IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 July 2022, accepted 21 July 2022, date of publication 25 July 2022, date of current version 3 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3193676

RESEARCH ARTICLE

Intelligent Detection Method of Forgings Defects Detection Based on Improved EfficientNet and Memetic Algorithm

TANG YU[1,4,5], WANG CHEN[1,2,4,5], GAO JUNFENG[3], AND HUA POXI[1,4,5]

[1]Department of Mechanical Engineering, Hubei University of Automotive Technology, Shiyan 442002, China
[2]Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Shanghai 200072, China
[3]Industrial Product Quality Inspection and Testing Institute, Shiyan 442002, China
[4]Hubei Zhongcheng Technology Industry Technique Academy Company Ltd., Shiyan 442002, China
[5]Shiyan Industry Technique Academy, Chinese Academy of Engineering, Shiyan 442002, China

Corresponding author: Wang Chen (20090011@huat.edu.cn)

This work was supported in part by the Humanities and Social Sciences Project of the Ministry of Education under Grant 20YJCZH150, in part by the Excellent Young and Middle-Aged Science and Technology Innovation Team Plan Project of Hubei Province Colleges and University under Grant T2020018, and in part by the Social Science Research Project of Hubei Provincial Department of Education under Grant 21Q17.

ABSTRACT In the process of production, automobile steel forgings are prone to various cracks, which affect the product quality. At present, forgings defects are mainly detected by fluorescent magnetic particle inspection and manual inspection. Aiming at the problems of low detection accuracy and efficiency in this method, an improved convolutional neural network model is proposed. The fluorescent magnetic particle inspection images of two typical forgings were intelligently inspected. Firstly, a deep learning model with EfficientNet as the backbone and Feature Pyramid Network (FPN) as the fusion layer is constructed. Secondly, in order to improve the convergence speed and detection accuracy, the calculation method of intersection over union is improved, and the network is improved by using the Attention Mechanism. Finally, Particle Swarm Optimization algorithm (PSO) with adaptive parameters is introduced to optimize the hyperparameters of neural network, and a fluorescent magnetic particle inspection image acquisition platform is built for verification. The mean Average Precision (mAP) of the best model of EfficientNet-PSO on the validation set is 95.69%. F1 score is 0.94 and FLOPs is 1.86B. Compared with other five deep learning neural network models, this method effectively improves the defect detection efficiency and accuracy of flange plate and cylinder head, which can meet the defect detection requirements.

INDEX TERMS Machine learning, industry applications, object detection.

I. INTRODUCTION

Automobile steel die forgings are widely used in automotive gearboxes, transmission systems, steering systems, front and rear axles, and engine interiors. Due to improper processes such as quenching, overheating, and overturning in the production process, there may be defects in the production process of forgings. In order to ensure the high quality of forgings produced by enterprises, it is necessary to carry out nondestructive testing on forgings. Currently, the fluorescent magnetic particle inspection method is used by company [1], [2] to detect defects by observing magnetic traces with the manual inspection. But this detection method is inefficient, expensive, and prone to errors when workers are overworked. Therefore, the computer vision technique to replace manual inspection in detection is one of the important trends in the current of industry [3], [4]. This technique is called object detection [5].

Object detection technology is mainly divided into two categories: traditional vision [6], [7] and deep learning [8]–[10]. The traditional machine vision method usually designs the corresponding feature template according to the

The associate editor coordinating the review of this manuscript and approving it for publication was Long Xu.

characteristics of defects, which has good accuracy and robustness, but the traditional machine vision is difficult to deal with scenarios with complex and changeable texture background. In addition, industrial production scenarios often need generalization models to cope with the adjustment of production demand. In this regard, traditional visual methods have natural disadvantages. Compared with traditional visual technology, there is more powerful feature extraction ability in deep learning-based detection. In particular, convolutional neural networks [11], [12] (CNN) are more suitable for object recognition in defect problem. CNN has a high level of fault tolerance, parallel processing, and self-learning ability, as well as the ability to operate with complex environmental data, ambiguous background knowledge, and ambiguous reasoning principles. Therefore, computer vision based on deep learning are more and more used in industry.

However, deep learning model training is difficult. Due to CNN network is called black box model, manual parameter adjustment can only be done by professionals, and the time cost of trial and error is high. In order to solve this problem, the use of hyperparameter optimization to improve deep learning is increasingly recognized [13], [14]. The optimization problem has the characteristics of ambiguous expression, nonconvex and high evaluation cost. Memetic algorithm is an effective global optimization algorithm [15]–[17].

An intelligent detection approach for forgings magnetic particle defect detection based on memetic computing is proposed in this paper to improve the CNN model. The main contributions of this paper are as following:

1) The real scene of forging production is simulated, the fluorescent magnetic particle image detection platform is built, and the defect sample data set is collected.

2) A CNN model based on improving EffficientNet backbone is proposed. There is a powerful feature extraction capability for complex high-level semantic features. For the application scenario of the model, Complete Intersection over Union (CIoU) and Attention Mechanism (AM) are innovatively introduced to improve the detection accuracy of the model. In order to improve the robustness of the model, the activation function is improved and Dropconncet is used.

3) Aiming at the difficulty of model parameter tuning, memetic computing is used to adjust parameters automatically. Based on the above dataset to validate the performance of six different deep learning models.

The rest of this paper is organized as following. In Sect.2. We review the literature closely related. In Sect.3 the details of model design are described. The details and parameters of the experiment is described in Sect.4. The performance of the model is analyzed in Sect.5. And some concluding remarks and future work are outlined in Sect.6.

II. LITERATURE REVIEW

To the best of our knowledge, since there exist a large number of existing works using convolutional networks for defect detection, but based on deep learning methods, there is few relevant literature for the research of magnetic particle inspection of forgings defect detection, so we review the closely related contributions.

A. DEFECT DETECTION BASED ON CNN

In recent years, CNN has been more and more widely used in the field of defect detection. Compared with traditional machine vision, CNN has the key advantage of automatically learning defect features. In the literature, Deep Learning-based industrial product defects, Park *et al.* [18] suggested a system for inspecting surface parts for dirties, scratches, burrs, and wears automatically. CNN is used to analyze sample images. Compared with traditional machine learning methods, it has better generalization performance. The technology outperforms manual detection in terms of both cost and efficiency, according to the findings of the experiments. Dai *et al.* [19] adopted an improved RCNN method for detecting surface defects in precision workpieces. The backbone of the system is ResNet101. The technique includes an object network extrusion and excitation (SENet) module, a channel feature fusion module, a feature pyramid network (FPN) module, and a ROI network module, which dramatically improves defect detection performance of CNN. A network model for PCB detection was presented by Shen *et al.* [20]. This model is named LD-PCB, the traditional model is difficult to handle complex components and a wide variety of problems are solved by LD-PCB. Therefore, the advantages of deep learning methods are highlighted. Tong *et al.* [21] proposed an integration of a fully convolutional network with a Gaussian-conditional random field (G-CRF), an uncertainty framework, and used for road defect detection, extraction of road location, type and other related information. This approach offered good accuracy and generalization capabilities, according to the test findings. And overcomes the shortcomings of traditional detection methods. Feng *et al.* [22] discussed a new object detection technique for detecting rail problems. The proposed network design of algorithm incorporates a MobileNet backbone network and numerous novel detection layers with multi-scale feature mappings. The results of the experiments reveal that the method has a fast inference speed, high precision, and a wide range of applications in industry.

Yu *et al.* [23] proposed a damage identification and location method of building structure based on deep convolution neural network. A variety of convolution kernels of different sizes are used to improve the performance of the detector, and LReLU and Dropout are used to improve the generalization ability of the detector, which will ultimately affect other machine learning methods. In conclusion, the effectiveness of CNN in defect detection has been widely proved. CNN training needs a large number of samples, but in the production detection process, the defect sample data is less, the image acquisition is time-consuming, and the labeling cost is high. When facing this problem, the most intuitive solution is to data augmentation. Du *et al.* [24] discussed the impact of data augmentation on network performance when trying to solve the problem of defect detection of automobile casting aluminum parts. Using a single data augmentation method to improve the network performance is limited. When the data increases to a certain amount, the network performance will decline. This means that we cannot blindly rely on data augmentation to solve the problem of small samples. Di *et al.* [25] proposed a semi-supervised deep learning-based steel surface fault classification system. The classification rate for hot rolled plates is increased by roughly 16 percent using the new CAE-SGAN method, which is based on Convolutional Autoencoder (CAE) and semi-supervised Generative Adversarial Networks (SGAN). CAE-SGAN can make full use of sample images of steel surface (labeled and unlabeled images), which improves the accuracy of defect classification with limited training samples. The deep learning model used for industrial detection is often aimed at a specific application scenario, while Ren *et al.* [26] proposed a new surface detection algorithm, which has certain universality in surface inspection tasks. In the small sample segmentation task, this method only needs 5 images for training, and the detection result reaches 0.0% error escape rate. In addition, it has the ability to automatically adapt to the complex industrial production scenarios with small samples and strong noise. This achievement brings some inspiration for the research of industrial detection model. In addition to the problem of the number of samples mentioned above, the small defect size is also a very difficult problem in industrial detection. To solve this problem, Hu and Wang [27] proposed a new method based on object-level attention mechanism, which not only reduces the dependence of the machine on accurate annotation, but also has a good effect in small object. Experiments show that this method can effectively realize real-time defect detection of castings in complex scenarios.

However, the above experience is mostly CNN networks with manual design and parameter adjustment. This method requires professionals to pay a lot of time cost.

B. NEURAL NETWORK HYPERPARAMETER OPTIMIZATION

In computer vision and object detection, with the improvement of production process and image complexity, evolutionary algorithms have been widely used in neural network structure optimization, image classification, and evolutionary multi-task image feature learning and other applications. By decreasing the connection parameters in deep networks, a model with greater generalization ability can be obtained. For feature selection and classification in data mining, Nekkaa and Boughaci [28] suggested a memetic method paired with support vector machine (SVM).The suggested strategy aims to identify the subset of features that improve SVM classification accuracy the most. Then SVM, as a shallow model, is far inferior to the deep learning model in feature extraction performance. However, combining the idea of optimizing SVM with meme algorithms also brings us some inspiration. Jia *et al.* [29] proposed a multi-objective optimization-based layerwise structure learning approach. This method can optimize each layer to find the structure with good generalization ability and high expression ability. Finally, an improved multi-objective memetic algorithm is designed to solve the model. Martin *et al.* [30] presented a new evolutionary technique called EvoDeep to modify the parameters and architecture of CNNs in order to maximize classification accuracy and preserved an efficient layer sequence. It was put to the test against a commonly used dataset of handwritten digit pictures, and it came out with a score of 98.83 percent accuracy. From the perspective of multi-objective evolution, an effective deep network compression method is proposed by Huang *et al.* [31]. A multi-objective compression deep learning model is built, as well as an approximate compression model generation mechanism, to reduce the high model training costs associated with the optimization process. Finally, the number of network parameters will be reduced, and the network evaluation process will be greatly accelerated. Atila *et al.* [32] Applied Efficientnet to plant leaf disease classification model and compared it with other deep learning models. The experimental results show that the performance of Efficientnet model achieves the highest value in plantvillage dataset. This work also implies that Efficientnet model has excellent feature extraction ability. Liu *et al.* [15] adopted a Multi-Object Evolutionary Algorithm Assisted Stacked Autoencoder (SAE MOEA/D), which can adaptively optimize the weights, activation functions and parameters such as balance factors and hyperparameters. Kim and Cho [33] suggested a PSO-based technique for CNN-LSTM Neural network optimization. PSO iteratively searches and optimizes CNN-complicated hyperparameters space of LSTM. Finally, the results show that the PSO optimized model outperforms other deep learning and machine learning models. Yu *et al.* [34] proposed a concrete crack detection method based on depth convolution neural network, and used the enhanced chicken swarm algorithm to optimize the super parameters of the network. After training and testing, good results are obtained.

As can be seen from the brief review above, object detection based on memetic algorithm optimization and deep learning has become an active research area in recent years. However, with the increasing complexity and real-time requirements of object detection tasks, how to further optimize the network performance of the deep learning model and how to use the memetic algorithm to improve the accuracy and efficiency of object detection to meet the actual detection needs of production requires further in-deep research.

III. ANALTSIS AND MODELING

A. ANALTSIS OF DATA SET

Forgings detection in this paper is cylinder head and flange plate, and the location of defects is determined. Based on practical experience, defects of cylinder head forgings are often on the edge of the bottom surface of the cylinder head, and the defects of flange forgings are near the bottom of the flange and the shaft hole, as showed in Fig.1. 450 original defect images were collected. First, 80% of the images are randomly selected as the training set, 10% as the test set, and 10% as the verification set. Then, the training set, test set and validation set are augmented independently. Using 8 different data augmentation methods including vertical flipping and adding noise, 1200 training sets, 150 test sets and 150 validation sets are finally obtained. In the actual production inspection process, it is only necessary to judge whether there is a defect in the forgings, and there is no need to distinguish the defect type. Therefore, all defects are classified as "crack" in the notes. Features of data set defects are as following:

1) There is different sizes and complex morphological features in forgings defects.

2) The defect background and defect distribution of different forgings is complex.

3) Poor light conditions lead to poor image quality.

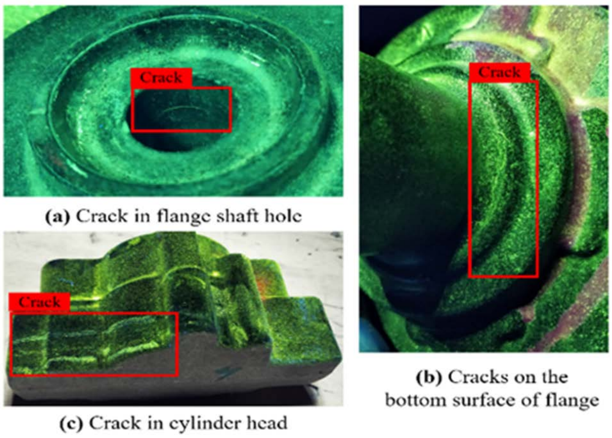


FIGURE 1. Schematic diagram of forgings defects.

B. MODEL AND DESIGN

According to the requirements of object detection algorithm, a new EfficientNet-PSO model is proposed in this paper. The highlight is as following, and its structure is shown in Fig.2

1) For problems with complex defect features, an improved EfficientNet [35] is used as the backbone.

2) Using FPN [36] as the feature fusion layer to improve the multi-scale object detection ability of the algorithm.

3) CIoU [37] is introduced to optimize the convergence speed and accuracy of the anchor box.

4) Aiming at the difficulty of model parameter adjustment, PSO [38] is used for automatic parameter search.

1) IMPROVED EFFICIENTNET

EfficientNet is a group of CNN based backbone released by Tan M *et al.* In 2019, there are 8 different models in this series (EfficientNet-B0 to EefficientNet-B7). The faster the model detection speed of this series, the lower the detection accuracy. According to the actual detection beats (20 to 26 piece/min), EfficientNet-B2 can both detection accuracy and efficiency, more suitable for the production of the enterprise. With the improved EfficientNet-B2, the model has stronger feature extraction ability and adapts to the dark magnetic particle flaw detection environment.

1) Efficientnet-B2 model includes Mobile Inverted Bottleneck Convolution module (MBConv) [39], and the Squeeze-and-Excitation Network module (SENet) [40] in MBConv adopts attention mechanism. The attention mechanism of model allows it to pay greater attention to the channel features that contain the most information while suppressing the less important channel features. MBConv has a structure that is similar to the residual connection. When the network depth is large, the gradient of MBConv is not easy to disappear, and the robustness of the model is better.

2) The original SENet uses the ReLU activation function, but forced sparse processing of ReLU will reduce the effective information received by the model. The negative gradient leads to the zeroing of ReLU, which may cause neuronal necrosis. In order to alleviate this problem, the Swish [41] activation functions is applied in the SEnet module. Swish activation function is a deformation of Sigmoid activation function. Formula 1 and Formula 2 are expressions of Sigmoid and Swish respectively, where x represent input.

$$Sigmoid\,(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$Swish\,(x) = x \cdot Sigmoid\,(x) \tag{2}$$

Swish function has no upper bound, so there will be no gradient saturation. It has a lower boundary and has a strong regularization effect, but it is less prone to neuronal necrosis than ReLU. Therefore, swish is more appropriative for the model. The function graph is shown in Fig.3

3) DropConnect [42] is applied in model to increase the generalization of model ability and mitigate the overfitting condition. The difference between DropConnect and DropOut is that DropConnect does not randomly discard the output of the hidden layer, but the input to the hidden layer. Both dropconnect and dropout can prevent overfitting and enhance the robustness of the network, but dropconnect is better in comparison. In addition, with the addition of scale parameters in EfficientNet model family, the model is easier to over fit and the drop rate of DropConnect gradually increases.

Efficientnet-B2 model is divided into 9 stages according to modules. Take Step 4, Step 6 and Step 9 as the effective layers, and its feature mAP will be connected to the FPN. The adjusted EfficientNet-B2 model structure is shown in Table 1.

2) FEATURE PYRAMID NETWORK

Although the position of the defect on the forgings is fixed, the position of the forgings on the image is not fixed and the size is different. Recognizing object with large size differences is one of the basic challenges faced by computer vision. The common solution is to use FPN as the multi-scale feature fusion layer. Different from the conventional CNN model, FPN integrates the feature maps of multiple prediction layers. It can be detected according to the characteristics of different

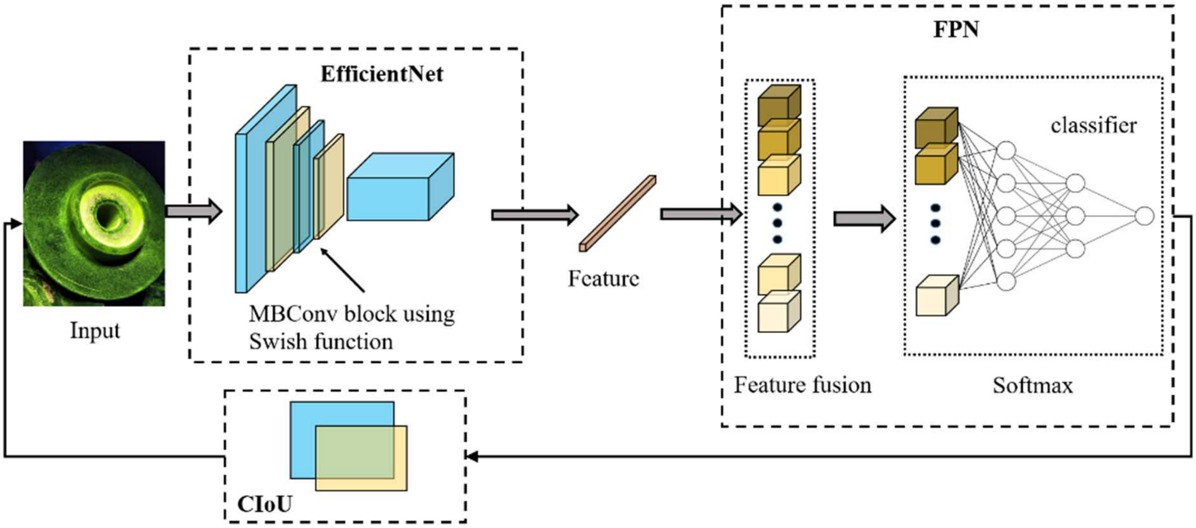


FIGURE 2. Model structure diagram.

TABLE 1. Structure of improved EfficientNet-B2 backbone.

Step	Module	Resolution $\hat{H}_i \times \hat{W}_i$	Channel $\hat{C}_i$	Layer $\hat{L}_i$
1	Conv3×3	208×208	32	1
2	MBConv1,k3×3	104×104	16	1
3	MBConv6,k3×3	104×104	24	2
4	MBConv6,k5×5	52×52	48	2
5	MBConv6,k3×3	26×26	88	3
6	MBConv6,k5×5	26×26	120	3
7	MBConv6,k5×5	13×13	208	4
8	MBConv6,k3×3	13×13	352	1
9	Conv1×1&Pooling&FC	13×13	1408	1

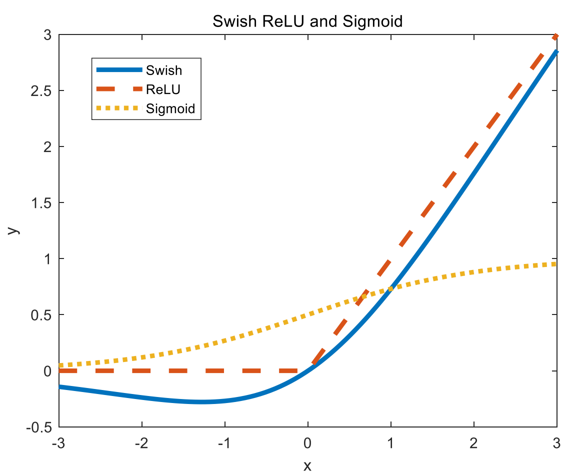


FIGURE 3. Three activation functions.

scale defects of flange plate and cylinder heads. If only the top-level features are used for regression prediction, the objects with small scale and weak features will be lost. Therefore, in order to deal with the multi-scale prediction task, FPN is used to predict in multiple independent feature layers. FPN combines high-resolution shallow feature layer and rich semantic information deep feature layer to realize multi-scale feature fusion.

In EfficientNet-PSO, FPN extracts three feature layers of EfficientNet: Step 4, Step 6 and Step 9. The corresponding feature mAP resolutions are 52 × 52, 26 × 26 and 13 × 13. After multiple convolution operations, one component of each of the three feature layers is utilized to output the feature related findings of layer, while the other is deconvoluted and fused with other feature layers.

3) COMPLETE INTERSECTION OVER UNION

In order to improve the accuracy of the loss function of the object position and speed up the convergence speed during training, CIoU is used as the intersection over union of the loss function in EfficientNet-PSO. The intersection over union can used to measure the loss between the predicted bounding box and the ground truth. The commonly used loss functions of intersection over union include IoU, Generalized Intersection Over Union (GIoU) [43] and Distance Intersection Over Union (DIoU). IoU converges slowly in training, and when there is no intersection between prediction box and real box, IoU is always zero and loses its affect as a loss function GIoU can reflect the relationship between the ground truth and the prediction bounding box when there is no intersection between the prediction bounding box and the ground truth, however the convergence speed of GIoU is not beneficial; DIoU is better than the first two kinds of intersection over union, but its disadvantage is that it does not consider the aspect ratio of the bounding box in the process of reasoning. CIoU overcomes the shortcomings of the above functions and takes into account the fast convergence speed while ensuring the accuracy and the convergence effect is better. The formula of CIoU is as following:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{3}$$

The formula of IoU is as follows:

$$IoU = \frac{A \cap B}{A \cup B} \tag{4}$$

The formula of α is as follows:

$$\alpha = \frac{v}{(1 - IoU + v)} \tag{5}$$

The formula of v is as follows:

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{6}$$

where, $\rho^2(b, b^{gt})$ is the Euclidean distance, c is the diagonal length of the minimum area, v is the consistency measurement parameter of aspect ratio, w and h is the width and height of the predicted bounding box, w^gt and h^gt is the width and height of the ground truth respectively.

4) MEMETIC ALGORITHM OPTIMIZTION AND PARAMETER AUTOMATIC SEARCH

In order to enhance the performance of EfficientNet-PSO model, some tricks such as Weight Decay [44], Mosaic Data Augmentation [45], Cosine Annealing Learning Rate Decay [46] and Loss Normalization are introduced. The model itself and the five tricks contain eight key parameters, which are the decisive factors determining the performance of the model. It shows in Table 2. How to choose a faster and more accurate model from many parameters is the key problem to be solved. Therefore, it is necessary to use PSO algorithm for parameter automatic search. The particle position is determined by the 8 key parameters, and the fitness is determined by the loss of model on the validation set. The inertia factor W and learning factors C_1 and C_2 of PSO change dynamically according to the number of iterations. In order to enhance the robustness of PSO, a parameter dynamic increase strategy based on the number of iterations is used. At the beginning of the iteration, the individual diversity of particles is enhanced to avoid falling into local optimization, and at the end of the iteration, the group sharing of particles is enhanced to speed up the convergence speed.

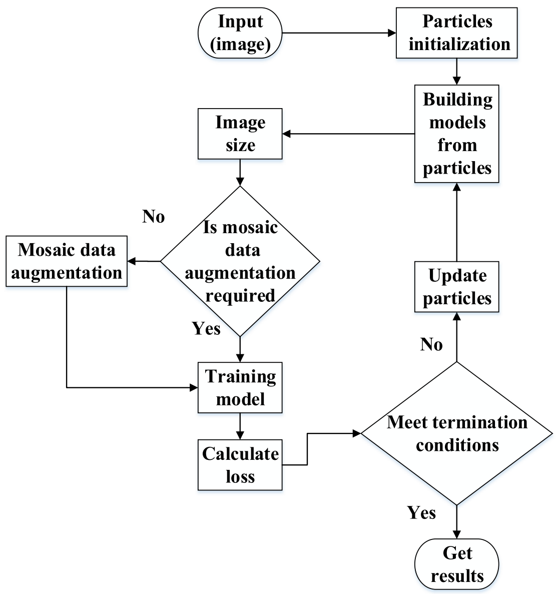


FIGURE 4. Model flow chart.

IV. EXPERIMENT

A. THE DESIGN OF INTELLIGENT DETECTION PLATFORM

In the traditional forgings defect detection process, firstly, magnetizing equipment will be used to magnetize the forgings, and then the magnetized forgings will be sent to the dark detection room. Finally, inspectors will carry out manual detection in the detection room and manually sort the defective forgings. In order to realize intelligent detection of forgings defects and simulate the actual production situation, an intelligent detection platform for forgings defects was designed and built. In order to obtain a better effect of black light illumination, in the case of high indoor brightness, a hood is usually added to the outside. According to actual production experience, the defect size of cylinder head and flange plate is less than 70mm usually. Therefore, the experiments mainly use the following devices. S4560-6K floating LED black light, UV(Light)-FLUX $\geq$ 6 000 μW/cm2; The industrial camera FLIR BFS-U3-89S6M with a resolution of 4096 $\times$ 2160, and the imaging range is 207.8mm $\times$ 106.4mm.The detection process is shown in Fig.5.

B. MODEL TRAINING

The models trained and tested by EfficientNet-PSO run on the workstation of Ubuntu20.0 operating system. The hardware of the workstation is Inter3. 10 GHz 64 core CPU, 128 GB memory, two NVIDIA Titan XP GPUs. The software environment is Ubuntu20.0 operating system based on 64 bit, pytorch1.7.1 framework, CUDA11.0, OpenCV2 and Visual Studio Code integrated development environment.

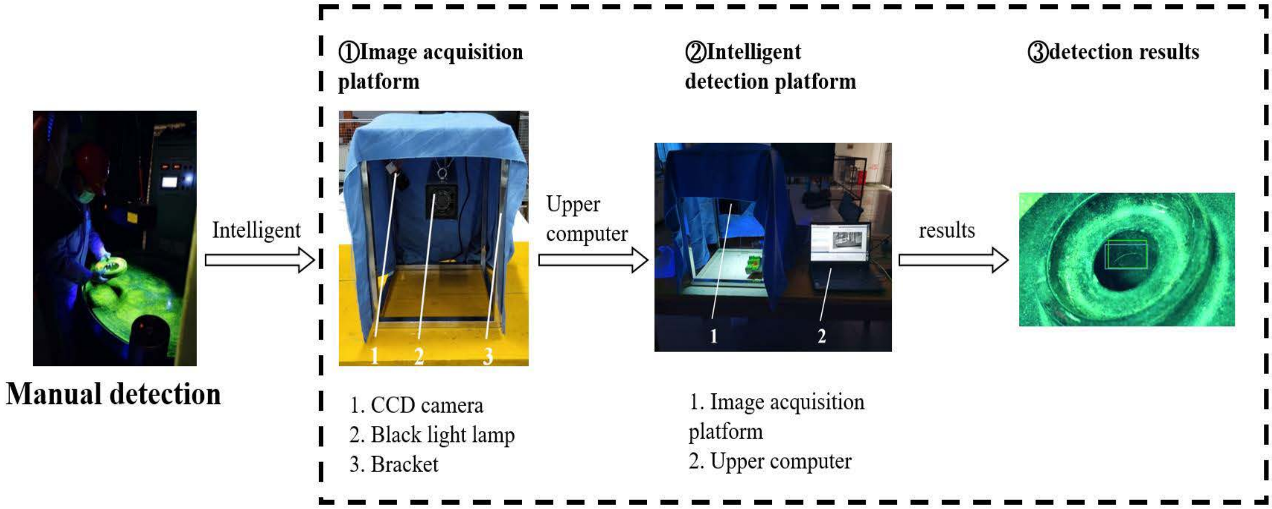


FIGURE 5. Schematic diagram of detection process.

TABLE 2. Key parameters of EfficientNet-PSO.

Parameter	Value range	Parameter type
Optimizer	"SGD", "Adadelta", "Adagrad", "Adam", "Adamax"	Discrete
Weight Decay	[0, 1e-3]	Consecutive
Normalize	True, False	Discrete
Input Shape	416, 608	Discrete
Mosaic	True, False	Discrete
Learning Rate	[1e-2, 1e-4]	Consecutive
Cosine lr	Ture, False	Discrete
Batch size	2, 4, 8, 16, 32, 64	Discrete

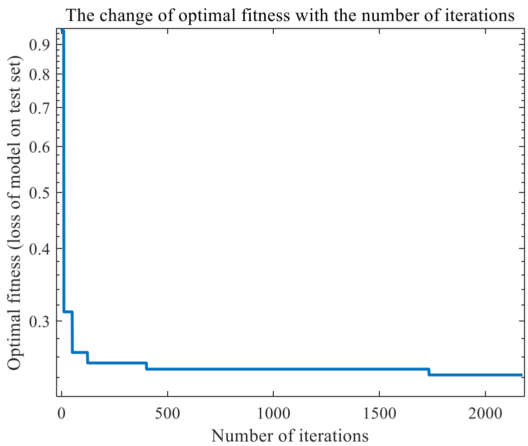


FIGURE 6. Change of optimal fitness with the number of iterations.

TABLE 3. Parameters of particle swarm optimization.

Parameter	Value Range
Population number	10
Number of iterations	30
W	[0.7,1.4]
C_1	[0.5,2.5]
C_2	[0.5,2.5]

The parameters of PSO algorithm are shown in Table 3. The optimal parameters of the CNN model obtained through

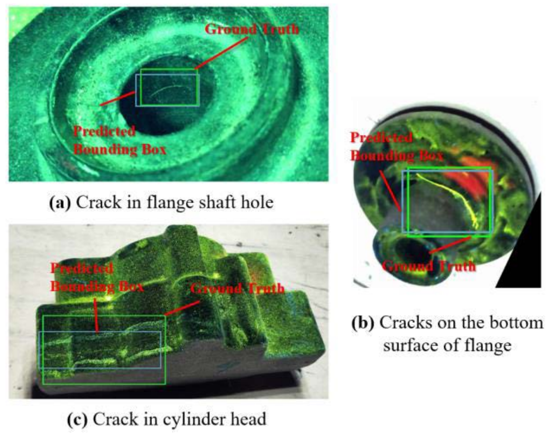


FIGURE 7. Schematic diagram of detection results.

TABLE 4. Optimal parameters of CNN model.

Parameter	Value Range
Optimizer	"Adadelta"
Weight Decay	0
Normalize	True
Input Shape	608
Mosaic	False
Learning Rate	1e-3
Cosine lr	False
Batch size	32

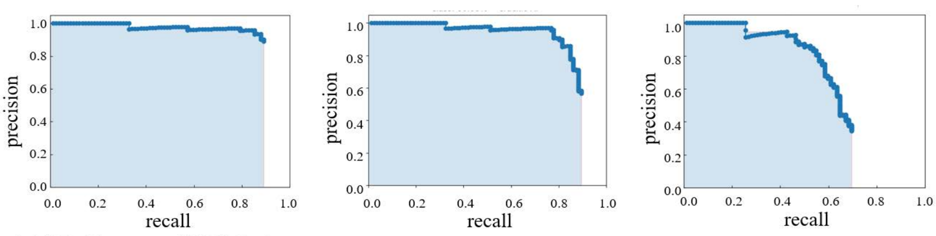
iterative calculation are shown in Table 4. Detection results of a single picture are shown in Fig.7. The iterative curve of PSO algorithm is shown in Fig.6.

V. MODEL PERFORMANCES

In this section, comparative experiments will be carried out from two aspects: the performance of CNN model and the performance of HPO algorithm.

A. COMPARATIVE EXPERIMENT OF CNN MODEL

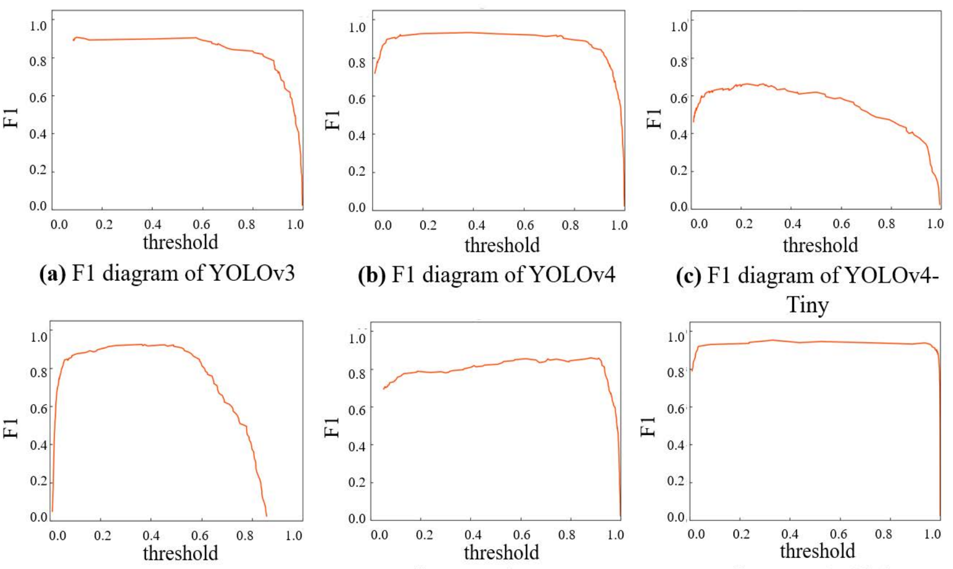
The accuracy average value of rate under various recall rates is AP; the average value of different categories of AP is mAP; the combined value of recall rate and accuracy rate is F1; the



(a) PR diagram of YOLOv3 (b) PR diagram of YOLOv4 (c) PR diagram of YOLOv4-Tiny

(d) PR diagram of CenterNet (e) PR diagram of Faster-RCNN (f) PR diagram of EfficientNet-PSO

FIGURE 8. PR diagram of six models.



(a) F1 diagram of YOLOv3 (b) F1 diagram of YOLOv4 (c) F1 diagram of YOLOv4-Tiny

(d) F1 diagram of CenterNet (e) F1 diagram of Faster-RCNN (f) F1 diagram of EfficientNet-PSO

FIGURE 9. F1 diagram of six models.

total parameters of the model scale are represented by total params. Model complexity and detection speed is expressed by floating-point operation FLOPs. The above evaluation indicators can accurately evaluate the trained model.

EfficientNet-PSO was compared with the following five models: YOLOv4, YOLOv3 [47], CenterNet [48], YOLOv4 Tiny and faster RCNN [49]. Using the same dataset, after 100 generations of training, the optimal results of model validation set test results. Figure 8 is a PR comparison chart of the 6 models. The larger the area enclosed by the curve in the figure, the higher the accuracy of the model; the more stable the curve, the better the performance of the model when the positive and negative samples are uneven. EfficientNet-PSO is the best of the six models. Figure 9 shows the F1 curve, the closer the F1 curve is to 1 and the more stable it is, the better the detection performance is. EfficientNet-PSO is the best of the six models. Figure10 shows the comparison of EfficientNet-PSO with the mAP and FLOPs of the other five models, and the closer to the lower right corner in the figure, the better the performance. Among the six models, the mAP and FLOPs of EfficientNet-PSO are the best; it boasts the fastest detection speed and the highest detection accuracy.

TABLE 5. Comparison of experimental results of six models.

Model	mAP	Total Params (Million)	FLOPs (Billion)	F1	Recall	Precision
YOLOv3	87.33%	61.9	32.8	0.90	85.37%	**95.89%**
YOLOv4	93.32%	64.0	29.9	0.92	89.02%	94.81%
YOLOv4-Tiny	61.23%	**5.9**	3.41	0.62	46.34%	92.86%
CenterNet	91.73%	32.7	3.8	0.91	82.93%	100%
Faster-RCNN	86.08%	43.2	23.9	0.82	84.88%	80.22%
EfficientNet-PSO	**95.69%**	10.7	**1.86**	**0.94**	**93.90%**	95.06%

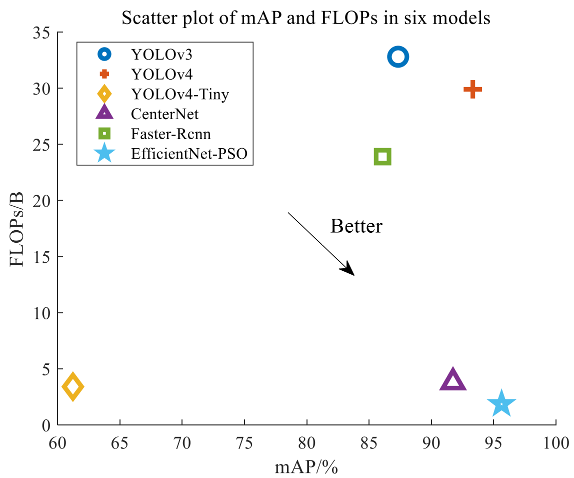


FIGURE 10. Scatter plot of mAP FLOPs in six models.

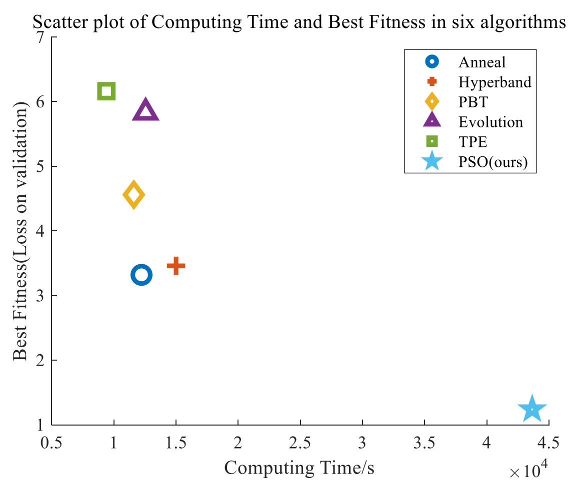


FIGURE 11. Scatter plot of computing time and best fitness in six algorithms.

B. COMPARATIVE EXPERIMENT OF HPO ALGORITHM

PSO was compared with the following five algorithms: Anneal, Evolution [50], Hyperband [51], PBT [52], and TPE [13]. The above five algorithms include heuristic algorithm and Bayesian algorithm. Fig.11 shows the comparison of PSO with the computing time and best fitness of the other five algorithms. In the figure, the closer to the bottom, the better the fitness, and the closer to the left, the less the computing time. The computing time of PSO application in HPO is not dominant, because PSO requires a certain scale of population, which will produce a large computational overhead. But the best fitness of PSO is obviously better than other algorithms. Compared with other algorithms, using PSO algorithm is a method that uses computing time to exchange the optimization effect. Considering that the absolute value of PSO computing time is only a few hours higher than other algorithms, this is an acceptable cost. Therefore, we believe that PSO is more suitable than other algorithms to solve the problem.

TABLE 6. Parameters of particle swarm optimization.

HPO Algorithm	Computing Time	Best Fitness (Loss on Validation)
Anneal	12227s	3.32
Evolution	15007s	3.46
Hyperband	11607s	4.56
PBT	12558s	5.83
TPE	**9404s**	6.16
PSO	43654s	**1.24**

VI. CONCLUSION

This paper proposes a new intelligent detection model for automobile steel die forgings defects. Taking the cylinder head and flange plate as the detection object, the EfficientNet deep learning network model is constructed as the main body, FPN is the feature fusion layer, and the parameters are automatically searched through the CIoU and PSO algorithms, which improves the defect detection accuracy of the model. EfficientNet-PSO model is not only high precision, but also an efficient, which avoids the complex image processing process in traditional object detection. It can realize end-to-end real-time detection. The main conclusions are as following:

1. For the efficient extraction of complex high-level semantic features, the EfficientNet-PSO model proposed in this paper introduces the Swish activation function and combines

the SENet self-attention module to effectively improve the detection performance of the model.

2. The high-level semantic features are introduced into the FPN feature fusion layer, which effectively enhances the multi-scale object detection ability of the model.

3. The description of the object frame overlaps is more accurate than traditional calculation methods, and effectively optimizes the convergence speed and accuracy of the bounding box by introducing CIoU.

4. PSO algorithm is used to solve the problem that it is difficult to adjust the parameters of the CNN model. It is proved that memetic computing is not only effective in dealing with the problems of computer vision and image processing. But also plays a practical role in engineering application.However, in the experimental results, it is found that very few samples have to overlap two predicted bounding box, which may be due to unreasonable optimization of IoU threshold of non-maximum Suppression module, which needs further research. In addition, despite the fact that the PSO technique is utilized in this paper, is PSO the best memetic algorithm? How to choose the appropriate memetic algorithm for different CNN models? These problems need to be further studied in order to better serve the actual forgings inspection need.

In the future work, it is planned to use neural architecture search technology (NAS) to improve the universality and performance of detection. If NAS is combined with HPO, the detector will be completely computer-generated rather than designed by professionals.

REFERENCES

[1] A. S. Bakunov, A. Y. Korolev, D. A. Kudryavtsev, and V. P. Petrov, "A set for magnetic fluorescent-penetrant inspection," *Russian J. Nondestruct. Test.*, vol. 41, no. 3, pp. 170–174, Mar. 2005.

[2] B. Potter, "NDT of precipitation hardened steels using fluorescent magnetic particle testing," *Mater. Eval.*, vol. 61, no. 12, pp. 1280–1287, 2003.

[3] W. Hou, Y. Wei, J. Guo, Y. Jin, and C. Zhu, "Automatic detection of welding defects using deep neural network," *J. Phys., Conf. Ser.*, vol. 933, Jan. 2018, Art. no. 012006.

[4] Q. Wang, R. Yang, C. Wu, and Y. Liu, "An effective defect detection method based on improved generative adversarial networks (iGAN) for machined surfaces," *J. Manuf. Processes*, vol. 65, pp. 373–381, May 2021.

[5] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: A survey," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 84–100, Jan. 2018.

[6] W. Nsengiyumva, S. Zhong, M. Luo, Q. Zhang, and J. Lin, "Critical insights into the state-of-the-art NDE data fusion techniques for the inspection of structural systems," *Struct. Control Health Monitor.*, vol. 29, no. 1, p. e2857, Jan. 2022.

[7] M. Li, S. Wan, Z. Deng, and Y. Wang, "Fabric defect detection based on saliency histogram features," *Comput. Intell.*, vol. 35, no. 3, pp. 517–534, Aug. 2019.

[8] X. Wang, B. S. Wong, G. T. Chen, P. K. Kai, and F. Foo, "Image enhancement for radiographic non-destructive inspection of the aircraft," in *Proc. 12th Asia–Pacific Conf. NDT*, 2006, pp. 1–6.

[9] H. Fang, M. Xia, H. Liu, Y. Chang, L. Wang, and X. Liu, "Automatic zipper tape defect detection using two-stage multi-scale convolutional networks," *Neurocomputing*, vol. 422, pp. 34–50, Jan. 2021.

[10] J. Shi, H. Zhu, S. Yu, W. Wu, and H. Shi, "Scene categorization model using deep visually sensitive features," *IEEE Access*, vol. 7, pp. 45230–45239, 2019.

[11] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020.

[12] H. Tang and Z. Hu, "Research on medical image classification based on machine learning," *IEEE Access*, vol. 8, pp. 93145–93154, 2020.

[13] J. Bergstra, R. Bardenet, B. Kégl, and Y. Bengio, "Algorithms for hyper-parameter optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1–9.

[14] A.-C. Florea and R. Andonie, "Weighted random search for hyperparameter optimization," *Int. J. Comput. Commun. Control*, vol. 14, no. 2, pp. 154–169, Apr. 2019.

[15] G. Liu, Y. Li, L. Jiao, Y. Chen, and R. Shang, "Multiobjective evolutionary algorithm assisted stacked autoencoder for PolSAR image classification," *Swarm Evol. Comput.*, vol. 60, Feb. 2021, Art. no. 100794.

[16] C. Wang, Y. Wang, K. Wang, Y. Dong, and Y. Yang, "An improved hybrid algorithm based on biogeography/complex and metropolis for many-objective optimization," *Math. Problems Eng.*, vol. 2017, pp. 1–14, Mar. 2017.

[17] T. Ling and C. Wang, "An MOEA/D-ACO with PBI for many-objective optimization," *Math. Problems Eng.*, vol. 2018, Mar. 2018, Art. no. 5414869.

[18] J.-K. Park, B.-K. Kwon, J.-H. Park, and D.-J. Kang, "Machine learning-based imaging system for surface defect inspection," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 3, no. 3, pp. 303–310, 2016.

[19] X. Dai, Y. Zhao, and C. Zhu, "A study of an improved RCNN network model for surface defect detection algorithm of precision workpiece and its realisation," *Int. J. Wireless Mobile Comput.*, vol. 19, no. 1, p. 95, 2020.

[20] J. Shen, N. Liu, and H. Sun, "Defect detection of printed circuit board based on lightweight deep convolution network," *IET Image Process.*, vol. 14, no. 15, pp. 3932–3940, Dec. 2020.

[21] Z. Tong, D. Yuan, J. Gao, and Z. Wang, "Pavement defect detection with fully convolutional network and an uncertainty framework," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 8, pp. 832–849, Aug. 2020.

[22] J. H. Feng, H. Yuan, Y. Q. Hu, J. Lin, S. W. Liu, and X. Luo, "Research on deep learning method for rail surface defect detection," *IET Electr. Syst. Transp.*, vol. 10, no. 4, pp. 436–442, Dec. 2020.

[23] Y. Yu, C. Wang, X. Gu, and J. Li, "A novel deep learning-based method for damage identification of smart building structures," *Struct. Health Monitor.*, vol. 18, no. 1, pp. 143–163, Jan. 2019.

[24] W. Du, H. Shen, J. Fu, G. Zhang, and Q. He, "Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning," *NDT & E Int.*, vol. 107, Oct. 2019, Art. no. 102144.

[25] H. Di, X. Ke, Z. Peng, and Z. Dongdong, "Surface defect classification of steels with a new semi-supervised learning method," *Opt. Lasers Eng.*, vol. 117, pp. 40–48, Jun. 2019.

[26] R. Ren, T. Hung, and K. C. Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 929–940, Mar. 2018.

[27] C. Hu and Y. Wang, "An efficient convolutional neural network model based on object-level attention mechanism for casting defect detection on radiography images," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10922–10930, Dec. 2020.

[28] M. Nekkaa and D. Boughaci, "A memetic algorithm with support vector machine for feature selection and classification," *Memetic Comput.*, vol. 7, no. 1, pp. 59–73, Mar. 2015.

[29] J. Liu, M. Gong, Q. Miao, X. Wang, and H. Li, "Structure learning for deep neural networks based on multiobjective optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2450–2463, Jun. 2018.

[30] A. Martín, R. Lara-Cabrera, F. Fuentes-Hurtado, V. Naranjo, and D. Camacho, "EvoDeep: A new evolutionary approach for automatic deep neural networks parametrisation," *J. Parallel Distrib. Comput.*, vol. 35, no. 3, pp. 517–534, 2018.

[31] J. Huang, W. Sun, and L. Huang, "Deep neural networks compression learning based on multiobjective evolutionary algorithms," *Neurocomputing*, vol. 378, pp. 260–269, Feb. 2020.

[32] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," *Ecol. Informat.*, vol. 61, Mar. 2021, Art. no. 101182.

[33] T.-Y. Kim and S.-B. Cho, "Optimizing CNN-LSTM neural networks with PSO for anomalous query access control," *Neurocomputing*, vol. 456, no. 2, pp. 666–677, Oct. 2021.

[34] Y. Yu, M. Rashidi, B. Samali, M. Mohammadi, T. N. Nguyen, and X. Zhou, ''Crack detection of concrete structures using deep convolutional neural networks optimized by enhanced chicken swarm algorithm,'' *Struct. Health Monitor.*, Jan. 2022, Art. no. 14759217211053546.

[35] M. Tan and Q. V. Le, ''EfficientNet: Rethinking model scaling for convolutional neural networks,'' 2019, *arXiv:1905.11946*.

[36] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, ''Feature pyramid networks for object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2117–2125.

[37] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, ''Distance-IoU loss: Faster and better learning for bounding box regression,'' in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 12993–13000.

[38] F. Marini and B. Walczak, ''Particle swarm optimization (PSO). A tutorial,'' *Chemometrics Intell. Lab. Syst.*, vol. 149, pp. 153–165, Dec. 2015.

[39] Z. Jiang, L. Zhao, S. Li, and Y. Jia, ''Real-time object detection method based on improved YOLOv4-tiny,'' 2020, *arXiv:2011.04244*.

[40] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, ''Squeeze-and-excitation networks,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[41] P. Ramachandran, B. Zoph, and Q. V. Le, ''Searching for activation functions,'' 2017, *arXiv:1710.05941*.

[42] Li Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, ''Regularization of neural networks using DropConnect,'' in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, 2013, pp. 1058–1066.

[43] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, ''Generalized intersection over union: A metric and a loss for bounding box regression,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.

[44] I. Loshchilov and F. Hutter, ''Decoupled weight decay regularization,'' 2017, *arXiv:1711.05101*.

[45] A. Bochkovskiy, C. Y. Wang, and H. Liao, ''YOLOv4: Optimal speed and accuracy of object detection,'' in *Proc. Comput. Vis. Pattern Recognit.*, 2020, pp. 1–17.

[46] I. Loshchilov and F. Hutter, ''SGDR: Stochastic gradient descent with warm restarts,'' 2016, *arXiv:1608.03983*.

[47] J. Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, *arXiv:1804.02767*.

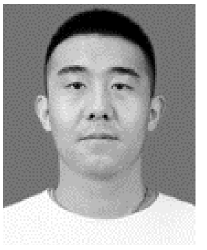
[48] X. Zhou, D. Wang, and P. Krähenbühl, ''Objects as points,'' 2019, *arXiv:1904.07850*.

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

[50] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, ''Large-scale evolution of image classifiers,'' in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, Sydney, NSW, Australia, 2017, pp. 2902–2911.

[51] L. Li, K. Jamieson, G. Desalvo, A. Rostamizadeh, and A. Talwalkar, ''Hyperband: A novel bandit-based approach to hyperparameter optimization,'' *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6765–6816, 2016.

[52] J. M. Elliot, C. Wojciech, G. T. F. Goldie, and D. V. Clement, ''Population based training of neural networks,'' 2017, *arXiv:1711.09846*.



TANG YU was born in Hebei, China, in 1996. He received the B.S. degree in energy and power engineering from the School of Science and Technology, Hebei University of Science and Technology, in 2020. He is currently pursuing the M.S. degree in mechanics with the Hubei Institute of Automotive Industry. His research interests include intelligent manufacturing and machine vision.

WANG CHEN was born in Shiyan, Hubei, China, in 1983. He received the B.S. degree from the Hubei University of Automotive Technology, in 2006, the M.S. degree from the Wuhan University of Science and Technology, in 2009, and the Ph.D. degree in mechanical engineering from Shanghai University, in 2019. Since 2018, he has been an Associate Professor with the Mechanical Engineering Department, Hubei University of Automotive Technology.

GAO JUNFENG was born in Shiyan, Hubei, China, in 1986. He received the B.S. degree from Air Force Engineering University, in 2008. He works with the Shiyan Industrial Product Quality Inspection and Testing Institute and obtained the title of an Engineer, in 2020. In 2022, he was a Chief Engineer with the Shiyan Industrial Product Quality Inspection and Testing Institute.

HUA POXI was born in Chengguan, Lanzhou, Gansu, China, in 1997. He received the B.S. degree in mechanical engineering from Fuzhou University, in 2019. He is currently pursuing the M.S. degree in mechanical engineering with the Hubei University of Automotive Technology. His research interests include intelligent manufacturing and machine vision.