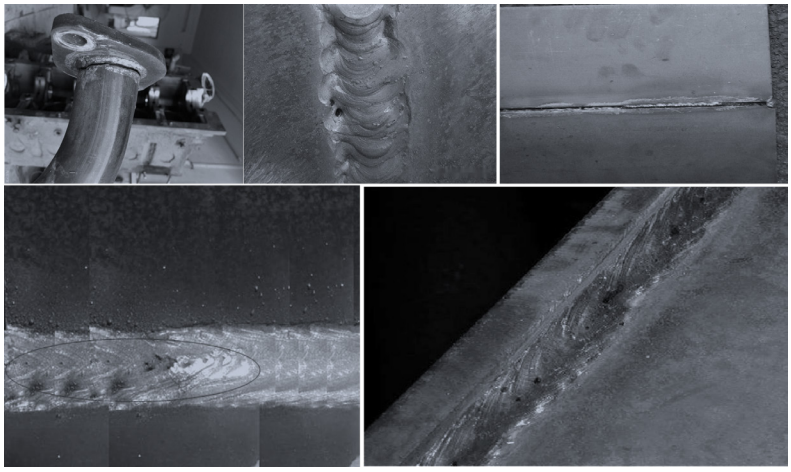

Fig. 10. Overall framework of the model

5.1 Model Description

The network testing environment is Python 1.11 and Python 3.8. The specific hardware configuration and model parameters are shown in Table 3.

Table 3. Models and parameters of cameras and computer hardware

Name	Parameter values
GPU	RTX4090
CPU	Intel Core i7 14700
CUDA	11.7
CuDNN	10.2
Pixel	640×640
Learning rate	0.001
Optimizer	Adam
Batch size	18

5.2 Model Description

The model training frequency is set to 480, and as the number of iterations increases, the model gradually converges. The convergence effect is shown in Fig. 11.

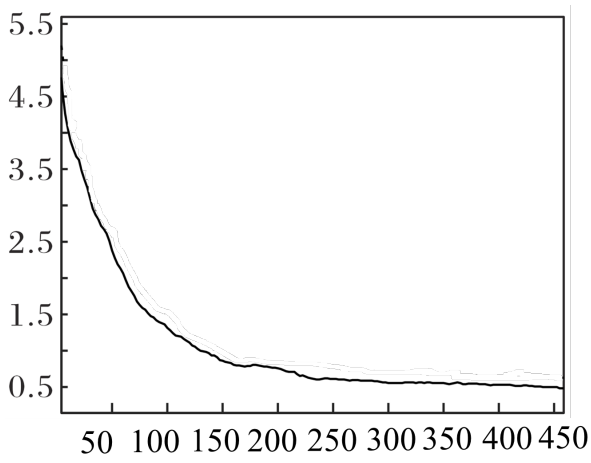

Fig. 11. Model convergence effect

After ablation experiments, the following results were obtained: Firstly, after incorporating the CA attention module into YOLOv7, the mAP50 value increased by 6.53%, and the accuracy of network detection was significantly improved, reducing the possibility of false detection and missed detection of sign language actions in nonlinear environments. The number of model parameters has been reduced by 0.29 × 106, and the mAP is 97.28%, which has almost no impact. However, the calculation speed of the model has been improved, reducing the hardware resources required for model deployment.

5.3 Comparison of Different Network Models

Under the same experimental environment and training parameters, this paper compares the improved YOLOv7 network model with other network models to demonstrate the advantages of the improved network model. For the convenience of demonstration, the improved YOLOv7 is referred to as I-YOLOV7,

To objectively evaluate the true effect of the experiment, universal object detection evaluation indicators are used, mainly including Precision (P), Recall ®, mean Precision (mAP), and the predicted results will be divided

pixel by pixel with the real labels and categories, including true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The calculation formulas are as follows:

$$P = \frac{TP}{TP + FP}. \tag{20}$$

$$R = \frac{TP}{TP + FN}. \tag{21}$$

$$AP_i = \int_0^1 P\left(R_i\right) dR. \tag{22}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k}. \tag{23}$$

In the formula, AP represents the average of the highest accuracy under different recall rates, and the experimental results are shown in Fig. 12.

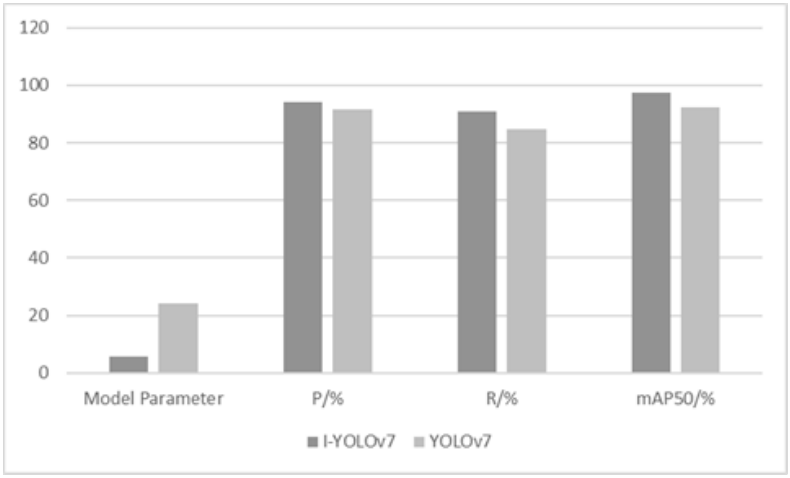

Fig. 12. model comparison

After analyzing the data in Fig. 7, it can be concluded that the parameter scale of the algorithm network model in this article is much lower than other network models, and the mAP50 value is 97.28%, which is 4.92% higher than YOLOv7 in the same series. This also indicates that the improved YOLOv7 algorithm model has high accuracy and detection speed in defect detection, achieving better detection results.

6 Conclusion

This article proposes an improved YOLOv7 based algorithm for detecting surface weld defects in engine exhaust pipes, which is characterized by small size, complex background, complex weld defects, and low contrast. Firstly, based on the analysis of the characteristics of the weld seam, a visual system was established, and the process of building the visual system model was elaborated in detail. Then, by introducing CA attention mecha-

nism in the algorithm model, an improved upsampling structure and an improved one were used to improve the feature extraction ability for small defects and reduce feature information loss. Secondly, in order to enhance the robustness of the algorithm in this article, the MP module is integrated into the module to improve the performance of the algorithm. Multiple comparative experiments and ablation experiments have shown that the improved YOLOv7 object detection algorithm has an average accuracy value of 97.28%, which is at least 4 percentage points higher than the original YOLOv7 algorithm. The detection speed has also been improved, which meets the actual production needs.

During the research process, the shortcomings of this article were also identified, which will serve as the direction for further in-depth research.

1) The network structure is not lightweight enough. In order to improve algorithm speed and reduce dependence on hardware device performance, further research direction is to replace YOLOv7 with YOLOv7 tiny as the basic model.

2) In order to improve the stability of the algorithm, improvements have been made in the selection of the loss function. When the IoU value between the predicted bounding box and the true bounding box is low, the gradient of the loss function may be small, resulting in a slower training process.

7 Acknowledgement

This work has been supported by Scientific and Technologial Innovation Programs of Higher Education Institutions in Shanxi (2022L534, 2022L533).

References

[1] M.-Z. Huang, W.-S. Zhang, X.-Q. Huang, S.-J. Lin, S.-M. Pang, R.-Y. Lu, Fracture analysis and optimization of vehicle exhaust system, Modern Machinery (3)(2020) 26-29.

[2] L. Chen, Intelligent Recognition of Small Diameter Pipe Girth Weld DR Image Defects based on Deep Learning, Chemical Equipment Technology 43(4)(2022) 30-35.

[3] L.-C. Xiong, H. Yao, Z. Zhang, Q.-H. Du, J. Liu, J. Du, Research on automatic identification method of dome electron beam weld defects based on DR, China Measurement & Test 49(S1)(2023) 91-96.

[4] F.-Q. Zhao, L. Sheng, Z.-Y. Niu, S.-Y. Wu, C.-J. Liang, Defect Identification Method of Weld Inspection Based on LBP and SVM, Welded Pipe and Tube 45(6)(2022) 33-38.

[5] Y. Zhang, Y.-F. Gao, H. Zhang, Path recognition of grid fillet welds in ship cabin based on binocular vision, Welding & Joining (7)(2022) 21-27.

[6] S.-Q. Wang, J.-Z. He, Y.-D Gao. X.-M. Liu, Feature Extraction Method of Intersecting Line Weld Based on Vision Sensor, Journal of Shenyang University of Chemical Technology 36(4)(2022) 368-375.

[7] X.-H. Hang, H. Wang, Research on weld image recognition method based on deep learning, Wireless Internet Science and Technology 20(24)(2023) 126-132.

[8] W.-K. Xiao, Y.-J. He, Similarity Detection of Recognition Based on Deep Learning, Microcomputer Applications 38(12)(2022) 124-127+135.

[9] Z.-Y. Yu, H.-Q. Yuan, X.-L. Wei, G.-F. Du, Defect Identification Method of Pipeline Weld Ultrasonic Testing Based on Deep Learning, Science Technology and Engineering 22(30)(2022) 13288-13292.

[10] L. Long, H. Chen, H. Liang, S. Zhao, Z. Liu, Z.-T. Li, Real-time X-ray Weld Defect Detection Based on Lightweight YOLO Network, Network New Media Technology 12(2)(2023) 30-38.

[11] H.-B. Ma, J.-B. Zhang, X.-W. Yang, F. Liang, Application of Ultrasonic Image Recognition Based on YOLOv5 in Rail Flaw Detection in Plateau and Cold Areas, China Railway(9)(2023) 85-89.

[12] Q. Sun, S.-H. Zheng, Method of Plant 3D Reconstruction Based on Binocular Vision and Its Application, Journal of Anhui Agricultural Sciences 49(24)(2021) 11-17.

[13] Y. Zhao, Y. Xu, Visual Interaction Simulation of Depth Information of Stadium Based on Binocular Fusion, Computer Simulation 40(11)(2023) 202-206.

[14] Z.-Y. Zhang, A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11)(2000) 1330-1334.

[15] L.-K. Lin, Q.-Q. Niu, Y. Li, Optimized Camera Parameter Calibration Method Based on Checkerboard, Computer Technology and Development 33(12)(2023) 101-105.

[16] J.-J. Ji. J. Wang, Y.-J. Chen, D.-H. Lu, Detection of Minor Defects on the Surface of Hydraulic Valve Block Based on Improved YOLOv7, Computer Engineering 49(11)(2023) 302-310.

[17] X.-B. Han, Q.-S. Hu, X.-F. Zhao, Q. Qiu, Research on sign language recognition algorithm based on improved YOLOv7-tiny, Modern Electronics Technique 47(1)(2024) 55-61.
[18] Y. He, M.-K. Lu, S. Gao, B. Cao, J.-P. Lu, Distracted Behavior Detection of Commercial Vehicle Drivers Based on the Mobile ViT-CA Model, China Journal of Highway and Transport 37(1)(2024) 194-204.
[19] H. Leng, J.-X. Xia, Metal surface defect detection method based on improved YOLOv7, Computer Era (9)(2023) 48-53+58.
[20] P.-H. Gui, T. Song, J.-B. Tang, Z.-P. Xu, S.-X. Cao, Q. Jiang, Surface Defect Detection Algorithm of Micro-Channel Aluminum Flat Tube Based on Improved FCOS Model, Computer Engineering and Applications 59(24)(2023) 298-308.
[21] F.-S. Shu, Z.-B. Xu, Y.-C. Bao, improved YOLOv7 garment stitch fault detection method based on attention mechanism, Wool Textile Journal 52(1)(2024) 107-115.