2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC) | 979-8-3503-0363-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/YAC59482.2023.10401826

Detection of small defects on gear surface based on improved YOLOv7

Jianghe Qiao, Yanqiu Che*, Guangyuan Gao
Tianjin Key Laboratory of Information Sensing & Intelligent Control
School of Automation and Electrical Engineering
Tianjin University of Technology and Education
Tianjin 300222, P. R. China
*yqche@tju.edu.cn

***Abstract*—In industrial quality inspection, false inspection or undetected inspection often occur in the identification of gear surface minute defects. This paper presents a novel detection framework for identifying surface defects on gears, utilizing an enhanced version of YOLOv7. Firstly, we add a small target detection layer upon the foundation of YOLOv7 to enhance the network's capability of accurately detecting smaller targets. Secondly, we replace the MP2 module with the MPCA module that integrates the CA (Coordinate Attention) mechanism, which can reduce the amount of computation and increase the network's attention to small targets by suppressing the interference of unnecessary information. Finally, for small and dense defects on the gear surface, the original Binary Cross Entropy Loss and CIoU Loss are substituted with the Varifocal Loss and SIoU Loss to train the network. The empirical findings demonstrate that the improved model significantly enhances the precision and recall rate of small defects detection. mAP@0.5 achieves 77.7%, exhibiting a notable improvement of 3.3% over the baseline YOLOv7 model, and better than other one-stage target detection networks.**

***Index Terms*—Gear defect detection, Small target detection layer, Attention mechanism, Varifocal Loss, SIoU Loss, YOLOv7**

I. Introduction

A. Background introduction

Gear system is an essential part of mechanical equipment [1], and its production quality will directly or indirectly affect the performance of mechanical equipment. Black patches and collisions are common surface defects in the gear production process. These defects are usually small and dense, which are easy to cause missed inspection in the quality inspection process, laying hidden dangers for the use of mechanical equipment [2]. So how to improve the recall rate while ensuring the accuracy holds immense importance in guaranteeing the production quality of gears.

At present, the algorithms for detecting surface defects are primarily classified into two distinct categories in application: approaches relying on the machine vision or advanced deep learning methodologies [3]. Shao et al. introduced a rapid and precise approach for detecting matching targets on gear surfaces, focusing on evaluating surface integrity. This method leverages digital image processing technology and employs the normalized cross-correlation coefficient [4]. Wang et al. put forth a computerized quantitative method to detect the state of external gears [5]. Firstly, a two-stage scheme is proposed to segment the meshing region of the gear teeth; Then the adaptive threshold and shape analysis are combined to detect surface defects. The appearance of convolution neural network has greatly promoted the utilization of deep learning methods in surface defects detection. Su et al. introduced a novel approach for detecting defects on gear end-face using an enhanced version of the YOLOv3 model [6]. Initially, a gray histogram-based technique for preprocessing gear images is introduced, followed by the development of a feature extraction backbone network called DWPW-ResNet34. Shi et al. proposed a technique for extracting regions of shaft gear components on the basis of Mask R-CNN [7]. To begin with, ResNet50 acts as the feature extraction network in order to extract the image's features. Subsequently, the feature pyramid network (FPN) is employed to amalgamate the information from feature maps of varying scales. Following this, the region proposal network (RPN) generates the regions of interest (ROI). Eventually, the head network is utilized to segment the targeted region. With sufficient light source and simple background, these methods have good detection effect on the surface defects of gears with fixed shape and size.

B. Algorithm introduction

In the dataset, the resolution of gear surface defects is usually less than 32 * 32, which belongs to the category of small targets. Although the above-mentioned methods enhance the precision of gear surface defect detection, they are not ideal for the identification of minute intensive defects on the surface of gears. A new model for detecting small defects on gear surface based on improved YOLOv7 is presented to solve the problems of weak information and easy to miss detection of dense small defects. By adding a small target detection layer to the YOLOv7 network, incorporating the attention mechanism and improving the loss function, the detection accuracy can be improved and the undetected rate can be reduced. First of all, we add a high resolution detector head to the detector head to detect undersized targets, which effectively improves the accuracy in identifying diminutive targets by the model. Secondly, we replace the MP2 module in the detection head

with the MPCA module integrated with the CA attention mechanism, so that the neural network allocates greater focus to the regions encompassing vital information within the targets and suppresses invalid information. Finally, we replace the original Binary Cross Entropy Loss and CIoU Loss by the Varifocal Loss and SIoU Loss to train the network, which enhances the significance of positive samples, diminishes the significance of negative samples, and concentrates the training process on superior-quality samples.

II. The YOLOv7 Algorithm

According to the different depth and width, YOLOv7 [8] model can be divided into YOLOv7-tiny, YOLOv7, YOLOv7X, YOLOv7-d6, YOLOv7-e6, YOLOv7-e6e and YOLOv7-w from small to large. The YOLOv7 model takes into account both speed and accuracy, so we chose YOLOv7 model as the experimental model. The YOLOv7 network architecture is shown in Fig. 1. At present, within the range of 5-160 FPS, the speed and accuracy of YOLOv7 model exceed all known target detection models.

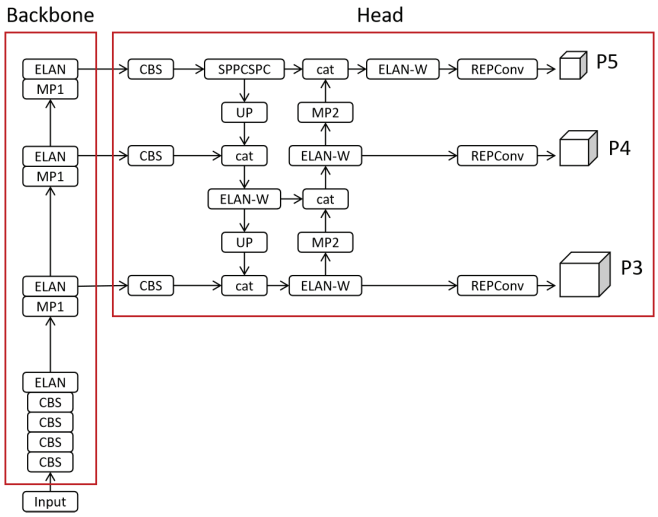

Fig. 1. An illustration of the YOLOv7 network architecture.

The YOLOv7 model consists of three components: Input, Backbone and Head. The backbone network consists of CBS, ELAN and MP1 modules. The CBS module consists of Convolution layer, Batch normalization layer and Silu activation function layer. According to the different size of convolution kernel (k) and step (s), CBS modules are divided into three types: (k=1, s=1), (k=3, s=1) and (k=3, s=2). The function of CBS module aims to change the size of input feature map while extracting noteworthy features. ELAN module consists of two branches. The upper branch is composed of a CBS module (k=3, s=1); The lower branch consists of 2 CBS modules (k=1, s=1) and 4 CBS modules (k=3, s=1). The structure of ELAN module is efficient. By manipulating the shortest and longest gradient paths, the convergence of the network can be accelerated and the generalization capability of the network can be enhanced. The architecture of ELAN module is shown in Fig. 2. The MP1 module is also composed of upper and lower branches. The upper branch consists of a maximum pooling layer and a CBS module (k=1, s=1). The lower branch consists of a (k=1, s=1) CBS module and a (k=3, s=2) CBS module. Both branches are used to perform subsampling, and adding up the outputs of the two branches gives a better subsampling result. The architecture of MP1 module is shown in Fig. 3.

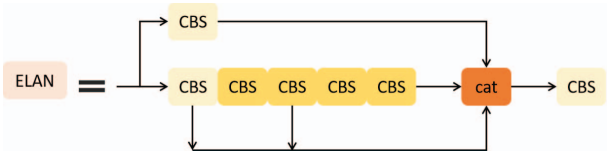

Fig. 2. Diagram of ELAN module architecture.

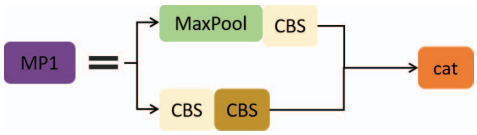

Fig. 3. Diagram of MP1 module architecture.

The head network adopts the PANet [9] structure, which is a top-down and bottom-up bidirectional fusion network structure. This structure can integrate features between different layers and maintain the integrity and richness of features. The head network consists of SPPCSPC, CBS, Up-sampling, Concat, ELAN-W, MP2 and REPConv modules. SPPCSPC is composed of upper and lower branches, and the upper branch consists of CBS modules and maximum pooling layers. By controlling the size of pooling core in the maximum pooling layer, we can obtain different sizes of receptive fields, so that the network can better distinguish objects of different sizes. The lower branch consists of only one CBS module. This structure of SPPCSPC can not only reduce the amount of computation, accelerate the convergence speed, and further augment the network's capacity for acquiring knowledge. The architecture of SPPCSPC module is shown in Fig. 4. ELAN-W and ELAN are basically the same in structure, but on the second branch, ELAN module selects three outputs to add, while ELAN-W module selects five outputs to add. The REPConv module draws on the idea of REPVGG [10] network. In order to minimize the model's parameters and enhance its performance, the REPConv module uses different network structures during network training and inference stage.

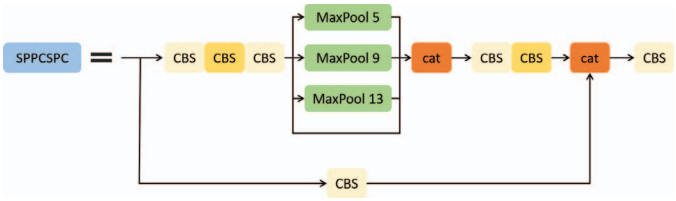

Fig. 4. Diagram of SPPCSPC module architecture.

III. IMPROVED YOLOV7 ALGORITHM

The YOLO algorithm, acclaimed as an exemplar of one-stage target detection algorithms, relies on deep neural networks for the identification and localization of objects. Its exceptional efficiency enables its application in real-time systems. YOLOv7 represents the pinnacle of the innovative YOLO series currently, surpassing its predecessors in both accuracy and speed. Although its accuracy and real-time performance are good, there are still some problems in small target detection in long distance and large background. Its feature extraction ability for small targets is weak, which is easy to lead to missed detection. So we have improved YOLOv7 model in terms of target detection layer, attention mechanism and loss function. The improved model architecture is shown in Fig. 5.

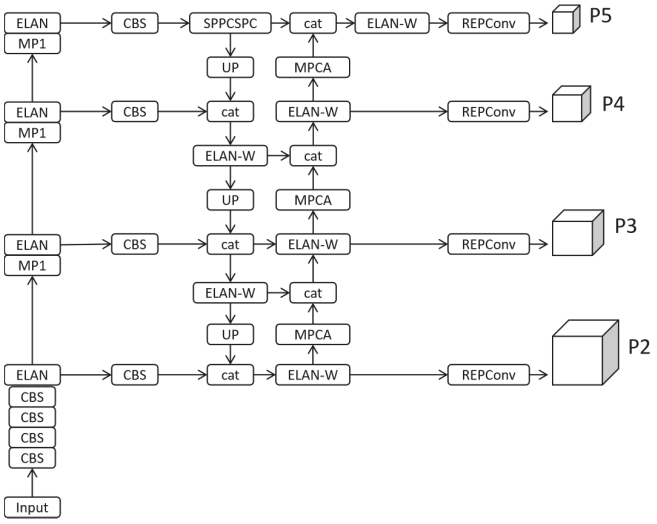

Fig. 5. An illustration of improved YOLOv7 network architecture.

A. Small target detection layer

Considering the small size of the surface defects of the gear in the data set, adding a small targets detector head [11] notably diminish the occurrence of undetected and erroneous detection of small defects. The original YOLOv7 model can detect at three scales, and the detection scale is relatively balanced. If a 640 * 640 size picture is input into the network, the final output scales of feature map are 80 * 80, 40 * 40 and 20 * 20. Respectively correspond to P3, P4 and P5 in Fig. 5, where Pi indicates that the resolution size is $1/2^i$ of the initial image size. In the gear surface defects data used in this paper, the pixel area of most defects is less than 0.12% of the original image area, so these defects are small targets. After many times of down-sampling, most of the characteristic information of such targets has been lost, and their detection remains arduous even when utilizing the P3 detector with exceptional resolution. With the purpose of enabling the network to detect these small targets better, we let the model do one more up-sampling and tensor splicing in the head network, so that the network can output an expanded feature map and reduce feature loss of these minute targets in the network. The improved model adds a detection scale, namely P2 as shown in Fig. 5. The detection scale is 160 * 160, which is more sensitive to small targets.

B. The MPCA module

In actual gear defects detection, gear defects have the characteristics of strong background texture, uneven illumination, more interference noise and weak defect information. In the process of detection, the characteristic information contained in different channels of the same layer of feature map is different. Conventional convolution operation can only collect local information in the image, lacking of global vision. Attention mechanism can enhance important targets feature information, make the model context-dependent, and suppress secondary feature information, which can enhance the accuracy of detecting gear surface defects.

The Coordinate Attention (CA) module [12] can be flexibly inserted into the classical detection network with little computing overhead. The CA module incorporates spatial details within the feature map into the mechanism of channel attention: The input characteristic map is decomposed into two feature maps by using two-dimensional global pooling operation, one is directional perception attention characteristic map and the other is position perception attention characteristic map. These two different direction perception attention characteristic maps are complementarily applied to input characteristic map. Therefore, CA module can capture the remote dependencies of feature maps from one dimension's spatial direction, while accurately retaining location details within the feature map along the spatial direction of another dimension to strengthen the feature representation of the improved network. The structure of CA module is shown in Fig. 6.

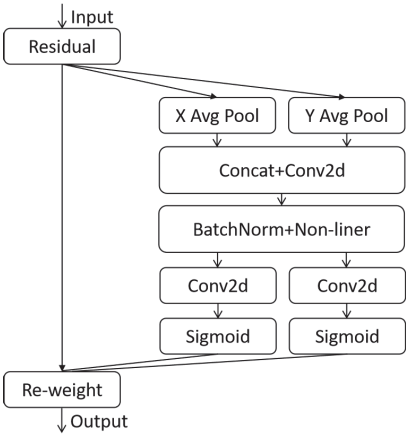

Fig. 6. The structure of CA module.

In order to better extract the targets features, we substitutes the MP2 module with the MPCA module in this paper: the CBS module (k=1, s=1) in the MP2 module does not alter the dimensions of the input feature map and has little impact on the network. So after replacing the CBS module (k=1, s=1) in the second branch with CA attention mechanism, it

can not only reduce the amount of computation, but also ensure the network to acquire further pivotal features and suppress unnecessary features. The structure of MP2 module and MPCA module is shown in Fig. 7.

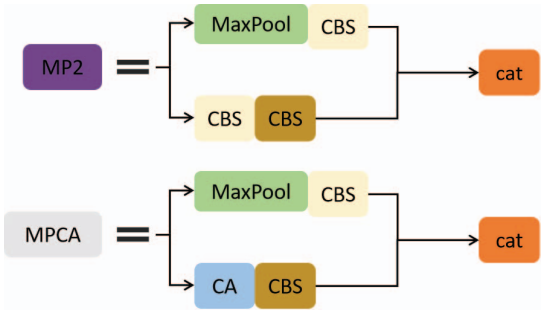

Fig. 7. The structure of MP2 module and MPCA module.

C. Improved loss function

Various defects in the gear data set are densely distributed, and there is a issue of sample disequilibrium between the defects and the background. At present, YOLOv7 uses the Binary Cross Entropy (BCE) loss function to compute the loss associated with classification and confidence. Simultaneously, the Focal Loss function can be employed to compute the loss, and the magnitude of the gamma value can be adjusted to tune the loss function. Although the binary cross entropy loss [13] improves the imbalance between positive and negative samples, there is no way to control the imbalance between dense samples due to its lack of distinction between difficult and easy samples. In an endeavor to address the issue of excessive disparity between the foreground class and background class during the training of dense targets detectors, the developer of Focal Loss has enhanced the BCE loss function. The following is the calculation formula of BCE Loss and Focal Loss [14].

$$BCE(P_t) = -\alpha_t \log(P_t) \tag{1}$$

$$FL(P_t) = -\alpha_t (1 - P_t)^{\gamma} \log(P_t) \tag{2}$$

In Equation (2), α_t is called the balance factor, and the weight of positive samples is increased when α_t decreases, which can ameliorate the imbalance between positive and negative samples. $(1 - P_t)^{\gamma}$ is employed for mitigating the imbalance between easy and difficult samples, where γ is greater than 0. As the confidence level of a positive sample increases, its weight decreases.

Varifocal Loss [15] draws on the weighting idea from Focal Loss and its calculation formula is shown below.

$$VFL(p,q) = \begin{cases} -q(q \log(p) + (1-q) \log(1-p) & q > 0 \\ -\alpha p^{\gamma} \log(1-p) & q = 0 \end{cases} \tag{3}$$

In Equation (3), p indicates the prediction score and the q is the IoU value between bounding box and ground truth box. Varifocal Loss scales the loss through p coefficient, which only reduces the weight of negative samples (q=0), without diminishing the loss contribution of positive samples (q>0). This is due to the scarcity of positive samples compared to negative ones, and we ought to retain these very precious learning signals. If the q value of a positive sample is large, its impact on the loss function will correspondingly increase, making the network pay more attention to the training of high-quality samples, thus improving the accuracy of dense target detection.

The original YOLOv7 model uses CIoU [16] as the loss function for bounding box regression. While CIoU Loss adequately considers factors such as the overlapping area, center point distance, and aspect ratio between the prediction box and the target box, it unfortunately neglects to account for the angular mismatch between the actual box and the predicted box. Thus, this paper introduces SIoU [17] Loss as the positioning loss function of bounding box regression. The loss function of SIoU consists of the Angle cost, Distance cost, Shape cost, and IoU cost. The formula of SIoU loss is as follows.

$$Loss_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{4}$$

In Equation (4), Δ is Distance cost considering the Angle cost condition and Ω is Shape cost. By introducing directionality in the cost of the loss function, SIoU diminishes the total degree of freedom concerning the loss function with effect. In contrast to the initial design, the refined model not only exhibits enhanced convergence during the training phase but also displays superior performance during inference.

IV. Experimental Validation

A. Experimental data and settings

The data set utilized in this research is the real data of the gear parts provided by FAW Hongqi Automobile in the production process. All the data are taken in the production line, with a total of 3400 pictures. The data set contains 3 classes of common gear surface defects: black patches on the top surface of the gear (hp_cm), black patches on the bottom surface of the gear (hp_cd), collisions on gear surface (kp). The partial enlarged view of these three defects is shown in Fig. 8.

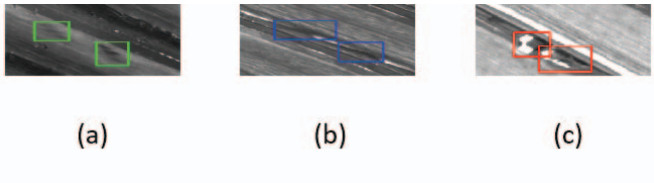

Fig. 8. Partial enlarged view of three kinds of defects, where (a) represents hp_cm, (b) represents hp_cd, (c) represents kp.

The initial data set is partitioned into a training set and a test set in an 8:2 proportion. Specifically, the training set consists of 2720 images, while the test set comprises 680 images. The number of labeled instances of each class in the original data set is shown in TABLE I.

TABLE I
NUMBER OF LABELED INSTANCES FOR EACH CLASS

Class	Number	Percentage
hp_cm	13151	31.5%
hp_cd	19363	46.4%
kp	9223	22.1%

The CPU of the computer utilized in the experiment is Intel (R) Xeon (R) CPU E5-2640 V4@3.60GHZ, NVIDIA Quadro M4000 graphics card with 8GB memory and Win10 operating system with 32GB memory. The software environment includes Pytorch1.8.1, CUDA11.3, and torchvision0.9.1. During network training, the initial learning rate of the network model is set to 0.01, the weight_decay is set to 0.0005, fl_gamma is set to 1.5, the batchsize is set to 8, and the epochs is set to 100.

B. Evaluation index

In order to assess the efficacy of the target detection algorithm, we employ P, R, AP, and mAP metrics to comprehensively evaluate the algorithm [18]. Precision (P) refers to the proportion of the true positive samples among the predicted positive samples, serving as an indirect evaluation of the model's false detection rate. Recall (R) signifies the proportion of the correctly predicted positive samples out of the total positive samples, which can indirectly evaluate the model's undetected rate. Average Precision (AP) represents the average precision of a category. Mean Average Precision (mAP) stands for the average precision across all categories. The calculation formulas are as follows.

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$AP = \int_0^1 P(r)\,\mathrm{d}r \tag{7}$$

$$mAP = \sum_k^n \frac{AP_k}{n} \tag{8}$$

In Equation (5) and Equation (6), TP indicates that positive samples are correctly predicted as positive samples; FP indicates that negative samples are incorrectly predicted as positive samples; FN indicates that positive samples are incorrectly predicted as negative samples.

C. Results and comparative analysis

In order to validate the efficacy of the algorithm discussed in this paper, based on the gear surface defects data set, YOLOv5 [19], RetinaNet [20], CornerNet and YOLOv7 are selected as the comparison models. The empirical outcomes are presented within the confines of TABLE II. TABLE II reveals that the mAP@0.5 of the improved YOLOv7 is much higher than that of YOLOv5, RetinaNet and CornerNet. In contrast to the initial YOLOv7, the increase of mAP@0.5 is 3.3%, which proves that the improved model has stronger detection ability for dense small targets. The mAP@0.5 curves of the enhanced YOLOv7 and the initial YOLOv7 on the gear surface defects data set are shown in Fig. 9.

TABLE II
COMPARISON OF DIFFERENT ALGORITHMS

Algorithm	AP			mAP@0.5
	hp_cm	*hp_cd*	*kp*	
RetinaNet	0.415	0.487	0.531	0.472
CornerNet	0.434	0.509	0.542	0.495
YOLOv5	0.598	0.632	0.673	0.634
YOLOv7	0.705	0.747	0.779	0.744
Improved YOLOv7	0.696	0.773	0.861	0.777

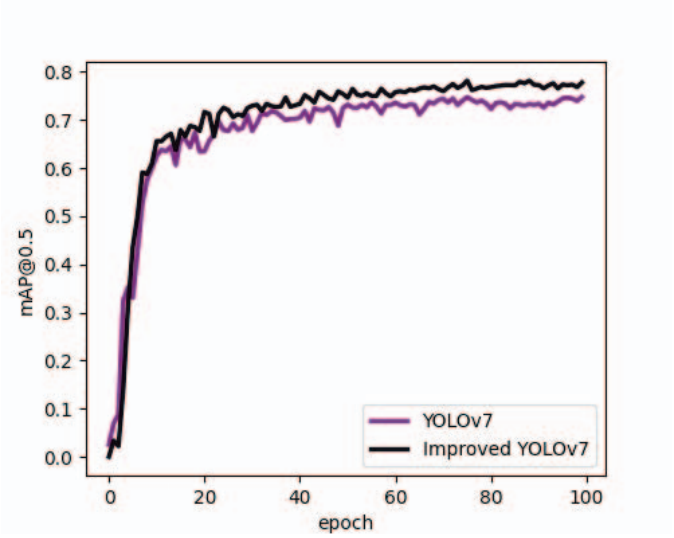

Fig. 9. Assessment of the mAP@0.5 curves between the enhanced and initial YOLOv7 network.

In Fig. 10, the paper shows some pictures of gear defects detection results, which demonstrates that the algorithm outlined in the study exhibits exemplary efficacy in detecting gear defects.

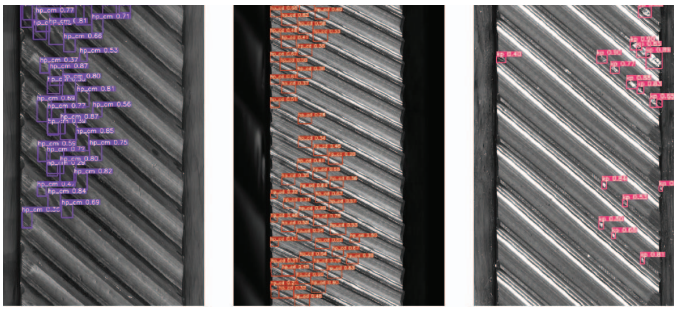

Fig. 10. Some gear defects detection results.

D. Ablation experiment

This section, based on the gear surface defects data set, explores the enhancement effect of new and improved modules

on the holistic model through ablation experiment. Based on the original YOLOv7, each module is added in turn for comparison. TABLE III lists the experimental results. According to TABLE III, the mAP@0.5 is increased by 1.2% after adding the small target detection layer, indicating that incorporating the high-resolution detection head can significantly augment the prowess of detection for small targets. On the basis of increasing the small target detection layer, adding MPCA module and improving the loss function separately, the mAP@0.5 has seen incremental improvements of 0.8% and 0.9%, respectively. When MPCA module is added and loss function is improved at the same time, mAP@0.5 increases by 2.1%. It shows that improving the disequilibrium between positive and negative samples while enhancing the feature information of important targets can better enhance the model's detection efficacy on dense diminutive objects.

TABLE III
RESULTS OF ABLATION EXPERIMENT

Moudle	mAP@0.5
YOLOv7	0.744
YOLOv7+P2	0.756
YOLOv7+P2+MPCA	0.764
YOLOv7+P2+Varifocal+SIoU	0.765
YOLOv7+P2+MPCA+Varifocal+SIoU	0.777

V. CONCLUSION

In this paper, YOLOv7 is improved for dense small target detection. First of all, a high resolution detector head has been integrated to detect undersized targets, which effectively improves the precision of the model in detecting diminutive targets. Secondly, we replace the MP2 module in the detection head with the MPCA module integrated with the CA attention mechanism, so that the neural network allocates greater focus to the regions of interest that encompass significant information and suppresses invalid information. Finally, the original Binary Cross Entropy Loss and CIoU Loss are replaced by the Varifocal Loss and SIoU Loss to train the network, which can amplify the significance of positive samples, diminish the influence of negative samples, and make the training focus on high-quality samples. The empirical findings demonstrate that this algorithm has certain advantages in the detection of dense small targets, and the mean Average Precision on multiple targets can reach 77.7%. Although in contrast to the initial model, the problem of undetected detection and false detection has been significantly improved, the detection effect of hp_cm is not ideal, which will be further explored in the future.

REFERENCES

[1] G. Cirrincione, R. R. Kumar, A. Mohammadi, S. H. Kia, P. Barbiero and J. Ferretti, "Shallow Versus Deep Neural Networks in Gear Fault Diagnosis," in IEEE Transactions on Energy Conversion, vol. 35, no. 3, pp. 1338-1347, Sept. 2020.

[2] F. Karpat, A. E. Dirik, O. Doğan, O. C. Kalay, B. Korcuklu and C. Yüce, "A Novel AI-Based Method for Spur Gear Early Fault Diagnosis in Railway Gearboxes," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey, 2020, pp. 1-6.

[3] Q. Ling and N. A. M. Isa, "Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey," in IEEE Access, vol. 11, pp. 15921-15944, 2023.

[4] W. Shao, Y. Shao, Q. Liu and R. Quan, "High-Speed and Accurate Method for the Gear Surface Integrity Detection Based on Visual Imaging," 2021 International Conference of Optical Imaging and Measurement (ICOIM), Xi'an, China, 2021, pp. 122-126.

[5] C. Wang, F. Li, Y. Li, H. Chen and X. Cao, "A Defect Status Detecting Method for External Gear in Railway," 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 2018, pp. 123-127.

[6] Y. Su and P. Yan, "A defect detection method of gear end-face based on modified YOLO-V3," 2020 10th Institute of Electrical and Electronics Engineers International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), Xi'an, China, 2020, pp. 283-288.

[7] Y. Shi et al., "Machining surface extraction method for shaft gear parts based on Mask R-CNN," 2021 IEEE International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT), Qingdao, China, 2021, pp. 75-79.

[8] S. Liu, Y. Wang, Q. Yu, H. Liu and Z. Peng, "CEAM-YOLOv7: Improved YOLOv7 Based on Channel Expansion and Attention Mechanism for Driver Distraction Behavior Detection," in IEEE Access, vol. 10, pp. 129116-129124, 2022.

[9] Y. Ma, "PANet: parallel attention network for remote sensing image semantic segmentation," ISCTT 2021; 6th International Conference on Information Science, Computer Technology and Transportation, Xishuangbanna, China, 2021, pp. 1-4.

[10] M. Nergiz, "Analysis of RepVGG on Small Sized Dandelion Images Dataset in terms of Transfer Learning, Regularization, Spatial Attention as well as Squeeze and Excitation Blocks," 2021 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 2021, pp. 378-382.

[11] K. -Y. Cao, X. Cui and J. -C. Piao, "Smaller Target Detection Algorithms Based on YOLOv5 in Safety Helmet Wearing Detection," 2022 4th International Conference on Robotics and Computer Vision (ICRCV), Wuhan, China, 2022, pp. 154-158.

[12] W. Li, S. Han and Y. Liu, "A Pedestrian Detection Algorithm Based on Channel Attention Mechanism," 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 2021, pp. 5954-5959.

[13] M. R. Rezaei-Dastjerdehei, A. Mijani and E. Fatemizadeh, "Addressing Imbalance in Multi-Label Classification Using Weighted Cross Entropy Loss Function," 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 2020, pp. 333-338.

[14] K. Doi and A. Iwasaki, "The Effect of Focal Loss in Semantic Segmentation of High Resolution Aerial Image," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018, pp. 6919-6922.

[15] X. Hu, B. Zhao, T. Sun, S. Cao, S. Fan and H. Xing, "A detecting algorithm for occlusion on the surface of photovoltaic modules based on improved YOLOv5," 2021 China Automation Congress (CAC), Beijing, China, 2021, pp. 3131-3135.

[16] S. Du, B. Zhang, P. Zhang and P. Xiang, "An Improved Bounding Box Regression Loss Function Based on CIOU Loss for Multi-scale Object Detection," 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 2021, pp. 92-98.

[17] Q. An, K. Wang, Z. Li, C. Song, X. Tang and J. Song, "Real-Time Monitoring Method of Strawberry Fruit Growth State Based on YOLO Improved Model," in IEEE Access, vol. 10, pp. 124363-124372.

[18] X. Hong, F. Wang and J. Ma, "Improved YOLOv7 Model for Insulator Surface Defect Detection," 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2022, pp. 1667-1672.

[19] B. He, J. Zhuo, X. Zhuo, S. Peng, T. Li and H. Wang, "Defect detection of printed circuit board based on improved YOLOv5," 2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT), Yichang, China, 2022, pp. 1-4.

[20] J. Sales, J. Marcato Junior, H. Siqueira, M. De Souza, E. Matsubara and W. N. Gonçalves, "Retinanet Deep Learning-Based Approach to Detect Termite Mounds in Eucalyptus Forests," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 586-589.