Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 2520933, 15 pages
https://doi.org/10.1155/2023/2520933

Research Article

CBAM-YOLOv5: A Promising Network Model for Wear Particle Recognition

Lei He[1,2] **and Haijun Wei**[1]

[1]*Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China*
[2]*Hefei University of Economics, Hefei 230031, China*

Correspondence should be addressed to Haijun Wei; haijun_welson@163.com

Received 27 November 2022; Revised 19 January 2023; Accepted 31 January 2023; Published 7 June 2023

Academic Editor: Danfeng Hong

The intelligent recognition technology for ferrography images is one of the important methods for diagnosis fault and state detection of machines. In allusion to these questions for the influences of wear particle images' blurring, background intricacy, wear particle overlapping and lack of light, and others which lead to be the reason for the difficulty of achieving accurate identification, missed detection, and false detection, an intelligent recognition algorithm for ferrography wear particle based on convolutional block attention module (CBAM) and YOLOv5 is proposed. Firstly, it needs enhancement to improve contrast for ferrography wear particle images and lower background interference by adaptive histogram homogenization algorithm. Then, under the framework of YOLOv5 algorithm, the depthwise separable convolution is added to improve the detection speed of the network, and the detection accuracy of the entire network is improved by optimizing the loss function. Moreover, increase weight ratio on wear particle in images by adding a convolution block CBAM model and increase feature representative capability in detection network with YOLOv5 algorithm detection network, which can improve detection accuracy for wear particle. Finally, compare the algorithm with the three classical homologous series object detection algorithm. The experimental results show that the detection accuracy of the model can reach 96.7%, and the detection speed is 32 FPS for the images with a resolution of 1280×720. It can be developed and applied to the fault diagnosis and condition monitoring of mechanical equipment.

1. Introduction

Fault diagnosis and condition monitoring of mechanical equipment is a technology for collecting, processing, and analyzing the information of the mechanical operation state [1]. Ferrographic analysis is an important part of equipment fault diagnosis, and the identification of wear particles is the core of ferrographic analysis. Ferrography image analysis technology is a technology to extract, classify, and observe the wear particles in the lubrication system and judge the lubrication condition, wear mechanism, and wear severity of the friction pair through its quantity, size, shape, and texture. Compared with other fault diagnosis technologies, it has the advantages of strong forward-looking, large detection range of abrasive particles, and direct reflection of the main wear mechanism [2, 3]. Ferrography image analysis technology currently lacks automation, and its application at this stage relies heavily on expert experience, and it is time-consuming and expensive [4, 5]. The above shortcomings restrict the large-scale promotion and application of this technology in the industry [6, 7]. This paper intends to study an intelligent recognition algorithm of ferrography wear particle image based on computer vision technology, to provide technical basis for intelligent and rapid recognition of ferrography wear particle and intelligent online monitoring system of lubricating oil.

Plenty of tribology experiments and practical applications show that different forms of wear and tear will produce different characteristics of wear particle. Apply computer vision technology, mathematical methods and other methods, and technology to grind grain shape, texture, and color of informationize quantification for characteristic

parameters and then use these feature parameters to train appropriate classification decision algorithm to realize intelligent recognition of wear particle [8, 9].

Many scholars have conducted in-depth research on this issue. Peng and Wang and Peng et al. used Inception-v3 to extract surface texture features of wear particle with high efficiency for wear particle classification. However, this model does not have the ability of target detection, and wear particle should be segmented first in the face of multiabrasive images [10, 11]. Zhang et al. proposed a convolution neural network model based on class center vector and distance comparison, which adopted two operations: point-to-point group convolution (PWConv) and channel shuffle. The number of model parameters is reduced, but the training speed is not significantly improved [12]. Peng and Wang used a relatively simple convolutional neural network FECNN to train with 420 wear particle images and obtained good accuracy but did not study the problem of wear particle segmentation under complex background [13]. Wang et al. have developed similar models to identify and classify cutting, spherical, and fatigue wear particle by combining CNN and SVM, but the samples used are still single-target images and the characteristics of spherical and cutting wear particle are more obvious and easy to be distinguished [14, 15]. Peng et al. proposed a method of ferrospectral image target detection and recognition based on YOLOV3, using batch normalization method instead of dropout method, which achieved good results when applied to small-sized and low-resolution wear particle images and increases the generalization ability, but still had some shortcomings such as low accuracy of similar wear particle classification and lack of ability to mark wear particle contour [16, 17]. Hong et al. proposed a novel convolutional neural network miniGCN. It trains large-scale GCNs in a more flexible minibatch fashion and can directly predict new input samples without retraining the network. Three fusion schemes, including additive fusion, elementwise multiplicative fusion, and concatenation fusion, were used to achieve better classification results in HS images [18]. Patel et al. proposed to design a novel feature descriptor involving multifeature fusion technology for human action recognition, which reduces the complexity of detection technology and has high detection speed and efficiency [19].

Patel et al. proposed a feature fusion technology to recognize human behavior based on the benchmark ASLAN data set and UCF11 data set, which has a good recognition accuracy [20]. Patel et al. and Bhatt et al. proposed an architecture common to any CNN, DBGC (dimension-based generic convolution block); provided a network that can intelligently select convolution kernels of various heights, widths, and depths; improved the accuracy of the results; and reduced the computational complexity [21, 22].

In recent years, researchers have been committed to target detection based on deep learning, so as to realize intelligent analysis of wear particle. This method can effectively solve the shortcomings of traditional target detection based on artificial features, such as low detection accuracy, vulnerable to environmental interference, and weak generalization ability. Deep learning target detection algorithms include Mask-R-CNN, Faster-RCNN, UIU-Net network, YOLO algorithm, and ORSIm detector. UIU-net is a new network for infrared small target detection, which embeds tiny U-Net into a larger U-Net backbone network, so as to realize multilevel and multiscale representation learning of objects. Compared with the YOLOv5 model, UIU-Net is more suitable for the application of accurate image segmentation. This model learns how to classify each pixel of an image into different object labels, ignoring feature extraction and learning [23]. Optical remote sensing image detector (ORSIm detector) integrates multiple channel feature extraction, feature learning, fast image pyramid matching, and enhancement strategies [24]. The ORSIm detector uses a novel air-frequency channel feature (SFCF) that combines the rotationally invariant channel feature constructed in the frequency domain with the original spatial channel feature (such as color channel and gradient amplitude) [25, 26]. Compared with YOLOv5, ORSIm detector pays more attention to feature extraction and learning as well as image pyramid matching and enhancement, ignoring image segmentation processing. He et al. and Ren et al. proposed a method based on Faster R-CNN to identify iron ferrography wear particle in gear boxes, which overcame the problem of wear particle crossing and could identify multiple wear particle with high accuracy but slow speed [27, 28]. An et al. proposed an intelligent segmentation and recognition method of ferrography wear particle based on Mask R-CNN, with a detection accuracy of 76.2% and good generalization ability, but the segmentation effect of overlapping wear particle is not ideal [29]. This kind of algorithm is mainly trained according to the specific position of the abrasive particles in the ferrography images, and the candidate region is extracted in advance. It overcomes the recognition problem caused by the abrasive particles crossing, but the recognition speed of the abrasive particles in overlapping and blurred images needs to be improved. Zhang et al. proposed a multitarget ferrography wear particle intelligent recognition algorithm based on the improved YOLO algorithm, which realized the recognition of multitarget wear particle under complex background but could not improve the recognition rate and detection speed of similar wear particles, especially layered wear particles [30]. This kind of algorithm is to target detection as a regression problem; direct end-to-end training of the network can ensure high detection accuracy and have good real-time performance [31]. However, such algorithms still have some shortcomings for wear particle image recognition with complex background, overlapping wear particles, and the lack of light and cannot meet the requirements of online monitoring temporarily.

In order to solve the problems of wear particle blurring, complex background, wear particle overlap, light effect, fewer ferrography pictures, and so on, which make it difficult to achieve accurate target detection, missed detection, and false detection in ferrography image, an intelligent recognition method of YOLOv5 ferrography image based on convolution block attention mechanism model CBAM (convolutional block attention module) is proposed. The main contributions and innovations of this paper are as follows:

(1) Adaptive histogram equalization is used to approach detect ferrography wear particle image to improve image contrast and image quality

(2) Aiming at the problems of complex background, overlap of wear particle, and illumination influence of wear particle in ferrography images, the convolution block attention model CBAM is introduced on the basis of the YOLOv5 detection network to enhance the weight ratio of the target, so as to improve the feature expression ability of the target to be detected and then improve the accuracy of abrasive particle recognition

(3) Accurately simplify the parameters of the original detection network model by introducing a depthwise separable convolution (DWConv). While it makes the network have high detection accuracy, this reduces the impact on the network detection speed after introducing the attention mechanism, thus improving the detection speed of the network model

(4) By using the method of optimizing the loss function to reduce the loss value of the network to speed up the convergence of the network, it can improve the detection accuracy of the whole detection network

2. Improved YOLOv5 Model Based on CBAM

2.1. YOLOv5 Algorithm Principle. The YOLOv5 model is mainly composed of four parts: input, backbone, neck, and prediction. Input generally consists of three parts: mosaic data enhancement, autolearning bounding box anchors, and adaptive image scaling [32]. Backbone consists of three parts: focus, cross-stage partial network (CSP), and spatial pyramid pooling (SPP). Among them, the focus structure slices the picture and obtains the sampled feature map under twice the information. The CSP structure is mainly designed to solve the problem of excessive computation in the process of reasoning from the perspective of the network structure design. The SPP layer is by way of maximizing the pooling of the characteristic layer after three convolutions and enlarges the receptive field of vision and enhances the nonlinear expression ability of the network. The neck part is a further optimization of the FPN structure to improve the speed of feature fusion and inference information transmission on a network. In the prediction part, the function GIOU loss (generalized intersection over union loss) is used, which is mainly used to evaluate the recognition loss of the target rectangular box. The overall network structure of the YOLOv5 algorithm is shown in Figure 1.

2.2. Based on the Improved YOLOv5 Wear Particle Detection Method. The detection framework of the algorithm proposed in this experiment is shown in Figure 2. Firstly, the size of the wear particle image obtained by the oil detection was adjusted to 640×640, and dealing with adaptive histogram equalization was performed. Then, it is sent to the designed detection network for training, so as to obtain the training weight of the detection model. Finally, the test data is used to verify the proposed detection network.

2.3. Detection Network Fused with Convolution Block Attention Model. In order to solve the problem of low significance to be detected caused by complex background and overlap of wear particle in ferrography images, channel and spatial convolution block attention model are introduced after the CSP module of the YOLOv5 network model, whose structure is shown in Figure 3.

As shown in Figure 3, M_c represents channel attention in the convolution block attention model and M_s represents spatial attention. Given a feature map $F \in R^{C\times H\times W}$, where R is the number of channels of the C feature map, $H \times W$ represents the size of the feature map. The CBAM module will initially send F into the channel attention module and at the same time use the average pooling method and the maximum pooling method to obtain the information of all channels, and finally, the obtained parameters are superimposed by the multilayer perceptron and then activated by the Sigmoid function, the channel attention characteristics $M_c(F)$ are obtained, and its calculation formula is shown below:

$$\begin{aligned} M_c(F) &= \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{\mathrm{avg}}\right)\right) + W_1(W_0(F^c_{\mathrm{max}}))\right), \end{aligned} \tag{1}$$

where $\sigma(\bullet)$ represents Sigmoid function; MLP is a multilayer perceptron; AvgPool$(\bullet)$ and MaxPool$(\bullet)$, respectively, represent the operation of average pooling and maximum pooling of the spatial information of the feature map by the module; and F^c_{avg} and F^c_{max}, respectively, represent the global average pooling and maximum average pooling operations of the channel attention mechanism. After the given feature F_X is sent into the spatial attention module, spatial information is gathered along the channel dimension through average pooling and maximum pooling; then, the spatial feature map $F_{\mathrm{avg}} \in R^{1\times H\times W}$ and $F_{\mathrm{max}} \in R^{1\times H\times W}$ are generated. Then, the spatial attention feature is obtained through 1×1 convolution and Sigmoid function activation. Then, multiply with each F_X element to get the spatial attention feature map F_s. The calculation formula is as follows:

$$F_s = \sigma\left(\mathrm{Conv}\left(\mathrm{Cat}\left(F_{\mathrm{avg}}, F_{\mathrm{max}}\right)\right)\right) \otimes F_X = \sigma\left(f^{7\times 7}\left(\left[F^s_{\mathrm{avg}}; F^s_{\mathrm{max}}\right]\right)\right), \tag{2}$$

where Cat indicates connection operation, $f^{7\times 7}$ represents a 7×7 convolution operation of size, and F^s_{avg} and F^s_{max} represent global average pooling and maximum average pooling operations of spatial CBAM, respectively. YOLOv5 has no attention preference during feature extraction and uses the same weighting method for features of different importance.

In this paper, the original network has no attention preference problem by introducing the CBAM module after the CSP module, so that the network can pay more attention to the target of interest during the detection process [33]. The

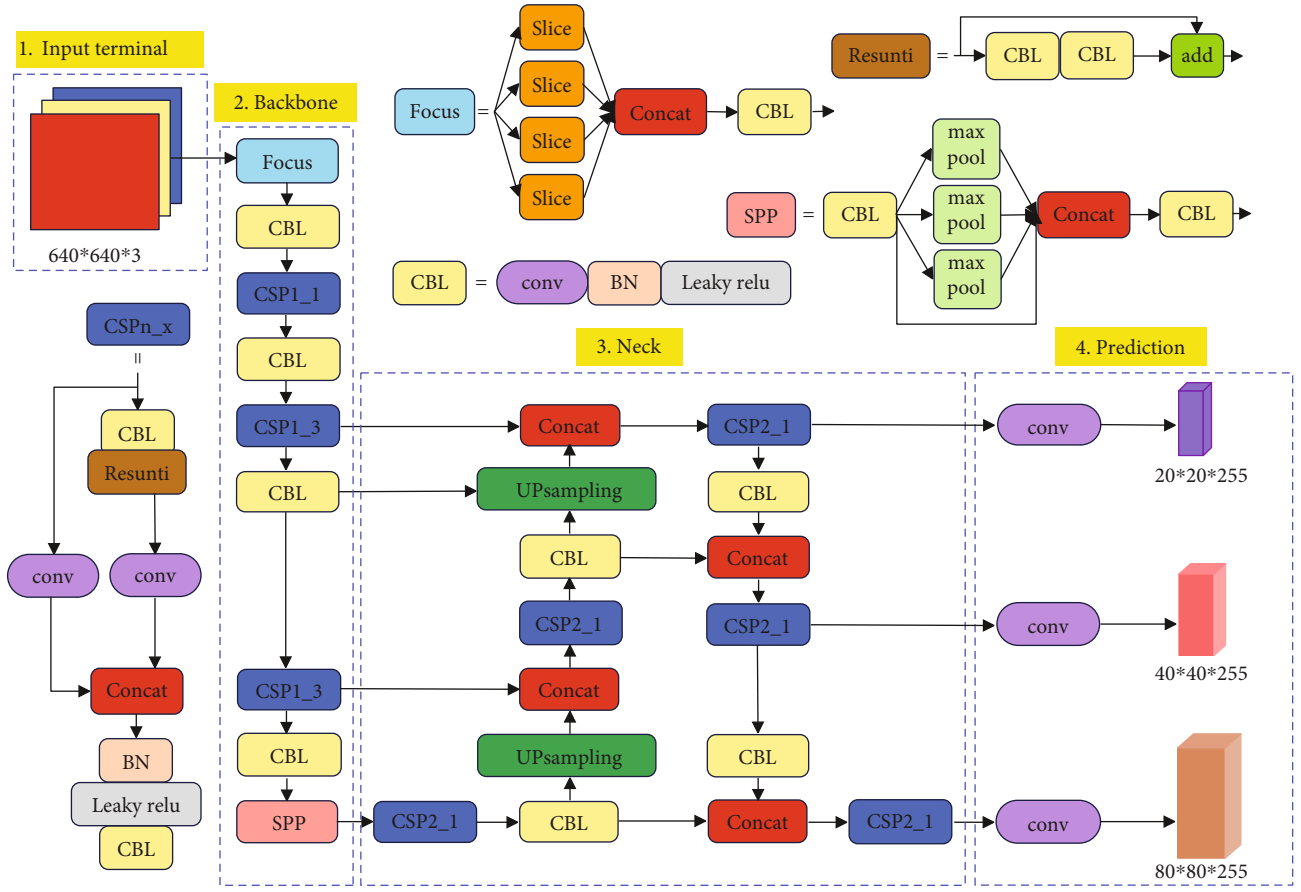


Figure 1: The YOLOv5 network structure diagram.

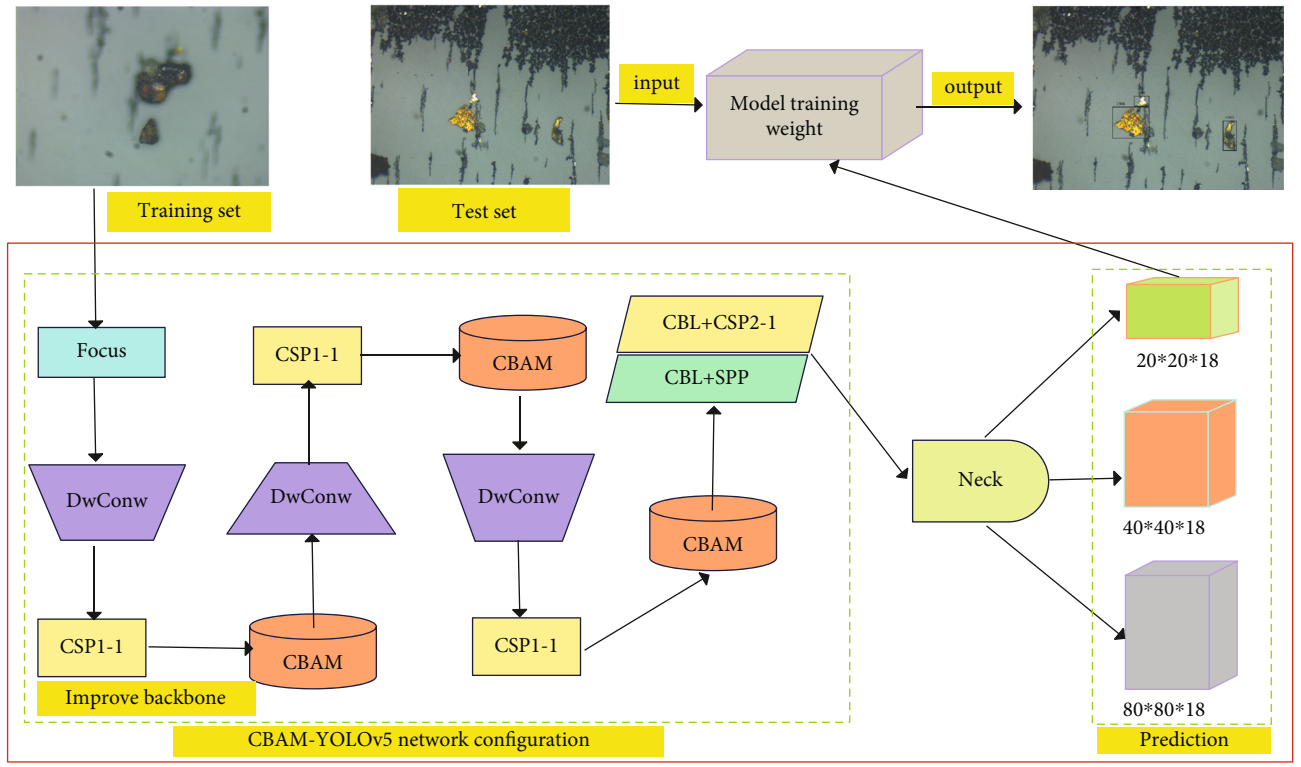


Figure 2: Detection framework of algorithm proposed in this paper.

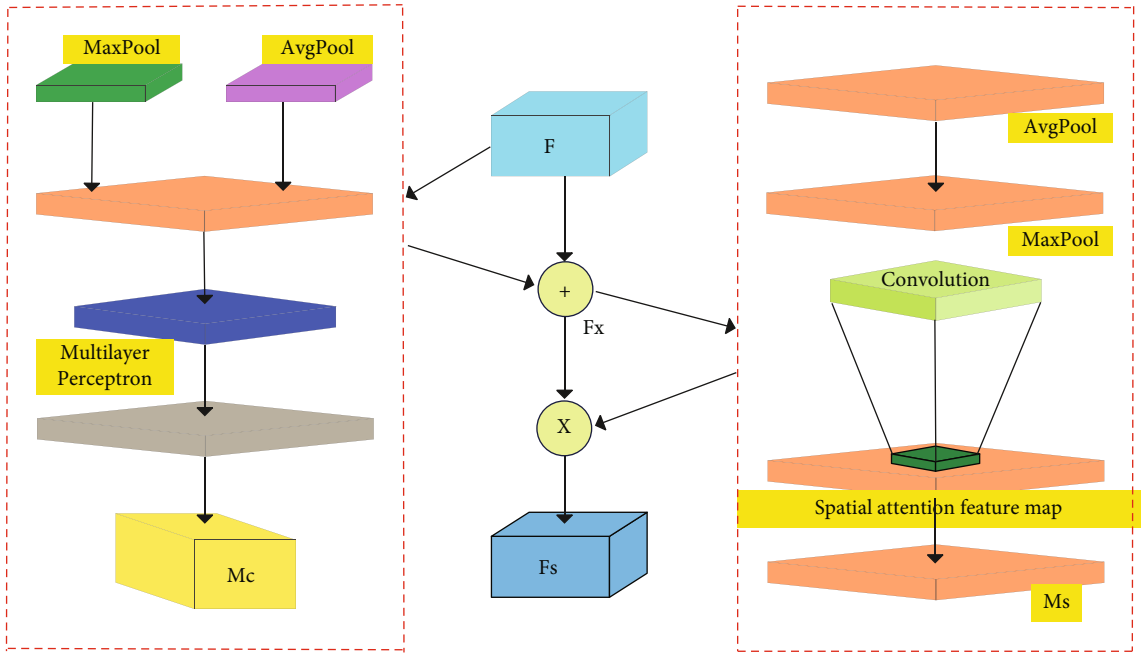


Figure 3: The CBAM module.

comparison results before and after the introduction of the CBAM model are shown in Figure 4. The red area in Figure 4 is the area with high salience, and the darker and clearer the color, the higher the salience.

As can be seen from the comparison of Figure 4, the addition of CBAM to the YOLOv5 model can enhance the saliency of the target to be detected in the complex background, which can lay a good foundation for the accurate detection of subsequent intelligent identification of abrasive grains. In addition, considering the accuracy and real-time nature of the detection network, three CBAM modules were finally introduced, as shown in Figure 2.

2.4. Depthwise Separable Convolution. In order to make the network have higher detection accuracy and reduce the impact on network detection speed caused by the introduction of attention mechanism, depthwise separable convolution (DWConv) is introduced to replace the original convolution operation in the backbone network, whose structure is shown in Figure 5.

As shown in Figure 5, M is the number of data input channels and N is the number of data output channels, D_x is the data input length, D_y is the data input width, D_k is the size of the convolutional kernel, and D_w is the data output length, which is the D_h data output width. The original convolution in the network is mainly to convolution the channel feature map; the calculation quantity Q_1 is as follows:

$$Q_1 = D_k^2 \cdot D_w \cdot D_h \cdot M \cdot N. \tag{3}$$

In Figure 5, a depthwise separable convolution splits the convolution operation into a 3×3 deep convolution and a pointwise convolution of 1×1. Suppose the input feature graph F is $M \times D_x \times D_y$. After deep convolution operation, the feature graph G of $N \times D_w \times D_h$ is obtained, and the calculation quantity Q_2 is as follows:

$$Q_2 = D_k^2 \cdot D_w \cdot D_h \cdot M + N \times M \times D_w \times D_h. \tag{4}$$

From formulas (3) and (4), it can be concluded that the ratio of the calculated quantity of the depthwise separable convolution to the standard convolution is as follows:

$$\frac{Q_2}{Q_1} = \frac{1}{N} + \frac{1}{D_k^2}. \tag{5}$$

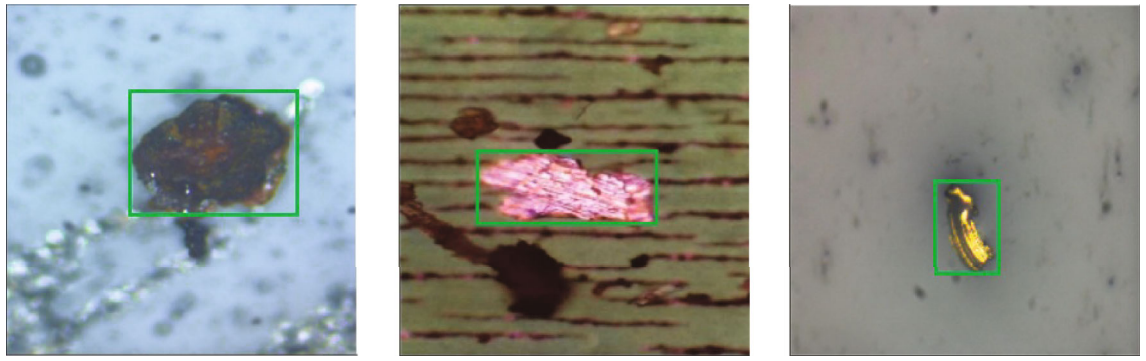
By introducing depthwise separable convolution, the calculation amount and parameters of the original network can be reduced; thus, the detection speed can be significantly increased [34]. This paper uses 3×3 convolution kernel, input channel 3, and output channel 256, which reduces the total network calculation to one eighth of that using standard convolution.

2.5. Optimize the Loss Function. Loss function can well reflect the difference between model and actual data. The bounding box regression loss in the YOLOv5 original network is calculated by GIOU function, and its calculation formula is shown in

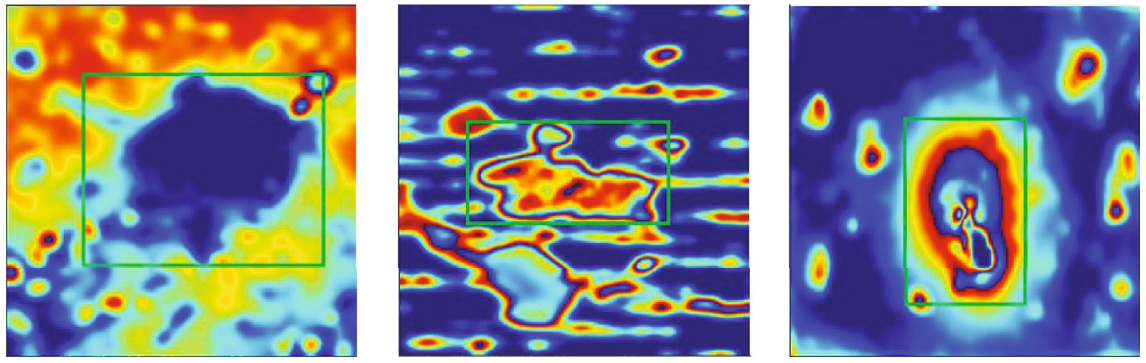
$$L_{\text{GIOU}} = 1 - \text{IOU} + \frac{\left|C - A \cup A^{gt}\right|}{|C|}, \tag{6}$$

where C is the smallest outer rectangle of two boxes and $A \cup A^{gt}$ is the union of two boxes.

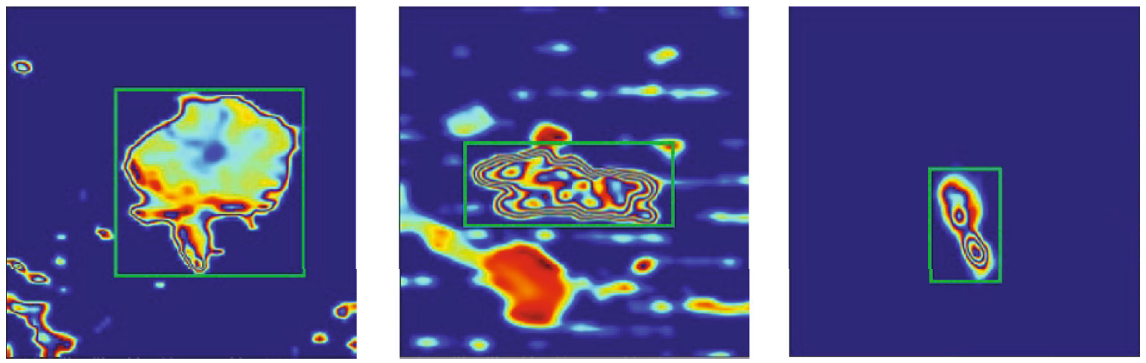
When the widest and highest aligns contained between the prediction box and the ground-truth box appear, the loss



(a) The original image



(b) Before adding CBAM



(c) After adding CBAM

Figure 4: Comparison results before and after adding CBAM.

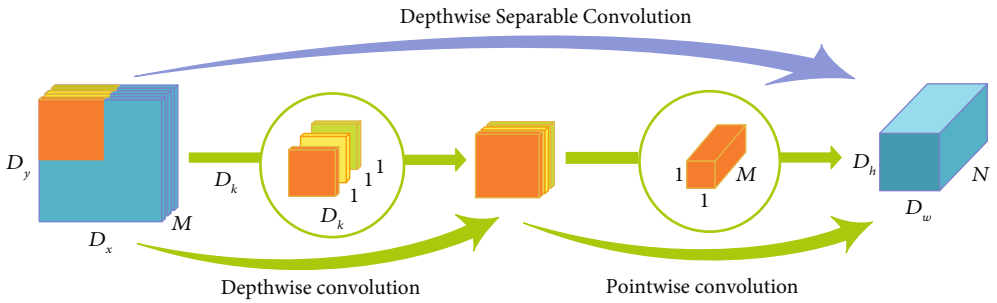


Figure 5: Depthwise separable convolution structure.

function degenerates into the IOU (intersection over union). At this time, it is impossible to evaluate the relative position of the prediction box and the ground-truth box, resulting in inaccurate target positioning, and the prediction box loses its convergence direction, affecting the detection accuracy. DIOU (distance intersection over union) takes into account

the distance, overlap rate, and scale influence between the prediction box and the ground-truth box, so as to ensure that the training process has a good convergence speed and is not easy to diverge. Thus, this paper uses DIOU replace of the original loss function to achieve a more efficient loss calculation between the ground-truth box and the prediction box, the calculation formula of which is shown in

$$L_{\mathrm{DIOU}} = 1 - \mathrm{IOU} + \frac{\rho^2(a, a^{gt})}{C^2}, \tag{7}$$

where a and a^{gt} represent the center points of the prediction box and the ground-truth box, respectively, ρ represents the Euclidean distance between two center points, and C is the slant distance covering the minimum rectangle between the prediction box and the ground-truth box.

2.6. Adaptive Histogram Equalization. Aiming at problems such as blurry and unclear ferrography images, the adaptive histogram equalization is used to process the data set, improve image contrast, enhance image quality, and optimize the image. Adaptive histogram equalization is mainly to divide the input image into small blocks and equalize the histogram of pixels within a rectangular range around each pixel, so as to balance the image gray scale and enhance the edge of the image. The number of pixels of the input image is counted as x and takes values in the range $[0, L-1]$, and L is the discrete gray level in the dynamic range. The histogram $p(r_k)$ of the input image can be represented as

$$p_r(r_k) = \frac{n_k}{x}, \quad k = 0, 1, 2, 3, \cdots L-1. \tag{8}$$

In the formula, r_k represents the k gray scale, which n_k represents the number of pixels present in the image. The grayscale cumulative distribution function S_k can be expressed as

$$S_k = \sum_{j=0}^{k} p(r_j). \tag{9}$$

The histogram equalization transformation function is

$$r_k' = \mathrm{round}((L-1), S_k), \quad k = 0, 1, 2 \cdots L-1. \tag{10}$$

In the formula, the gray level after the histogram r_k' is equalized and round(•) is rounded. The local effect comparison plot after the adaptive histogram equalization is shown in Figure 6.

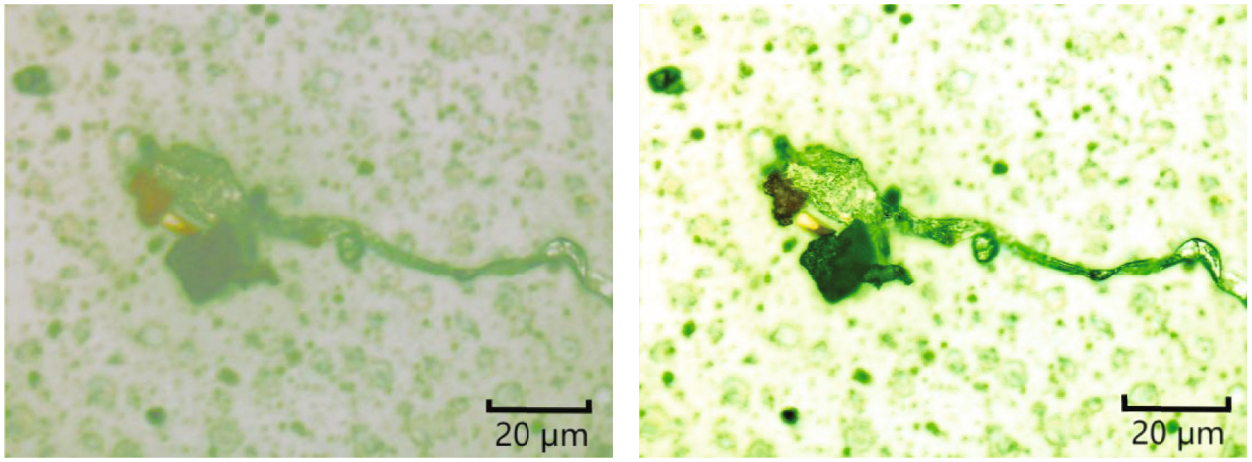
3. Experiments

3.1. Collection and Fabrication of Wear Particle Data Sets. Various wear particles are made using Bruker's UMT friction and wear test mechanism, as shown in Figure 7. The ambient temperature is 22°C, and the relative humidity is 50%. In the disc-pin experiment, the upper pin is a standard 416 stainless steel cylinder, the disc is alloy steel E52100, and the lubricant is Mobile Gard 412 lubricating oil, which lasts at a speed of 900 r/min at a load of 294 N for 25 h. The disc-pin friction experiment is mainly used to generate serious sliding wear particles and cutting wear particles. In the four-ball experiment, the material of the ball is GCr15 (hardness 63HRC), the maximum load and speed are set to 900 N and 300 r/min, respectively, and the experimental time is 50 h. Four-ball experiments are mainly used to produce fatigue wear particles, including spherical wear particles, fatigue wear particles, and laminar wear particles.

The prepared abrasive particles were passed through the SPECTRO-T2FM500 ferrography analyzer to make a spectrum piece, and then, the original wear pictures are taken by the microscope observer, and the equipment is shown in Figure 8. Since the size of the pictures taken by the optical microscope used in the experiment is 2568×1912, the resolution is too high, so these original wear particle images are rotated and cropped, and the data is expanded by OpenCV, 4867 640×640 wear particle image are obtained, 3880 are randomly selected as the training set, and 987 are the testing set. Finally, these wear particle images are labeled and classified by using the labeling tools and organized into VOC data set format [35, 36]. According to the generation mechanism and wear severity of the wear particle, the abrasive granules are divided into six categories, fatigue wear particle, layered wear particle, severe sliding wear particle, cutting wear particle, spherical wear particle, and oxidized wear particle, which can meet the needs of the network model for the number of training samples, and the data distribution of the abrasive granules label is shown in Table 1.

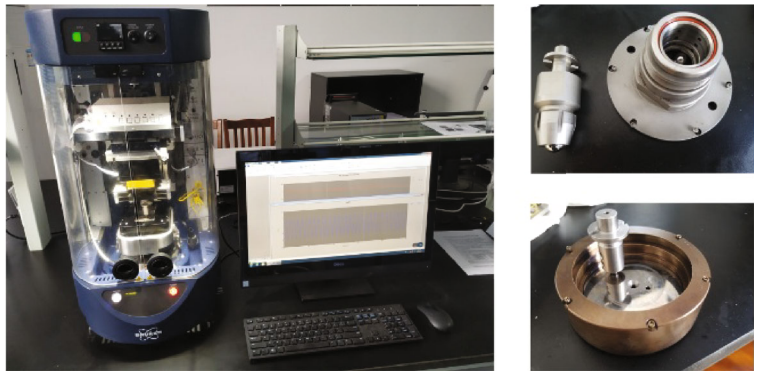
3.2. Experimental Platform and Model Training. This experiment builds a deep learning framework based on Ubuntu 18.04 LTS, Python 3.7.7, and PyTorch 1.6.0, and the main hardware parameters are as follows: GPU is NVIDIA GeForce GTX 1660Ti and the CPUs are Intel Core i5-10400F @2.90 GHz CPUs, CUDA 10.2, and CuDNN 7.6.5. During model training, the momentum factor is set to 0.937 to avoid the model falling into the local optimal solution or skipping the optimal solution. Set the learning rate for the first 300 rounds of network training to 0.01 and the learning rate for the last 200 rounds of training to 0.001. Set the weight decay regular term to 0.0005 to prevent overfitting of the network during training. Finally, after 500 rounds of iterative training of the model, the optimal model weight is obtained. The overall flowchart of this experiment is shown in Figure 9.

3.3. Evaluation Indicators. In order to verify the effectiveness and feasibility of this experimental detection model, evaluate from both qualitative and quantitative aspects [37], and for qualitative evaluation, the performance of the model will be evaluated by using the difference between the detection image of the algorithm in this paper and the control group algorithm, that is, comparing the positioning accuracy of the ground-truth box, and whether there is missing or false detection [38]. For quantitative evaluation, the main selected indicators are as follows: precision (P), recall (R), average

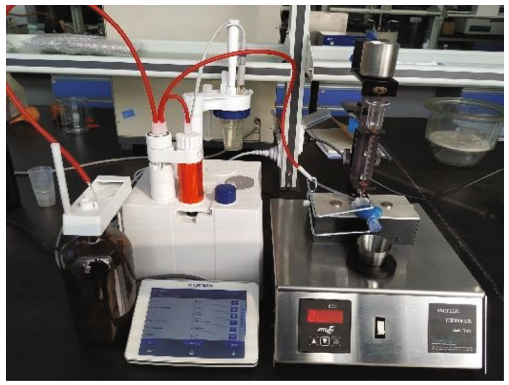


(a) The original image (b) The effect image after processing

Figure 6: Adaptive histogram equalization comparison diagram.



(a) Bruker Universal Mechanical Tester, four-ball module, and pin module



(b) SPECTRO-T2FM500 oil circulation system for simulating online collection of wear particles

Figure 7: Friction experimental equipment.

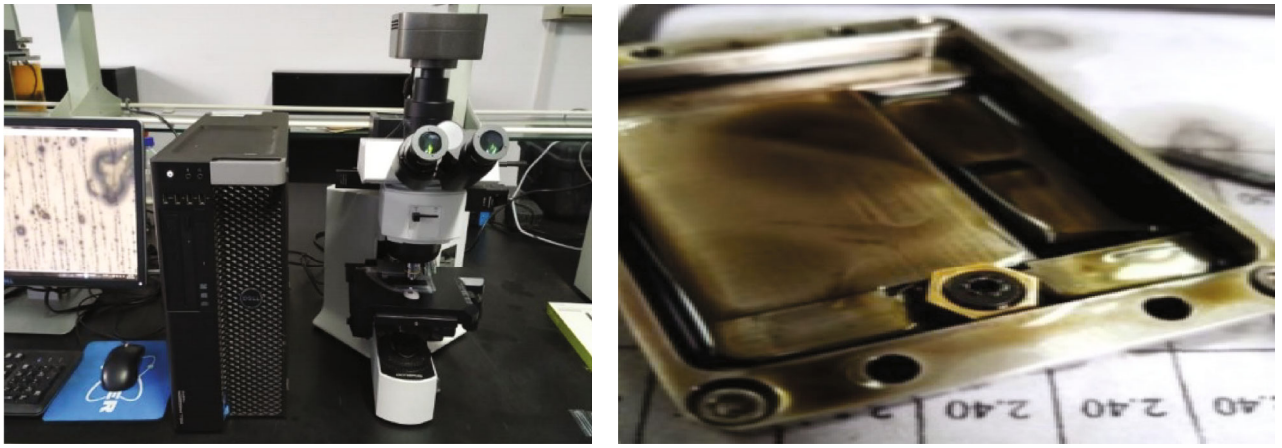
precision (AP), and mean average precision (mAP). The formulas are as follows:

$$P = \frac{T_P}{T_P + F_P} \times 100\%, \tag{11}$$

$$R = \frac{T_P}{T_P + F_N} \times 100\%, \tag{12}$$

$$\mathrm{AP} = \int_0^1 P(r)dr. \tag{13}$$

Taking the detected fatigue wear particles as an example, the formula T_P indicates the number of correctly recognized by the detection model, F_P represents the number of recognition errors or unrecognized, F_N represents the number of fatigue particle targets incorrectly detected as oxidation or slip, and r is taken as the parameter function $P(r)$. The



(a) Image collection of wear particles using Olympus BX51 (b) Residual lubricating oil in the oil tank after the experiment

Figure 8: Experimental equipment for the friction and wear part.

Table 1: Distribution of grinding label data.

Wear debris	Fatigue	Laminar	Sliding	Cutting	Spherical	Oxide
Training sets	896	793	580	551	634	426
Test sets	254	195	149	153	142	94

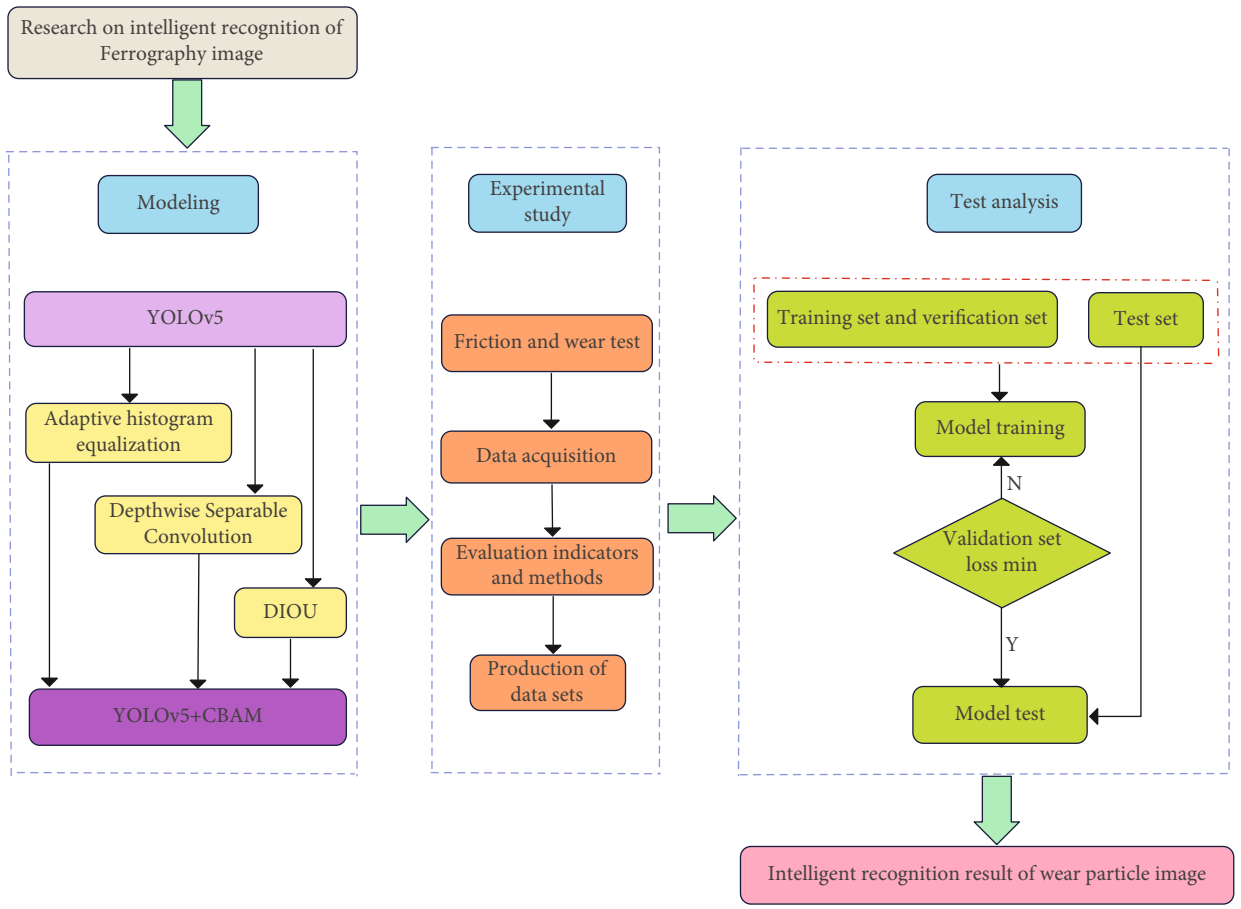


Figure 9: The overall flow chart of this experiment.

average precision (AP) indicates the average score of the accuracy rate to the recalled value, and the two indicators of P and R are generally used to judge the good or bad quality of the model. For each category of APs averaged as the average accuracy of mAP, mAP can measure the performance of the entire detection model.

4. Results and Analysis

4.1. Quantitative Analysis and Results. In order to more intuitively reflect the performance of the algorithm which is in the paper, the experimental algorithm is compared with the YOLOv5 object detection algorithm by randomly extracting images; in order to ensure the accuracy and rigor of the test, the two object detection algorithms are trained and tested on the same test platform, and the results are shown in Figure 10.

From Figure 10, it can be concluded that the YOLOv5 algorithm iterates to about 40 rounds before the accuracy rate rises to about 0.80 and finally stabilizes at about 0.84. However, after 40 rounds of iteration, the accuracy of the algorithm in this paper is about 0.956 and finally stabilizes at about 0.967. The comparison result of the mAP curve and the loss curve is shown in Figure 11.

From Figure 11(a), it can be concluded that the mAP curve of the proposed algorithm is above the mAP curve of the YOLOv5 detection algorithm; that is, the value of the mAP of the proposed algorithm is significantly higher than that of YOLOv5 networks. As can be drawn from Figure 11(b), the YOLOv5 loss gradually decreased to about 0.04 after 50 iterations and finally stabilized at about 0.041. After the introduction of the CBAM module, the initial loss of the network is about 0.099 and finally stabilizes at about 0.055. After optimizing the loss function, the network loss value is significantly reduced and the convergence speed is accelerated, and the initial loss value is about 0.085 and finally stabilizes at 0.033, which can be summarized that the detection model proposed in this paper has achieved good training effect.

4.2. Ablation Experiment. Based on the original YOLOv5 detection framework, the algorithm in this paper carried out adaptive histogram equalization, introduced CBAM and depthwise separable convolution, respectively, and optimized the loss function. In order to comprehensively analyze the advantages of various improved modules in CBAM-YOLOv5 for abrasive wear particle detection, ablation experiments were designed based on the original YOLOv5. Based on the original algorithm as the control group, the specific experimental content and test results are shown in Table 2. Table 2 analyzes the contribution of each improvement strategy to the network in this paper. It is found from the experiment that each module improves the overall performance of the model to different degrees.

In model 2, an adaptive histogram equalization module is added to the original network to improve the local contrast of the image, obtain more details of the image, and reduce the image blur interference. Compared with the data of model 1, it is easy to find that the introduction of the adaptive histogram equalization module successfully increases the detection accuracy by 1.81%, the recall rate by 1.78%, the average accuracy by 1.07%, and the detection speed does not change significantly.

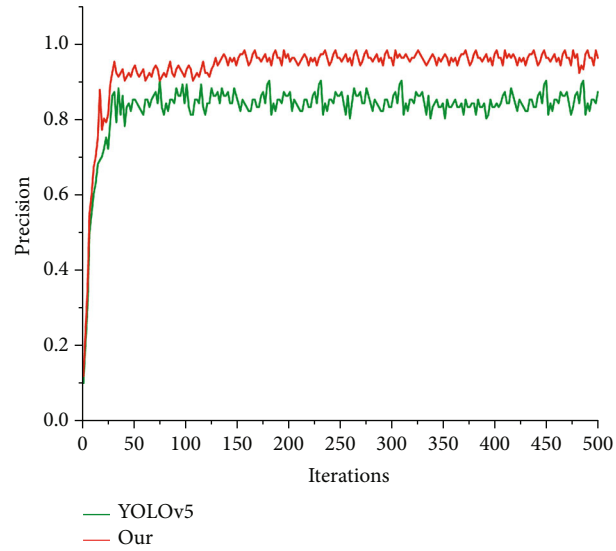


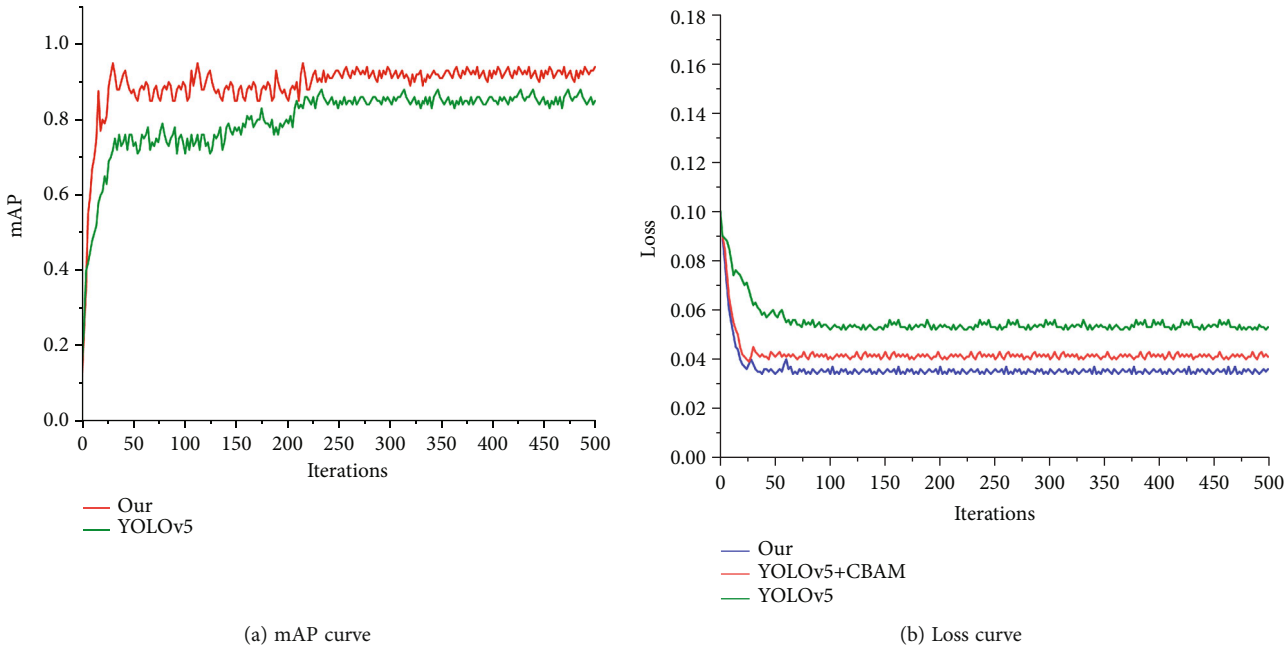
Figure 10: Accuracy change curve.

In model 3, CBAM was introduced into the backbone network to enhance the weight ratio of the abrasive region in the wear particle image, highlight the abrasive features, improve the feature expression ability of detection wear particle targets in the complex background, and effectively solve the problem that the complex background of lubricating oil image leads to the difficulty in feature extraction and the loss of network propagation feature information. The introduction of attention mechanism successfully optimizes the recognition performance of small targets. Compared with model 1, the detection accuracy is increased by 5.98%, recall rate by 6.17%, and average accuracy by 5.87%.

In model 4, after introducing depthwise separable convolution into the backbone network, the features of channel dimension and spatial dimension are mapped, respectively, and the results are combined. While retaining the learning ability of ordinary convolution representation, the number of parameters is reduced and the operational efficiency is improved. Compared with the original network, the detection accuracy is only improved by 0.09%, but the detection speed is improved by 23.95 FPS, in order to meet the new model more in line with the real-time and concise object detection needs in industry.

In model 5, the convolution module of model 3 is replaced by the depthwise separable convolution module, and its detection accuracy does not change greatly, but its speed increases by 17.16 FPS. The introduction of this module can reduce the amount of computation, obtain more characteristic information, and reduce the influence of the introduction of attention mechanism on the network detection speed.

After the loss function was changed to DIOU-LOSS in model 6, the distance, overlap rate, and scale effects between the prediction frame and the target frame were taken into



(a) mAP curve

(b) Loss curve

Figure 11: Evaluation indicators.

Table 2: Results of ablation experiments.

Model	Baseline network	AHE	CBAM	DWConv	DIOU-LOSS	*P*	*R*	mAP	*t*/*s*	FPS
1	YOLOV5					0.886	0.843	0.8376	0.029	34.87
2	YOLOV5	√				0.902	0.858	0.8468	0.029	34.87
3	YOLOV5		√			0.929	0.842	0.8663	0.038	26.32
4	YOLOV5			√		0.894	0.875	0.8675	0.017	58.82
5	YOLOV5		√	√		0.928	0.839	0.8673	0.023	43.48
6	YOLOV5				√	0.910	0.844	0.8415	0.034	29.41
7	YOLOV5	√	√	√	√	0.967	0.916	0.9262	0.031	32.26

account, so as to ensure a good convergence speed in the training process while not diverging easily. Based on the original network, the detection accuracy was increased by 2.71%, the recall rate by 3.68%, and the average accuracy by 5.24%.

By comparison of ablation experiments, it is found that the performance improvement of model 7, namely, CBAM-YOLOv5 proposed in this paper, is the most significant after the addition of various improved modules. The proposed algorithm synthesizes the advantages of each module, and the detection accuracy reaches 96.7%, the recall rate 91.6%, and the average accuracy 92.62%. It has greatly improved the problem of missing and misdetecting overlapping wear particles and fine wear particles, and the six kinds of detection targets have achieved good detection results. Based on the original network, the detection accuracy is increased by 9.14%, the recall rate is increased by 7.3%, the average accuracy is increased by 10.58%, and the average detection accuracy value is increased by 6.42%, which verifies the effectiveness of the proposed algorithm on the identification of wear particles. And the detection speed can reach 32 FPS, with good real time.

4.3. Qualitative Analysis and Results. In order to verify the advantages of this algorithm, four kinds of images with typical image blur, wear particle cross-overlapping, complex background, and lighting influence are selected for test verification. For ease of analysis, the objects to be inspected in the figure have been marked with different colored wireframes. The three classic detection algorithms of YOLOv3, YOLOv4, and YOLOv5 are selected and compared with the detection algorithm proposed herein, and the results after detection are shown in Figure 12.

Figure 12(a) is an image of the original ferrography wear particle, Figures 12(b)–12(e) are a plot of the detection results of different algorithms, and the experimental results of the four groups from left to right are analyzed as follows:

From the first set of experiments, it can be concluded that the proposed algorithm enhances the image contrast after the adaptive histogram equalization operation and also improves the clarity of the image, which can reduce the interference caused by image blur to a certain extent, YOLOv3, YOLOv4, and YOLOv5 model algorithms have different degrees of missed detection, and this algorithm can effectively solve the missed detection behavior in the

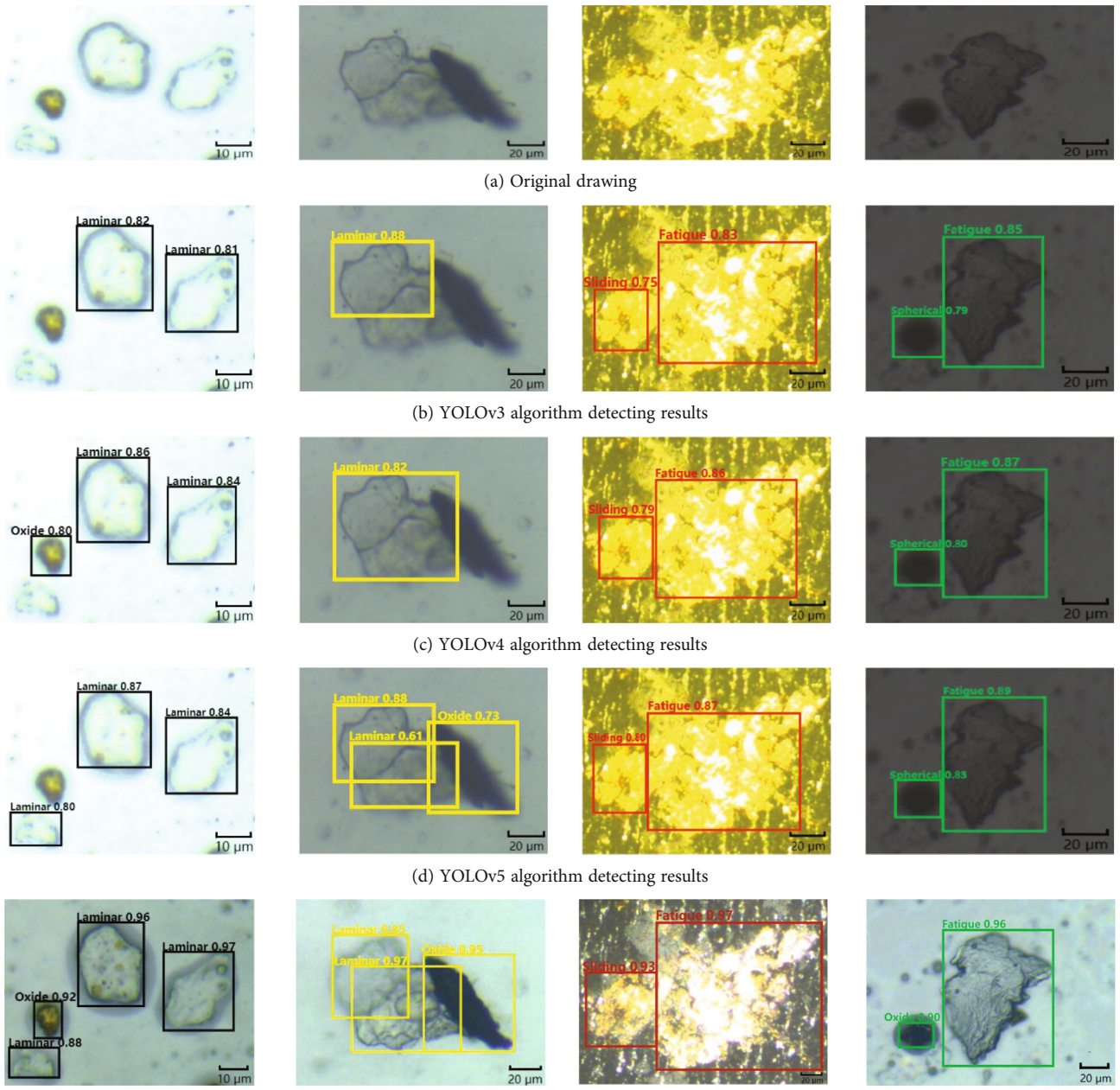


Figure 12: Comparison of the results of the different algorithms.

intelligent identification of wear particles, and the detection accuracy is as high as 0.92.

From the second set of experiments, it can be concluded that there is a strong overlapping interference of the wear particle and cannot be accurately distinguished. The YOLOv3 and YOLOv4 model algorithms have missed detection behavior for the fuzzy overlap of the wear particle. The YOLOv5 model can accurately identify the a wear particle type and effectively improve the missed detection behavior, but the detection accuracy is not high; the algorithm in this paper joins the CBAM module, by using the channel attention mechanism of this module to make the network automatically obtain the importance of each channel, and the information of each channel is superimposed with the multilayer perceptron sharing weight. The different weights of each channel given herein are used to strengthen the target characteristics, which can enhance the weight ratio of different wear particles, making the wear particle characteristics obvious. Then, the type of wear particle can be detected efficiently and with high precision, and there is without missed detection behavior.

From the third group of experiments, it can be concluded that the image background is complex and the wear particles are almost fused with the background. By comparing the

Table 3: Effect comparison of different models in the same series.

Model	P (%)	R (%)	T (s)	mAP (%)	Model size (MB)
YOLOv3	79.56	75.30	0.1420	82.10	236
YOLOv4	78.68	77.53	0.0851	84.70	156
YOLOv5	88.60	84.30	0.0516	83.76	20.5
Our	96.70	91.60	0.0313	92.62	25.5

images, it can be seen that all the tested algorithm models can accurately identify the wear particles, but the detection boxes of the YOLOv3, YOLOv4, and YOLOv5 model algorithms do not completely include the particles significantly smaller than the real frame; the detection frame of the algorithm in this paper is closer to the real size of the real frame. Not only that, for such complex background images, the proposed algorithm also has high detection accuracy.

From the fourth group of experiments, it can be concluded that the area brightness of the wear particle image to be detected is low due to the influence of illumination, especially the wear edge in the image is close to the background gray level. Model algorithms of YOLOv3, YOLOv4, and YOLOv5 have some degree of mischecking behavior, which makes oxidation wear particle mischecked into spherical wear particle. The algorithm in this paper can effectively solve the impact of light and can accurately identify wear particles with high detection accuracy.

It can be seen that the algorithm in this paper joins the CBAM module. By using the space attention mechanism in the CBAM module, all positions in the feature map are generated weights and output, which can enhance the target specific area while weakening the irrelevant background area, thus further enhancing the feature expression ability of the wear target, so that the target can be detected accurately. And under the operation of adaptive histogram equalization, it can greatly reduce the false detection and missed detection and improve the accuracy. Mean accuracy is also higher than other detection algorithms, which not only has high detection accuracy but also can solve more complex background, improve the detection rate of small wear particle and overlapping wear particle, and have strong robustness.

4.4. Comparison Experiment. In order to further verify the advantages of the algorithm in this paper, it is compared with three different model algorithms in the same series. To ensure the credibility of the experiment, all experiments are compared with objective data indicators in the same data set and the same training environment. The detection and comparison results are shown in Table 3.

It can be seen from Table 3 that the algorithm in this paper is significantly higher than the other three detection algorithms in terms of accuracy and average precision, and the detection speed is also significantly higher than that of the other three models. Although the size of the model is larger than YOLOv5, the recall rate is higher than that of other detection algorithms, which has obvious advantages in general.

5. Conclusions

The CBAM-YOLOv5 detection model is proposed to solve the problems of ferrography wear particle image blurring, complex background, wear particle overlapping, and illumination influence in the condition monitoring and fault diagnosis technology of mechanical equipment. Through a series of friction and wear experiments to collect ferrography wear particle images for intelligent detection experiments, the following conclusions are obtained:

(1) Adding CBAM attention mechanism to the YOLOv5 detection model can effectively solve the problems such as ferrography image blurring, wear particle overlapping, complex background, and lack of light influence, which lead to weak target saliency, difficult wear particle detection, missing detection, and false detection. Using adaptive histogram equalization to preprocess the image can effectively reduce the interference of wear particle recognition caused by ferrography image blurring and improve the quality of the data set. The detection speed of the network is improved by introducing the depthwise separable convolution, and the detection accuracy of the network can be improved by optimizing the loss function

(2) Through a large number of experimental results, it is proved that the accuracy of the algorithm in this paper can reach 96.7% for images with a resolution of 1280×720, the average accuracy is 92.62%, and the detection speed is 32 FPS, superior to YOLOv3, YOLOv4, and YOLOv5 algorithms

(3) Due to the limitation of the number of experimental data sets, the accuracy of the detection results in the experimental results is limited. However, with the expansion of the data sets, the detection accuracy of the algorithm proposed in this paper will be further improved. In consideration of both speed and accuracy, it has high application value and provides important reference and theoretical and practical basis for the subsequent intelligent fast identification of ferrography wear particles and the intelligent mechanical equipment oil online monitoring system

In the future research, we will take online monitoring technology as the center to design a deep learning model with higher accuracy, smaller model, faster speed, and stronger generalization ability and deploy it into the online oil analysis system, so as to solve the problem of online

intelligent recognition in the field of ferrographic image recognition to the greatest extent.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This work was supported by the Shanghai Engineering Research Center of Intelligent Ship Operation and Energy Efficiency Monitoring, Shanghai Science and Technology Program (20DZ2252300) and Key Project of Natural Science Foundation of Anhui Province (KJ2020A1180).

References

[1] T. H. Loutas, D. Roulias, E. Pauly, and V. Kostopoulos, "The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 25, no. 4, pp. 1339–1352, 2011.

[2] J. M. Wakiru, L. Pintelon, P. N. Muchiri, and P. K. Chemweno, "A review on lubricant condition monitoring information analysis for maintenance decision support," *Mechanical Systems and Signal Processing*, vol. 118, pp. 108–132, 2019.

[3] M. D. Haneef, R. B. Randall, W. A. Smith, and Z. Peng, "Vibration and wear prediction analysis of IC engine bearings by numerical simulation," *Wear*, vol. 384-385, pp. 15–27, 2017.

[4] R. K. Upadhyay, "Microscopic technique to determine various wear modes of used engine oil," *Journal of Microscopy and Ultrastructure*, vol. 1, no. 3, pp. 111–114, 2013.

[5] A. Kumar and S. K. Ghosh, "Size distribution analysis of wear debris generated in HEMM engine oil for reliability assessment: a statistical approach," *Measurement*, vol. 131, pp. 412–418, 2019.

[6] W. Cao, G. Dong, Y. B. Xie, and Z. Peng, "Prediction of wear trend of engines via on-line wear debris monitoring," *Tribology International*, vol. 120, pp. 510–519, 2018.

[7] F. Xie and H. J. Wei, "Research on controllable deep learning of multi-channel image coding technology in Ferrographic Image fault classification," *Tribology International*, vol. 173, no. 107656, 2022.

[8] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020.

[9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[10] P. Peng and J. Wang, "Wear particle classification considering particle overlapping," *Wear*, vol. 422-423, pp. 119–127, 2019.

[11] Y. Peng, J. Cai, T. Wu, G. Cao, N. Kwok, and Z. Peng, "WP-DRnet: a novel wear particle detection and recognition network for automatic ferrograph image analysis," *Tribology International*, vol. 151, article 106379, 2020.

[12] T. Zhang, J. Hu, S. Fan, and Y. Yu, "CDCNN: a model based on class center vectors and distance comparison for wear particle recognition," *IEEE Access*, vol. 8, pp. 113262–113270, 2020.

[13] P. Peng and J. Wang, "FECNN: a promising model for wear particle recognition," *Wear*, vol. 432-433, article 202968, 2019.

[14] S. Wang, T. H. Wu, T. Shao, and Z. X. Peng, "Integrated model of BP neural network and CNN algorithm for automatic wear debris classification," *Wear*, vol. 426-427, pp. 1761–1770, 2019.

[15] S. Wang, T. Wu, P. Zheng, and N. Kwok, "Optimized CNN model for identifying similar 3D wear particles in few samples," *Wear*, vol. 460-461, article 203477, 2020.

[16] Y. Peng, J. Cai, T. Wu et al., "A hybrid convolutional neural network for intelligent wear particle classification," *Tribology International*, vol. 138, pp. 166–173, 2019.

[17] Y. Peng, J. Cai, T. Wu et al., "Online wear characterisation of rolling element bearing using wear particle morphological features," *Wear*, vol. 430-431, pp. 369–375, 2019.

[18] D. Hong, L. Gao, J. Yao, B. Zhang, P. Antonio, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2021.

[19] C. I. Patel, D. Labana, S. Pandya, K. Modi, H. Ghayvat, and M. Awais, "Histogram of oriented gradient-based fusion of features for human action recognition in action video sequences," *Sensors*, vol. 20, no. 24, p. 7299, 2020.

[20] C. I. Patel, S. Garg, T. Zaveri, A. Banerjee, and R. Patel, "Human action recognition using fusion of features for unconstrained video sequences," *Computers and Electrical Engineering*, vol. 70, pp. 284–301, 2018.

[21] C. Patel, D. Bhatt, U. Sharma et al., "DBGC: dimension-based generic convolution block for object recognition," *Sensors*, vol. 22, no. 5, p. 1780, 2022.

[22] D. Bhatt, C. Patel, H. Talsania et al., "CNN variants for computer vision: history, architecture, application, challenges and future scope," *Electronics*, vol. 10, no. 20, p. 2470, 2021.

[23] W. Xin, D. Hong, and J. Chanussot, "UIU-Net: U-Net in U-Net for infrared small object detection," *IEEE Transactions on Image Processing*, vol. 32, pp. 364–376, 2023.

[24] D. Hong, L. Gao, N. Yokoya et al., "More diverse means better: multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, 2021.

[25] D. Hong, N. Yokoya, G.-S. Xia, J. Chanussot, and X. X. Zhu, "X-ModalNet: a semi-supervised deep cross-modal network for classification of remote sensing data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 12–23, 2020.

[26] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, "Learnable manifold alignment (LeMA): a semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 147, pp. 193–205, 2019.

[27] H. Beibei, C. Chenggang, and G. Weimin, "Ferrography wear particle recognition of gearbox based on faster R-CNN," *Lubrication Engineering*, vol. 45, no. 10, pp. 105–112, 2020.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[29] A. Chao, W. Haijun, and L. Hong, "Ferrographic wear debris intelligent segmentation and recognition based on mask R-CNN," *Lubrication Engineering*, vol. 45, no. 3, pp. 107–112, 2020.

[30] Z. Zhang, H. Wei, and H. Liu, "Intelligent recognition of multi-objective ferrographic wear particles based on improved YOLO algorithm," *Lubrication Engineering*, vol. 46, no. 5, pp. 27–33, 2021.

[31] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, Honolulu, HI, USA, 2017.

[32] Z. Weiguo and X. Yunxia, "Detection of dangerous driving behavior based on CG-YOLOv5," in *2022 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–6, Xi'an, China, 2022.

[33] H. Chang, P. Borghesani, and Z. Peng, "Automated assessment of gear wear mechanism and severity using mould images and convolutional neural networks," *Tribology International*, vol. 147, article 106280, 2020.

[34] S. Suh, J. Jang, S. Won, M. S. Jha, and Y. O. Lee, "Supervised health stage prediction using convolutional neural networks for bearing wear," *Sensors*, vol. 20, no. 20, p. 5846, 2020.

[35] X. Xu, Z. Tao, W. Ming, Q. An, and M. Chen, "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," *Measurement*, vol. 165, article 108086, 2020.

[36] H. Wu, N. M. Kwok, S. Liu, R. Li, T. Wu, and Z. Peng, "Restoration of defocused ferrograph images using a large kernel convolutional neural network," *Wear*, vol. 426-427, pp. 1740–1747, 2019.

[37] L. Xueyi, L. Jialin, Q. Yongzhi, and H. David, "Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning," *Chinese Journal of Aeronautics*, vol. 33, no. 2, pp. 418–426, 2020.

[38] J. Wang, P. Yao, W. Liu, and X. Wang, "A hybrid method for the segmentation of a ferrograph image using marker-controlled watershed and grey clustering," *Tribology Transactions*, vol. 59, no. 3, pp. 513–521, 2016.