INTERNATIONAL JOURNAL OF MATERIALS
DOI: 10.46300/91018.2022.9.5 Volume 9, 2022

Improving the algorithm study of YOLO in steel surface defect detection

Fu Su[1], Siying Wang[1]
[1]College of Electrical Information, Southwest Petroleum University, Chengdu 610500, Sichuan, China

Abstract—To solve the problem of low detection accuracy caused by background interference and diverse target forms, a series of improvements are proposed to improve the detection accuracy. According to the various characteristics of steel surface defects, this paper presents the K-Means clustering algorithm to optimize the clustering results and quickly and accurately obtain the size of the prior box. In view of the small proportion of the target defect area in the overall image and background interference, a two-way attention module (TWA-Block) is proposed to establish the long-distance dependence of the spatial domain and channel domain features, and a background suppression function is designed to realize the division of defect areas. Experiments of the proposed improvements in the NEU-DET dataset based on the YOLO series model show that the detection accuracy of all the improved YOLO series models has improved, and the number of parameters will not increase substantially.

Keywords—Defect Detection, A K-Means Clustering Algorithm for the Optimized Cuckoo Search, Dual-way Attention Module

I. INTRODUCTION

In the actual production process of steel, due to the simple manufacturing process, rough production mode, and other problems, the steel surface will have scratches, scratches, cracks, mottled, and other defects, which significantly impact the performance of steel products[1].

With the increasing popularity of deep learning in various industries, object detection-based methods in steel surface defect detection have gradually replaced traditional machine learning[2]. Still, the detection accuracy of this method needs to be improved. Therefore, this paper introduces the attention mechanism in the target detection model to improve the detection accuracy of steel surface defects.

At present, the convolutional neural network is widely used for feature extraction and classification in steel surface defect detection. In the current target detection task, there are two mainstream frameworks: two-step detection framework, that is, obtain candidate regions first, and then classify and regression, such as Fast R-CNN (Region-Conventional Neural Network)[3], Faster R-CNN[4], etc.; the one-step detection framework, such as YOLO (You Only Look Once)[5], SSD[6], etc. The two-step detection framework is accurate but slow and unsuitable for real-time detection; the one-step detection framework is slightly less accurate than the two-step method but fast detection[7].

As the depth of the network gradually deepens, the features of the target become less evident in the deep convolution, so the addition of an attention mechanism to the shallow network can strengthen the target features in the shallow network while effectively inhibiting the interference of background features. The traditional attention mechanism, SE-Block [8], channel aggregates and scales with unified weights for all pixels, but the method does not fully use the global context. The CBAM [9] provides a spatial attention mechanism based on the channel attention mechanism, but it still does not realize the inhibition of the interference features. Although these methods were very effective in the classification task, poor detection in the Yolo series model [10], [11]. Woo et al. [12] proposed that simultaneous weighting processing in the channel and spatial dimensions of the feature layer can highlight the target features. Wang et al. [13] proposed non-local neural networks that enhance spatially local feature information by modeling the connections between locations. In all the target detection models of the YOLO series, the K-means clustering algorithm for the training data is needed to obtain the size of the a priori box. However, the traditional K-means clustering algorithm has always had several shortcomings:

1. Blindness and randomness of the initial cluster of central

constituencies;

2. The clustering results converge to the local optimum.

Because of the above K-means clustering algorithm deficiency, many scholars at home and abroad have conducted the relevant research and put forward the improvement scheme. More of these solutions are based on improvements in group intelligence algorithms. Lin [14] proposed the GA-K-means clustering algorithm based on the genetic algorithm, which improves the global search capability of the traditional K-means clustering algorithm. Liu et al. [15] proposed a K-means clustering algorithm based on the particle group algorithm with a higher global search power than the existing genetic algorithm-based K-means clustering algorithms. Liu et al. [16] introduced the density information determination of the initial clustering center to improve the K-means clustering algorithm and improve the execution speed of the algorithm.

To improve the accuracy of steel surface defect detection, we propose a K-means attention model and clustering algorithm for optimizing cuckoo search. In this paper, relevant contrast experiments were designed for validation on the NEU-DET dataset. Finally, the comparative experimental results show that the improved YOLO series algorithm has a higher accuracy than the original YOLO series model.

II. A K-MEANS CLUSTERING ALGORITHM FOR OPTIMIZED CUCKOO SEARCH

The YOLOv3, YOLOv4 models adopt the idea of anchor boxes used in the Faster R-CNN, a set of initial candidate boxes with complete and high fixation, and the selection of the initial anchor boxes directly affects the detection accuracy and speed of the network on the target, [17].

A. The K-means clustering algorithm

The K-means clustering algorithm is a distance-based clustering algorithm, which belongs to the unsupervised linear clustering algorithms category. The clustering algorithm judges the similarity of two samples by calculating the distance between samples.[18] If the distance between two models is smaller, the similarity.

The basic flow of the proposed algorithm is:

K cluster centers were set to calculate the distance between each sample in the sample set to the cluster center and select the sample points in the cluster center according to the principle of the minimum distance. The cluster centers were recomputed based on the within-class distance between the sample points in each cluster, so consistently iterated until the termination conditions were met.[19]

The specific steps of the K-means clustering algorithm are shown as follows:

(1) randomly select k points as the initial value of the cluster center;

(2) Calculate the distance from each sample to the k cluster centers, and then classify these sample points into the class corresponding to the cluster center according to the principle of the shortest distance, forming the k clusters;

(3) Update the center of each cluster;

(4) Repeat steps (2) and steps (3) until the target function tends to converge.

The K-means algorithm is improved by the K-means clustering algorithm and its convergence to local optimization.

B. A cuckoo search algorithm

The cuckoo search algorithm has evolved based on the breeding habits of the cuckoo in nature. The main idea is to find the nest by random selection and update the nest position with random step size, thus bringing the final selected nest position to or near the optimal global solution.[19]

Under the premise of three idealization rules, the updated formula for the location and path of the nest is as follows:[20]

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \otimes L(s,\lambda) \quad (1)$$

$$L(s,\lambda) = \frac{\lambda\Gamma(\lambda)\sin(\pi\lambda/2)}{\pi}\left(\frac{1}{s^{1+\lambda}}\right), (s \gg 0) \quad (2)$$

Where $x_i^{(t)}$ represents the position of i the nest at the t-iteration, the step length factor is used to control the step length during the search, and the value follows a normal distribution.

After the nest position is updated by formula _ (1), a random number $\gamma(\gamma \in [0,1])$ will be compared with Pa. If $\gamma >$ pa, the nest position will be updated randomly using Levi flight, otherwise the nest position remains unchanged. The random walk strategy for Levy flight is shown in Equation (3):

$$x_k^{(t+1)} = x_k^{(t)} + r(x_i^{(t)} - x_j^{(t)}) \quad (3)$$

C. A K-means clustering algorithm for optimized cuckoo search

The alternating use of size steps in the traditional CS algorithm enhances the global search ability, but the late convergence rate is slow and lacks vitality for the main reasons:

1. The parameter pa and the convergence rate can be used to adjust the algorithm. The traditional CS algorithm adopts fixed values for both parameters and cannot be changed in the new generation.
2. If the value of pa is small and large, the algorithm will perform poorly and significantly increase in the number of iterations. If the pa value is large and small, the convergence rate is higher, but the best solution may not be found.

Therefore, both pa and parameters are needed to enable the CS search algorithm to find the best solution with high convergence rates. This paper introduces the adaptive discovery probability pa, which changes with the number of iterations, improves the algorithm's search ability, presents the adaptive search step with the number of iterations, and improves the convergence speed of the algorithm.

(1) Discovery of probabilistic adaptation

In the early stage of the algorithm, the value of pa needs to be large enough to enhance the population diversity. In the late stage, the value of pa needs to be reduced to adjust the number of solutions better. Therefore, this paper gives the adaptive strategy of pa according to the above rules:

$$p_a = 1.1 - \exp[-(\frac{t_i}{t_{\max}})^{\theta}] \quad (4)$$

(2) Search for step length adaptive

The adaptive strategy of the search step length designed in this paper considers the fitness value of each individual and the current number of iterations, which is together used to determine the change of the search step length. The adaptive strategy is as follows:

$$\alpha = \exp(-\frac{t_i - 1}{t_{\max}})[\alpha_{\max} + (\alpha_{\max} - \alpha_{\min})(1 - \frac{f_{\max}^{t)}}{f_i^{t)}})] \quad (5)$$

The nests in the optimized cuckoo search algorithm correspond to divided categories, and the location of each nest corresponds to k cluster centers to improve the search of the data cluster center. The problem of the optimal local solution of the traditional K-means algorithm is overcome by integrating the optimized cuckoo algorithm and the K-means clustering algorithm.

The implementation of the K-means clustering algorithm for optimizing the cuckoo search is as follows:

Step 1: Determine the number of clusters, k;

Step 2: Construct the fitness function;

$$f = \frac{\sum_{i=l}^{k} \sum_{j=l}^{n_k} I_{IoU}(B, C)}{n} \quad (6)$$

Step 3: Initialize the positions of N bird's nests randomly and calculate the fitness value of each bird's nest according to the fitness function. And keep the good bird's nest to the next generation.

Step 4: Update the next generation of bird's nest location according to the adaptive cuckoo algorithm.

Step 5: A random number U (0,1) is randomly generated. If> pa, new nest positions are randomly generated, otherwise the nest position remains unchanged, thus leaving the nest with the highest fitness.

Step 6: If the algorithm termination condition is reached, output the position of the optimal global nest, that is, output the optimal solution of the cluster center; otherwise, return to the fourth step cycle.

III. Dual-way attention module design

To effectively extract the texture and profile characteristics of steel surface defect images and reduce the interference of spatial attention and channel attention serial connections, a two-way attention module (TWA-Block) is designed, as shown in Figure 1. The module also selects the critical features in the feature map from the spatial and channel dimensions in parallel. It organizes the background suppression function to suppress the background of the feature map and enhance the defect features. The calculation of the model is shown in equation (7).

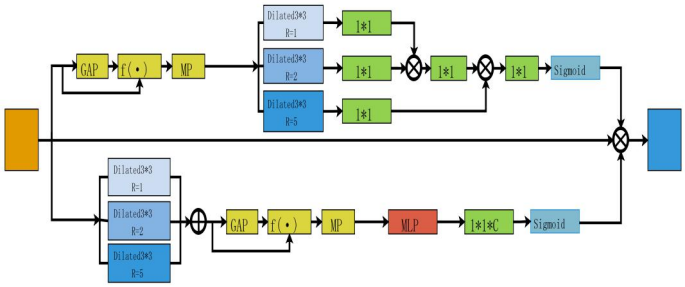


Fig. 1 two-way attention block（TWA-Block）

$$Z = F \otimes F_S \otimes F_C \quad (7)$$

The representation feature is multiplied by the element $\otimes$, F_S represents the spatial attention feature map, and F_C represents the channel attention feature map.

A. *Spatial attention module*

In the spatial attention model, first, after obtaining the background approximate value $V_{background}$ through global pooling, the original image and $V_{background}$ are passed through the background suppression function to get the feature map after background suppression, and finally through the maximum local pooling, the contour and details of the defect Information are enhanced. Second, the feature maps after background inhibition were convolved through expansion at different expansion rates to improve the receptive fields. Calculate the similarity between the current position and all other positions, and get the contribution degree of all other places to the current situation. Finally, the image and feature map processed by spatial attention are used to perform pixel-by-pixel multiplication to obtain feature maps with attention weights.

Because the contour and texture of the defect area are very different from the background, the module assigns greater weight to the contour and texture of the defect, thereby enhancing defect detection. The spatial attention module is shown in Fig. 2.

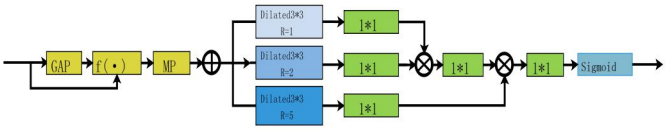


Fig. 2 Spatial attention module

The calculation steps of the spatial attention feature F_S diagram are as follows:

1) Conduct background suppression pooling on the feature map $F \in R^{C\times H\times W}$, as shown in formula _ (8);

$$\begin{aligned} dist(V_o, V_b) &= |V_o - V_{GA}| \\ F_1 &= F + [\frac{V_o - V_{GA}}{|V_o - V_{GA}| + \mu} \times S(dist(F, V_{GA}))] \\ S(x) &= \alpha \times \frac{1}{x+1} \end{aligned} \quad (8)$$

Vo represents the pixel values in the feature graph, and VGA represents the background approximation, S (.) Represents the inhibitory function, as the repressor.

2) The feature map after background inhibition is convolved through the expansion to enhance the receptive fields through different expansion rates;

3) The similarity of the current position to all other positions is then calculated to obtain the contribution of all other positions to the current position, as shown in Equation (9);

$$\begin{aligned} f_{c-Dilated1} &= Conv\{\underset{rate=1}{D}[F_1]\} \\ f_{c-Dilated2} &= Conv\{\underset{rate=2}{D}[F_1]\} \\ f_{c-Dilated5} &= Conv\{\underset{rate=5}{D}[F_1]\} \\ f_{T-Att1} &= f_{c-Dilated1} \otimes f_{c-Dilated2} \\ f_{T-Att2} &= Conv[f_{c-Dilated2}] \otimes f_{c-Dilated5} \\ F_S &= sigmoid\{conv[f_{T-Att2}]\} \end{aligned} \quad (9)$$

B. *Channel attention module*

Different receptive fields are required for defects of various sizes, so parallel dilated convolution is used to obtain other receptive fields and capture multi-scale information. Feature fusion is performed on feature maps of different receptive domains, and background suppression pooling is performed on the fused feature maps to suppress background information. After background inhibition pooling, MLP treatment after another 1 * 1 convolution was normalized using the Sigmoid function to obtain channel attention weights. After using the background inhibition pooling, the output feature map can better extract the features of the defects and differ significantly from the feature values of the background region, etc. The channel attention module is shown in Fig. 3.

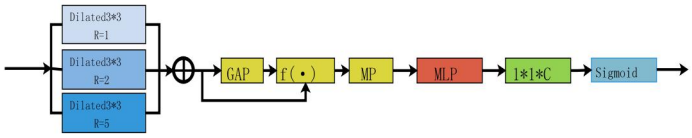


Fig. 3 Channel attention module

The calculation steps of the channel attention feature F_C diagram are as follows:

1) Use a parallel expansion convolution for the feature graph $F \in R^{C\times H\times W}$ to obtain different receptive domains and capture multi-scale information, as shown in Equation (10);

$$f_{c-Dilated} = C\{\underset{rate\in\{1,2,5\}}{D} Conv[F]\} \quad (10)$$

2) Fusion of the output of step 1);

3) Background suppression pooling of the fused feature map to suppress background information, such as formula (11);

$$F_2 = f_{c-Dilated} + [\frac{V_o - V_{GA}}{|V_o - V_{GA}| + \mu} \times S(dist(f_{c-Dilated}, V_{GA}))] \quad (11)$$

4) After background inhibition pooling, MLP treatment and then 1 * 1 convolution were normalized using the Sigmoid function to obtain the channel attention weight, as shown in Equation (12);

$$\begin{aligned} f_{ch-Att} &= MLP\{F_2\} \\ F_C &= sigmoid[f_{ch-Att}] \times F \end{aligned} \quad (12)$$

IV. Experimental Results and Analysis

The hardware configuration of this experiment is: Core ADM R5-3600, the memory is 16G, NVIDIA GeForce RTX 2060 GPU, and the software environment is CUDA11.0, cudnn8.0, Python3.7, and TensorFlow-gpu2.4.

A. Experimental dataset and the training model

In the field of steel surface defect detection, the commonly used image dataset is the NEU-DET dataset of Northeastern University. The dataset contains six categories of defects, namely plaque (patches), crack (crazing), hemp point (pitted surface), inclusion (inclusion), scratch (scratches), pressed into iron oxide skin (rolled-in scale). There were 300 images for each defect with 1800 images, each size 200200.

During the training process, the training and test sets were randomly divided by 8:1:1, with a batch size (batch size) of 16, learning rate decay with 0.001 as the initial learning rate, and optimized using the Adam optimizer.

B. Interpretation

YOLOv3 and YOLOv4 are algorithms that can balance detection accuracy and speed in small target detection. Therefore, to verify the influence of the improved algorithm on the detection results of steel surface defects, the proposed method was embedded into YOLOv3 and YOLOv4 models to obtain the enhanced YOLOv3 and YOLOv4 algorithms, and compare the improved YOLO and the original YOLO on the same data set. Mean Average Precision (mAP) and fps were used as performance indicators to evaluate the detection effect of the network model. The formula for the mean accuracy (mAP) is shown in formula (13), and fps is shown in (14).[21]

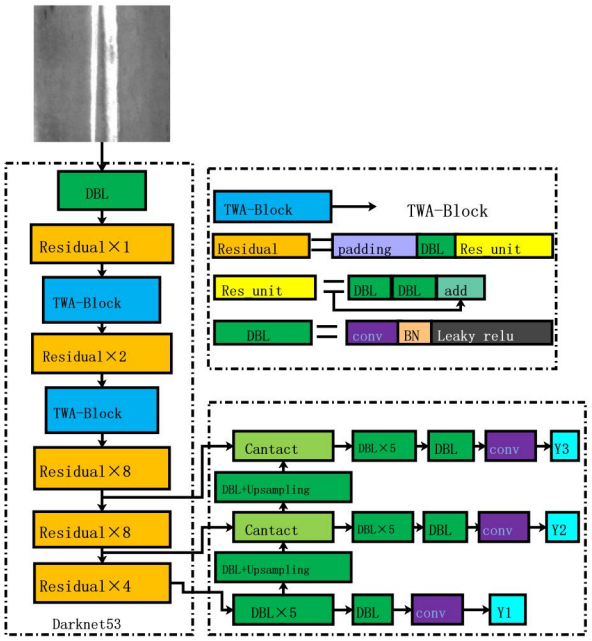
$$\begin{aligned} \mathrm{Re}call(R) &= \frac{TP}{TP+FN} \\ \mathrm{Pr}ecision(P) &= \frac{TP}{TP+FP} \\ AP &= \int_0^1 P(R)dR \\ mAP &= \frac{\sum_{i=0}^{n} AP(i)}{n} \end{aligned} \quad (13)$$

$$fps = \frac{NumFigure}{TotalTime} \quad (14)$$

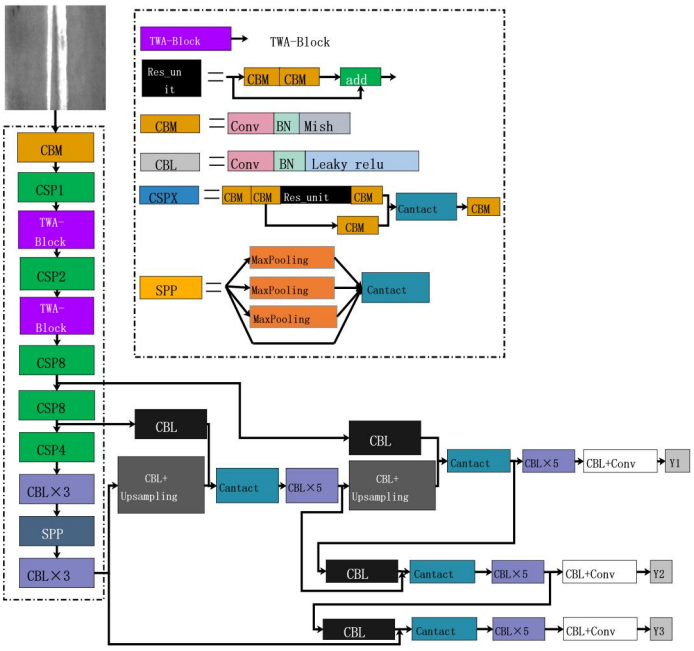
TP indicates the number of correct targets, FP represents the number of false targets incorrectly identified, FN represents the number of incorrectly detected targets, Recall (R) represents the recall rate, Precision (P) represents the accuracy, AP represents the detection accuracy of a category, n represents the total number of types, NumFigure represents the whole number of pictures, and TotalTime represents the total time spent in detection.

To illustrate the effectiveness of the improved algorithm, this paper designs related comparative experiments. They are the comparative experiments of the improved K-Means clustering algorithm and the traditional K-Means clustering in the YOLO model, the comparative experiments of the two-way attention model embedded in the YOLO model and the original YOLO model, and the comparison between the

improved YOLO detection algorithm and the standard target detection algorithm. YOLO model embedded with two-way attention model is shown in Fig. 4.



(a) Improved YOLOv3 model



(b) Improved YOLOv4 model

Fig. 4 Improved YOLO model

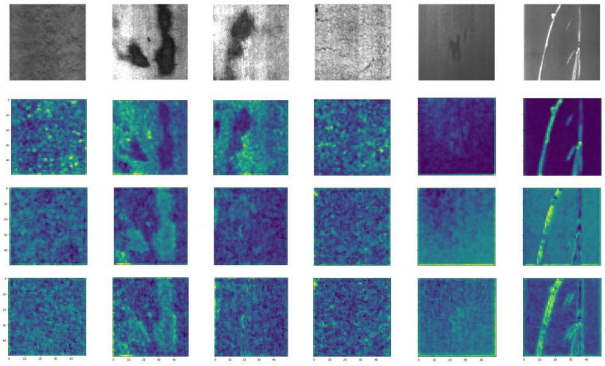
Table 1 shows the plot of experimental results obtained from the K-Means and traditional K-Means clustering algorithm applied in the original YOLO detection algorithm. The YOLO series's existing a priori box size is obtained through K-Means clustering on the COCO data set, but the NEU-DET [15] data set used in this article is quite different from the COCO data set in terms of target shape and size. In addition, the defects in the data set have diverse shapes and large scale changes. Therefore, the K-means clustering algorithm to optimize the cuckoo search algorithm on the NEU-DET dataset to obtain the anchor frame size suitable for this dataset. As can be seen from the table, the prior boxes and target boxes after K-Means clustering using the optimized cuckoo search are closer in size and shape and have more overlapping parts, thus improving the Recall and mAP of the model. The clustering algorithm integrated with optimized cuckoo search makes the clustering results more accurate and reduces the impact of falling into the local optimum. The final results showed that the K-Means clustering method with optimized cuckoo search improved 7.3% in mAP and 11.8% in Recall compared to conventional K-Means clustering methods. The proposed K-Means clustering optimized for cuckoo search effectively enhances the detection accuracy of some categories, reduces the impact of unbalanced samples on the clustering effect, and enhances the detection ability and applicability of the algorithm.

Table 1 The experimental results for the different clustering algorithms

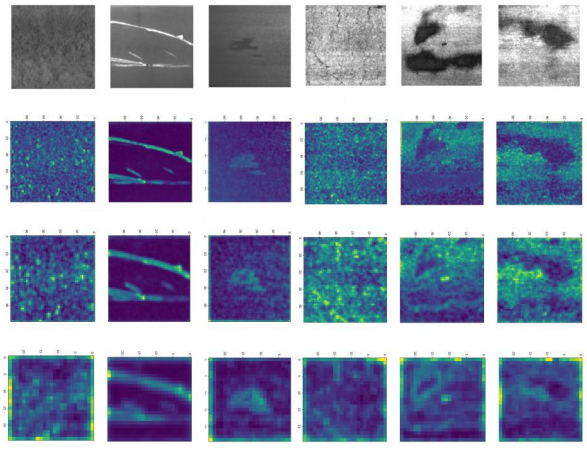
Object detection algorithm	Clustering Algorithm	mAP	Recall
YOLOv3	K-Means	65.1%	67.3%
YOLOv3	K-Means optimized for cuckoo search	74.2%	79.1%
YOLOv4	K-Means	76.5%	83.5%
YOLOv4	K-Means optimized for cuckoo search	77.3%	85.2%

The addition of two-way attention model (TWA-Block) to the shallow backbone feature extraction network increases the weight of defect features in the channel and spatial dimension of space, respectively, making the detection model pay more attention to defect targets in foreground areas and reduce the impact of complex background on defect detection. As shown in Fig. 5, they are the feature maps of the first three layers of the YOLOv3 model before and after the addition of attention, respectively. The two-way attention model is strengthened for the defect features similar to the background texture.Table 2 shows the experimental results obtained on the NEU-DET data set of the improved YOLO model, introducing the two-way attention model (TWA-Block) and the original YOLO model. It can be seen from the table that the mAP of

the YOLOv3 detection model with the attention mechanism is increased by 12.2% compared with the YOLOv3 model, and the parameter amount is only increased by 2%. The mAP of the YOLOv4 model with the attention mechanism increased and the number of parameters increased. Therefore, the addition of the two-way attention module (TWA-Block) has little impact on the detection speed of the network.



(a) Characteristics of no attention mechanism in the first three layers



(b) Attention mechanism features are shown in the first three layers

Fig. 5 the feature maps of the first three layers of the YOLOv3 model

Table 2 Comparison of improved performance of YOLOv3 algorithm

Detection model	YOLOv3	YOLOv3-TWA-Block	YOLOv4	YOLOv4-TWA-Block
Initialization method	Pre-training weight	Partial pre-training weights	Pre-training weight	Partial pre-training weights
Parameter	6.125×10^7	6.274×10^7	6.4×10^7	6.53×10^7
fps	50	45	65	62
mAP（%）	65.1	76.5	79.3	83.2

Table 3 shows the experimental results obtained by different detection algorithms on the NEU-DET data set. This table reflects the detection situation of the commonly used target detection algorithm and the improved YOLO network with the two-way attention model and the K-Means clustering algorithm optimized for cuckoo search in detecting steel surface defects. As shown by the experimental results, the mAP of the improved YOLOv3 model reached 81.7%, detection speed of 45fps, the mAP of improved YOLOv4 model reached 84.9%, and detection speed of 62fps. The improved model significantly improved the AP of crack (crazing), iron oxide skin (rolled-in scale) compared with SSD, Faster R-CNN and inclusion. However, since Faster R-CNN is a two-step detection framework, its detection speed is only 7 fps. Compared with the YOLO model, the improved model proposed in this paper improves by nearly 16.6% and 4.4% respectively in mAP, and only reduces the detection speed by 1 / 10. The improved YOLO model improves the accuracy and recall rate based on almost maintaining the original detection speed, and has a better detection effect on steel surface defects. Pictures of the detection results of each defect category were randomly selected from the experimental results are shown in Fig. 6, where green is the predicted value and blue is the actual value.

Table 3 Performance comparison of different algorithms

algorithm	mAP(%)	fps	AP(%)					
			patches	crazing	pitted surface	inclusion	scratches	rolled-in scale
SSD	72.3	55	85.78	48.41	78.36	67.96	84.85	68.69
Faster R-CNN	73.5	7	86.84	50.38	77.56	68.41	87.52	70.23
YOLOv3	65.1	50	82.04	40.86	70.84	58.12	82.38	56.44
Improve YOLOv3	81.7	45	88.92	72.32	86.13	74.36	90.57	77.64
YOLOv4	79.3	65	94.3	53.8	87.4	87.9	91.5	61.0
Improve YOLOv4	84.9	62	96.8	69.3	85.0	89.4	92.7	75.9
DDN+ResNet50	82.3	< 10	90.3	62.4	89.7	84.7	90.1	76.3

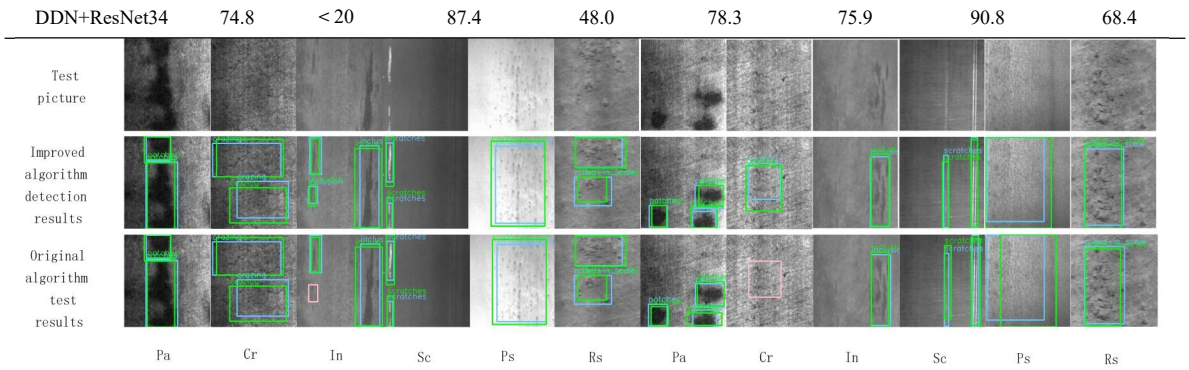


Fig. 6 Detection results of each defect category

V. CONCLUSION

Aiming at the current problem of steel surface defect detection, an improved YOLO detection algorithm is proposed. The algorithm uses the k-means algorithm to optimize cuckoo search to cluster the parameters of the target box, which makes up for the defect of low detection accuracy of traditional K-means clustering in unbalanced samples. At the same time, the two-way attention module (TWA-Block) was introduced into the model. This module uses background suppression pooling to solve defect textures and contours that are similar to background features and replaces ordinary convolution with hole convolution to expand the receptive field and enhance the detection of minor target defects. The verification results on the NEU-DET data set show that the improved method based on YOLO series models can effectively improve the detection accuracy of steel surface defects without significantly increasing the number of parameters. At the same time, it can meet the requirements of real-time detection. Finally, compared with YOLOv3 and YOLOv4 models, mAP improved by nearly 16.6% and 5.4%, respectively, and the detection speed remained within the range of 40-70fps.

ACKNOWLEDGMENTS

This work is supported by the National Science and Technology Major Project: Large-scale oil and gas fields and coalbed methane development(2017ZX05019003-002).

References

[1] QU E Q, CUI Y J, XU S, etc. Improved Gabor filter strip steel surface defect detection[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2017, 45(010):12-17

[2] FENG X Y. Two-stage target detection method based on deep learning and its application in surface defect detection[J]. Automation application,2020(8).

[3] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[4] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[5] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision & Pattern Recognition. IEEE, 2016.

[6] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer, Cham, 2016.

[7] YANG Q L、ZHOU B H、ZHENG W、LI M T. Dim target detection method based on fully convolutional recurrent network[J]. Acta Optics, 2020, 40(13):13.

[8] Jie H , Li S , Gang S , et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99).

[9] H. Li, L. Deng, C. Yang, J. Liu and Z. Gu, "Enhanced YOLO v3 Tiny Network for Real-Time Ship Detection

From Visual Image," in IEEE Access, vol. 9, pp. 16692-16706,2021,doi:10.1109/ACCESS.2021.3053956.

[10] LUO X Q, PAN S L. Improve YOLOV3's fire detection method[J]. Computer Engineering and Applications, 2020, v.56;No.960(17):192-201.

[11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, ‘‘YOLOv4: Optimal speed and accuracy of object detection,’’ 2020, arXiv:2004.10934. [Online]. Available: http://arxiv.org/abs/2004.10934

[12] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[J]. Springer, Cham, 2018.

[13] X Wang, Girshick R, Gupta A, et al. Non-local Neural Networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[14] L. Lin, "Improved K-means Algorithm Based on Genetic Algorithm", Electronic Technology and Software Engineering, vol. 01, pp. 111-112, 2020.

[15] J. Liu, L. Han and L. Hou, "K-Mean Clustering Algorithm Based on Particle Swarm Optimization", System Engineering Theory and Practice, vol. 06, pp. 54-58, 2005.

[16] R. Liu, H. Wang and X. Yu, "Shared-nearest-neighbor-based clustering by fast search and find of density peaks", Inf. Sci., vol. 450, pp. 200-226, Jun. 2018.

[17] Yu F, Bl B, Xl A, et al. Deep Learning-based Fast Recognition of Commutator Surface Defects[J]. Measurement, 2021.

[18] He Y, Song K, Meng Q, et al. An End-to-end Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features[J]. IEEE Transactions on Instrumentation and Measurement, 2019, PP(99):1-1.

[19] Lin T Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.

[20] Mcneely-White D, Beveridge J R, Draper B A. Inception and ResNet features are (almost) equivalent[J]. Cognitive Systems Research, 2020, 59:312-318.

[21] ZHU S W, HANG R L, LIU Q S. Underwater target detection based on class-weighted YOLO network[J]. Journal of Nanjing Normal University: Natural Science Edition,2020,43(1):129-135.