M-YOLO: A Detection Method of Substation Pointer Instrument

Zhaoming Hao, Meng Xu, Hongyan Li, Xiaoqiong Zhang, Ziyang Zhang, Weifeng Wang

Abstract—**With the advent of Industry 4.0, substations are undergoing intelligent upgrades, including the intelligent detection and reading of various instruments within the substations. To overcome challenges such as complex backgrounds, detection difficulties, and the requirements for high-speed and accurate detection, this paper introduces a detection algorithm called M-YOLO (Meter YOLO) for pointer instruments. The proposed algorithm is a modified version of YOLOv5, incorporating the following modifications. Firstly, a spatial attention mechanism called SAC3 is introduced in the backbone part to enhance features and reduce interference from complex backgrounds during instrument detection. Secondly, the C3 convolution of the backbone part is replaced by the improved RepBlock module to enhance the extraction of instrument feature information and reduce the loss of feature information. Additionally, the algorithm incorporates an RFB feature enhancement module to improve multi-scale prediction ability and enhances the BiFPN for improved instrument detection accuracy. The experiment shows that the mAP of M-YOLO reaches 95.9 %, which is 6.2 % higher than that of the original YOLOv5 network. It effectively detects dashboards in complex substation backgrounds while maintaining a speed of 62 FPS, enabling accurate reading of subsequent instrument data. In summary, the algorithm provides a practical and effective solution for the intelligent instrument detection of substations to meet the actual needs.**

Index Terms—**Substation pointer instrument, Instrument detection, YOLOv5, BiFPN**

I. INTRODUCTION

With the advent of Industry 4.0, substations are also upgrading towards intelligence and digitization [1]. In the context of China's electric power industry, the pointer instrument remains the primary tool for real-time monitoring due to its strong anti-interference capability and high reliability [2]. However, due to the lack of data output interface, the pointer instrument needs to be read manually, which has many problems, such as large manual error, low efficiency and poor real-time performance. Therefore, there is a pressing need to enable intelligent reading of pointer instruments in substations.

Since the 21 st century, the field of instrument detection has received extensive attention and development, and more and more related research has been carried out. There are two main categories. The first category encompasses traditional instrument detection methods based on computer vision techniques, while the second category consists of instrument detection methods based on target detection algorithm. Traditional instrument detection methods include approaches such as template matching, feature point matching algorithms, and support vector machines. For example, Sablatnig [3] proposed a method for extracting the instrument dial area through manual design. Peng [4] utilized the established LARK feature in conjunction with the PCA algorithm to accurately locate instruments by implementing a grid motion statistics method that effectively eliminates any mismatched points. Gao [5] improved the ORB feature matching algorithm for instrument identification. However, these traditional image methods rely on manually designed features, which are susceptible to interference from complex backgrounds and different types of instruments, leading to less than satisfactory detection results. While these methods offer fast detection speeds, their overall detection accuracy is often unsatisfactory.

With the rapid development of artificial intelligence, the focus of target detection has shifted to using the power of artificial intelligence. More and more visual problems, including instrument detection [6], are adopting deep learning methods [7]. Deep learning, based on machine learning techniques [8], allows for learning the necessary features to recognize objects from a large variety of labeled samples, eliminating the limitations of manual feature design. Moreover, deep neural networks have a high capacity to learn complex feature mappings, meeting the demands of instrument detection in complex backgrounds. Because the target detection method based on deep learning shows better robustness and high precision in complex environments, such as substations, it is widely used in various fields. At present, there are many kinds of target detection methods based on deep learning, but they can be divided into two categories: one-stage algorithms, such as YOLO [9], and two-stage algorithms, such as Faster R-CNN [10]. Generally, two-stage algorithms have higher accuracy in object detection compared to one-stage algorithms. However, the complex network structures of two-stage algorithms pose challenges

Manuscript received August 28, 2023; revised December 1, 2023. This work is supported by the National Key R & D Program (2021YFE0105000), Natural Science Foundation of China (52074213).

Zhaoming Hao is an Associate Professor of Electrical and Control Engineering, Xi’an University of Science and Technology, Xi’an 710054, China (e-mail: 2278285824@qq.com).

Meng Xu is a postgraduate student in Electrical and Control Engineering, Xi’an University of Science and Technology, Xi’an 710054, China (e-mail: xumeng9808@foxmail.com).

Hongyan Li is a Senior Engineer of Electrical and Control Engineering, Xi’an University of Science and Technology, Xi’an 710054, China (e-mail: lihongyan@xust.edu.cn).

Xiaoqiong Zhang is a postgraduate student in Electrical and Control Engineering, Xi’an University of Science and Technology, Xi’an 710054, China (e-mail: 2321255257@qq.com).

Ziyang Zhang is a postgraduate student in Electrical and Control Engineering, Xi’an University of Science and Technology, Xi’an 710054, China (e-mail: 1119308629@qq.com).

Weifeng Wang is a Professor of Electrical and Control Engineering, Xi’an University of Science and Technology, Xi’an 710054, China (e-mail: wangwf03@126.com).

in meeting the online requirements of instrument detection. Otherwise, one-stage algorithms can provide faster detection speed while still meeting the accuracy requirements of instrument detection. Consequently, more researchers choose to use one-stage algorithms for instrument detection. Jin [11] proposed the use of YOLOv5 to detect instrument dial areas, focusing on the problem of low accuracy in pointer instrument detection under different background and distance conditions. However, they did not make improvements to YOLOv5, and both accuracy and speed still need to be enhanced. Liang [12] introduced an improved Modi-YOLO v3-Tiny network by incorporating the feature extraction convolution layer from the VGGNet network into YOLOv3-Tiny, considering the network's lightweight and accuracy. However, it overlooks the interference of complex backgrounds in instrument detection. Tao [13] added residual modules to the YOLOv4-Tiny network structure to enhance the model's robustness, followed by instrument dial area detection. However, its detection performance is poorer in low-light backgrounds.

Although the one-stage algorithm performs better in the field of substation instrument detection, most of the existing instrument detection algorithms do not consider the influence of complex background on instrument detection. In complex environments such as substations, ordinary algorithms struggle to distinguish targets from the background, resulting in suboptimal detection results. Aiming at the shortcomings of the above substation instrument detection algorithm, this paper proposes an M-YOLO pointer instrument detection algorithm. Aiming at the instrument detection under a complex background, YOLOv5 is modified. Finally, the improved model shows better detection performance in the instrument detection under the complex background of the substation. The main contributions of this paper are:

1) A M-YOLO substation instrument detection model is proposed. The SAC3 spatial attention mechanism module is added to facilitate the detection of the pointer instrument in a complex background. So as to improve the multi-scale feature fusion quality of the network, the backbone network C3 is replaced with the RepBlock module. C3TR is placed in the last layer of the trunk to better extract global information. So as to improve the accuracy of instrument detection, we improve the BiFPN.

2) In this paper, the instrument pictures collected at the substation site are expanded to construct the substation pointer instrument data set.

3) The final experiments show that the proposed M-YOLO algorithm meets the requirements of substation pointer instrument detection. Without losing the detection accuracy of the one-stage target detection algorithm, the detection speed is improved, facilitates subsequent reading work, and achieves real-time requirements.

The residual sections of this article are structured as follows. The second part mainly introduces the YOLOv5 network structure. The third section focuses on the innovations proposed in this paper. The construction of the dataset and the evaluation metrics for experimentation are detailed in the fourth chapter. In the fifth part, the experimental results are obtained and analyzed. Finally, Chapter 6 summarizes the current work and outlines future directions.

II. YOLOv5 NETWORK STRUCTURE

YOLOv5 follows a specific architecture composed of input, backbone, neck, and detect components. It takes a three-channel RGB image with a feature size of 640×640×3 as input. To improve the target detection image background, data reinforcement, and other technologies are used, which also reduce the dependence of the model on batch size. The feature extraction part of YOLOv5 is mainly played by CSPDarknet53. It is a combination of multiple Conv and C3 modules with SPPF and bottleneck templates. The Conv template in this context refers to a combination of convolution, batch normalization (BN), and the activation function SiLU. On the other hand, the C3 convolution is a combination of the Conv, bottleneck module, and concatenation splicing. The neck of YOLOv5 utilizes the characteristic pyramid structure of PANet (Path Aggregation Network). This structure helps capture multi-scale features and improve detection performance. For reference. Depicted in Figure 1 is the YOLOv5 network architecture that can be referred to.

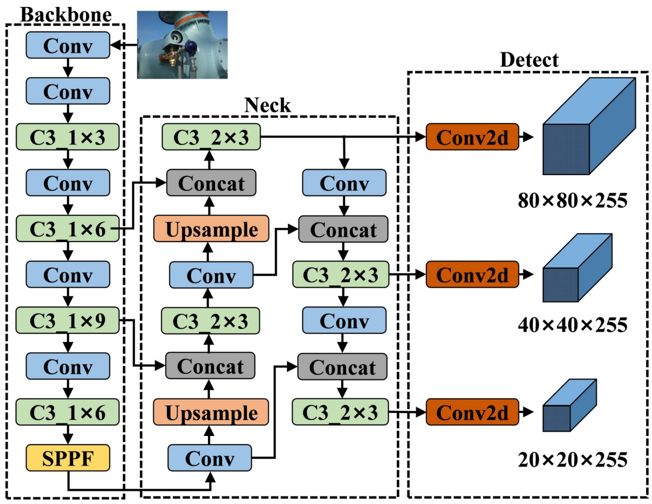


Fig. 1. YOLOv5 structure diagram

III. M-YOLO NETWORK STRUCTURE

In order to solve the problem of detecting pointer instruments in various backgrounds of substations, we have improved the network. In order to improve the ability of the network to detect instruments in complex backgrounds, SAC3 spatial attention is introduced. This module helps the network focus on important regions during the detection process. To increase the network's competency in fusing multi-scale features, the C3 module in the backbone network is replaced with the RepBlock module. This replacement enhances the network's capacity to capture features at different scales and improves its overall performance. To extract global information more effectively, the last layer C3 module of the backbone network is replaced by C3TR. This modification enables the network to better utilize global context information, which can be beneficial in instrument detection tasks. Finally, BiFPN is modified for instrument detection to improve the accuracy of instrument detection. The feature pyramid network is responsible for generating feature maps of three sizes, which is the key to detecting targets of different sizes. These modifications aim to improve the performance of the feature pyramid network, leading to improved detection accuracy. Figure 2 shows the structural details of the model M-YOLO in this paper.

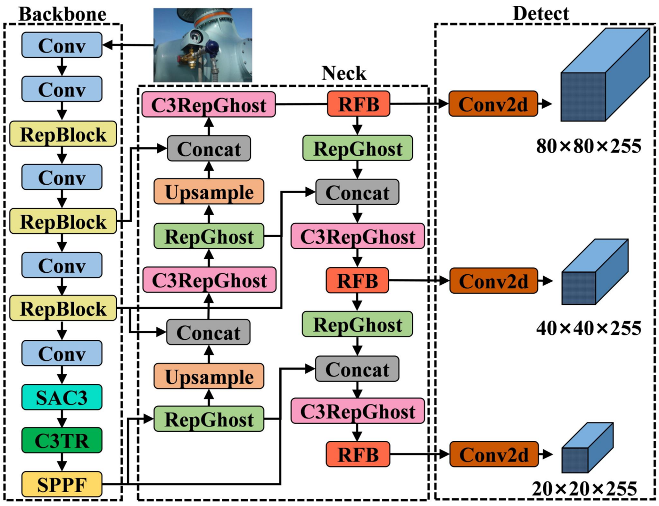


Fig. 2. M-YOLO structure diagram

A. SAC3 attention mechanism

In the context of complex backgrounds in substations, the feature information of pointer instruments may not be prominent enough, leading to potential detection misses. Therefore, it is crucial to strengthen detail extraction during the detection of pointer instruments. The attention mechanism is integrated into the detection network so that it can focus on the relevant information and ignore the irrelevant details. To achieve accurate detection of pointer instruments in the complex substation environment, an attention module can extract more instrument-specific details, enabling the network to prioritize instrument detection. This paper presents an improvement to the SAM (Spatial Attention Module) [14] and introduces a SAC3 (Spatial Attention Conv3d) attention mechanism module. The structural details of SAC3 are shown in Figure 3. By incorporating this attention mechanism module, the network can enhance its ability to capture instrument-specific details, enabling more focused and accurate detection of pointer instruments in the challenging substation environment.

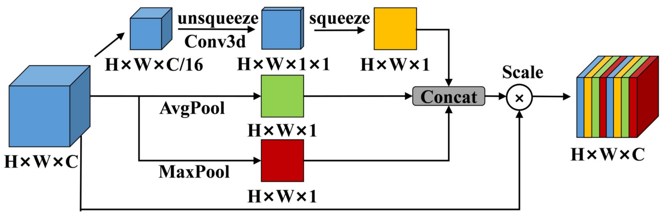


Fig. 3. SAC3 module structure diagram

The SAC3 attention mechanism first applies Conv3d, MaxPool, and AvgPool on each channel to obtain three H × W × 1 feature maps, and then concat the channel and send it to a standard convolutional layer. After activating the function, a Scale operation is performed with the input image, and the obtained channel weights are multiplied by the 2D matrix of the consistent feature map channel to obtain the result. This is the spatial attention SAC3, the formula is:

$$
\begin{aligned}
M_s(F) &= \sigma(f^{3\times3}([AvgPool(F);\\
&\qquad MaxPool(F);Conv3d(F)])) \\
&= \sigma(f^{3\times3}([F^s_{avg};F^s_{\max};F^s_{con}]))
\end{aligned} \tag{1}
$$

F^S_{avg} is the average pooling result, F^s_{max} is the maximum pooling result, F^s_{con} is the convolution result, and $M_S(F)$ is the final result.

The specific operation of Conv3d is as follows. Firstly, the amount of channels is reduced to C / 16 by convolution, and the amount of calculation is reduced. The size is H×W×C/16, and then the unsqueeze operation is performed to achieve the size increase effect. At this time, the image dimension is H×W×C/16×1, and the SiLU activation function is used. Then the squeeze operation is performed to decrease the image size to H×W×C/16, and the Sigmoid activation function is used. Compared with the SAM module, the SAC3 attention mechanism module adds the Conv3d processing process, which pays more attention to the spatial information of each Channel, which is conducive to extracting more instrument information in a complex background. Conv2d processing reduces computational complexity and enhances network nonlinearity.

B. Improved Rep VGG module

The RepVGG module is a straightforward architecture proposed in Reference [15], consisting of 3×3 Convolutional layers and ReLU activation. It is designed to be well-suited for GPUs and dedicated inference chips. RepVGG uses a structural re-parameterization method. The main operation is to use a multi-branch model similar to ResNet-style during training and convert it into a single-channel model of VGG-style during reasoning. During training, a multi-branch structure is employed during training, a multi-branch structure is employed, while during reasoning, the multi-branch structure is transformed into a single-branch structure through structural re-parameterization. The converted single-path model not only maintains the original inference speed but also achieves the progress of the multi-branch model. The structure of the RepVGG network training model and inference model is shown in Fig.4.

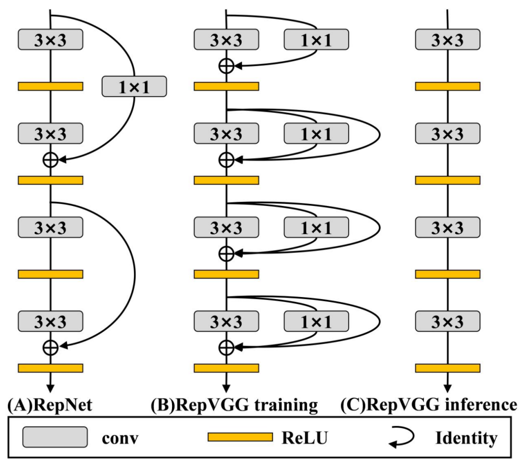


Fig. 4. RepVGG structure diagram

Li [16] proposed a method to increase the multi-scale feature fusion capacity and inference speed of the network by stacking the RepBlock module, which consists n the RepVGG blocks, as a replacement for the backbone C3 module. On this basis, we modify the activation function of the RepVGG network. In order to deepen the model and improve the accuracy, the activation function ReLU is replaced by the activation function SiLU, and the improved network is named RepVGG-S. Then the RepBlock module is formed by stacking n RepVGG-S modules. Figure 5 is the RepBlock structure diagram.

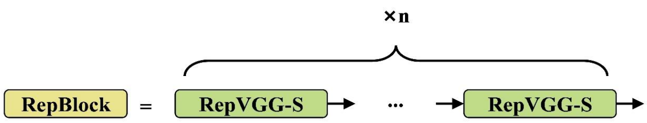


Fig. 5. RepBlock module structure diagram

In this paper, the YOLOv5 backbone C3 module is replaced by the above-improved module, which enhances the information extraction ability of the pointer instrument of the backbone network. The improvement deepens the network, greatly ensures the acquisition of pointer instrument feature information, and improves the inference speed.

C. Introduction of the RFB module

To improve feature expression and multi-scale prediction capability, an RFB feature enhancement module is incorporated into each output of BiFPN [17]. The structure of the RFB feature extraction module is primarily inspired by the inception concept, with hole convolution added on top of it to effectively increase the receptive field. The RFB module first constructs a multi-branch architecture with different sizes of convolutional layers and then uses extended convolutions to further increase the receptive field. The RFB structure is shown in Fig.6.

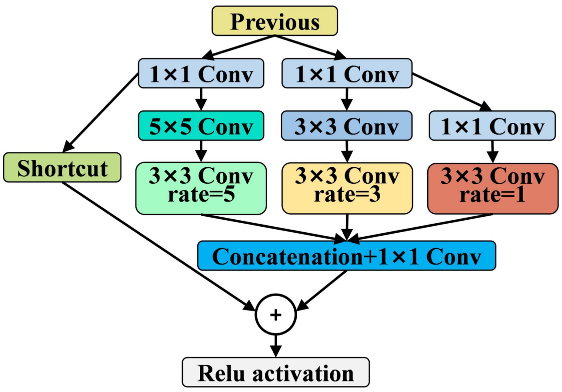


Fig. 6. RFB feature enhancement module structure diagram

D. RG-BiFPN

The conventional FPN [18] structure has a unidirectional flow of information from top to bottom, while the PANet network [19] incorporated in YOLOv5 incorporates a bottom-to-up pathway on top of FPN to enrich the information flow. This enhancement enables the retention of more shallow features effectively but does not take into account the different importance of different input features. To this end, the Google team posed a bipolar weighted feature pyramid structure (BiFPN) [20]. Considering that the resolution of the input feature image is different, the influence of the output feature map is different. When BiFPN is superimposed and fused at different levels of networks, learnable weights are introduced into each layer of the network, so that the network in the process of constantly adjusting the weights to learn different importance. However, due to the relatively complex structure and many parameters of the BiFPN module, the running speed of the network will be affected, which will eventually lead to a decrease in the detection speed of the instrument. In order to improve the detection speed, we have made the following lightweight improvements to BiFPN.

In 2020, Han [21] proposed a convolution module Ghost convolution, which is mainly divided into three steps, which are ordinary convolution and group convolution operation respectively. The third step is Identity, which adds the feature map obtained by ordinary convolution to the feature map obtained by group convolution.

Chen developed an efficient RepGhost module [22] through structural reparameterization technology. In order to save model inference time, a major improvement of RepGhost module is to replace Concat operation with the Add operation. The Add operation is performed first, and then the activation function ReLU is released to meet the structural reparameterization rules, which can be used for fast inference. The BN operation is placed in the identity mapping branch of identity mapping, which brings nonlinearity to the training process and can be merged for fast inference. The structure is shown in Figure 7.

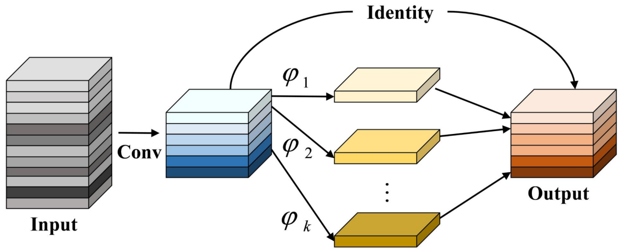


Fig. 7. RepGhost convolution module schematic diagram

In this paper, the BiFPN structure is improved, mainly replacing the ordinary convolution module and C3 module with the lightweight RepGhost convolution module and C3RepGhost module. The improved RG-BiFPN structure is shown in Figure 8 (b).

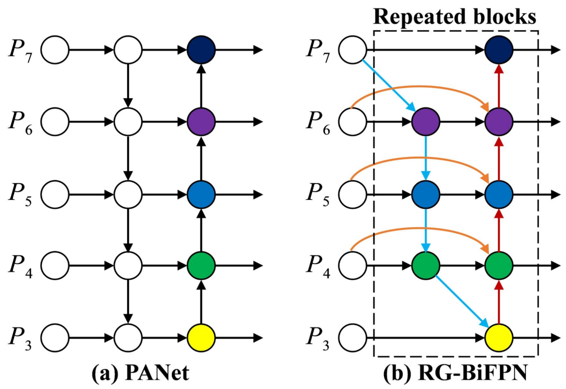


Fig. 8. PANet, RG-BiFPN network structure diagram

The RG-BiFPN network introduces learning weights for each path based on PANet and adjusts the weighted importance between different input features in the improved network, compared to the original feature extraction network of YOLOv5. To aggregate features of varying resolutions in each layer and increase the semantic representation of instrument image features, multi-scale feature fusion is accomplished through the use of a skip connection architecture. RG-BiFPN performs lightweight processing based on BiFPN, which improves the speed of the model. The RG-BiFPN feature fusion module replaces PANet and is used in the substation pointer instrument detection model in this paper for multi-scale feature fusion, which effectively connects the backbone part and the prediction part.

IV. Network training and testing

In this section, first, we introduce the source and composition of the dataset. Then the experimental environment and parameters are introduced. Finally, the evaluation index is briefly introduced.

A. Experimental data sets

The dataset used in this study consists of images captured on-site at a substation in Xi'an, Shaanxi Province, China. The images primarily contain pointer-type instruments. The on-site captured instrument images include six categories: SF6 pressure gauge, voltage meter, current meter, temperature gauge, pressure gauge, and transformer oil temperature gauge. To augment the dataset, several techniques were applied to the instrument images. These techniques included rotation, stitching, and mirroring to expand the dataset. Additionally, Gaussian noise and low illumination were introduced to certain portions of the dataset to enhance its generalization capability. The final dataset comprises a total of 4,237 images from the six instrument categories, captured under various backgrounds. To create a training and testing split, the training set and the test set are divided by 8:2.

B. Experimental environment and parameters

TABLE I
Units for Magnetic Properties

Lab environment	Environment configuration
Operating system	Windows 11
CPU	Intel i9-13900HX
GPU	NVIDIA GeForce RTX 4060
RAM	8G
Deep learning framework	Pytorch
CUDA version	11.8
Python	3.10

Table I shows the experimental environment parameters. The parameters of this experiment are based on YOLOv5. Under the support of experimental data, the optimal parameters were selected and some parameters were adjusted. After comparative analysis, this experiment in the network training, batch set to 4, epoch set to 200 when the model training is the best.

C. Experiment evaluation

We continue to use the common indicators of the target detection model, such as loss function curve, average precision (mAP), recall rate (Recall), and frame per second (FPS), to evaluate network performance.

Precision represents the percentage of the pointer instrument correctly predicting all instruments. The formula is:

$$P = \frac{TP}{TP + FP} \tag{2}$$

The quantity of correct instruments detected in the model prediction results is represented by the recall rate. It is expressed by the following formula:

$$R = \frac{TP}{TP + FN} \tag{3}$$

In the formula, TP represents the quantity of correctly detected instruments, FP represents the quantity of incorrectly detected instruments, and FN represents the quantity of unpredicted real instruments.

The average accuracy rate mAP is an intuitive standard for the detection performance results of a single category. The higher the value of mAP, the higher the detection accuracy of the model, and the better the detection effect in the complex background of the substation. The calculation formula (4) is:

$$mAP = \frac{\sum_{i=1}^{C} AP_i}{C} \tag{4}$$

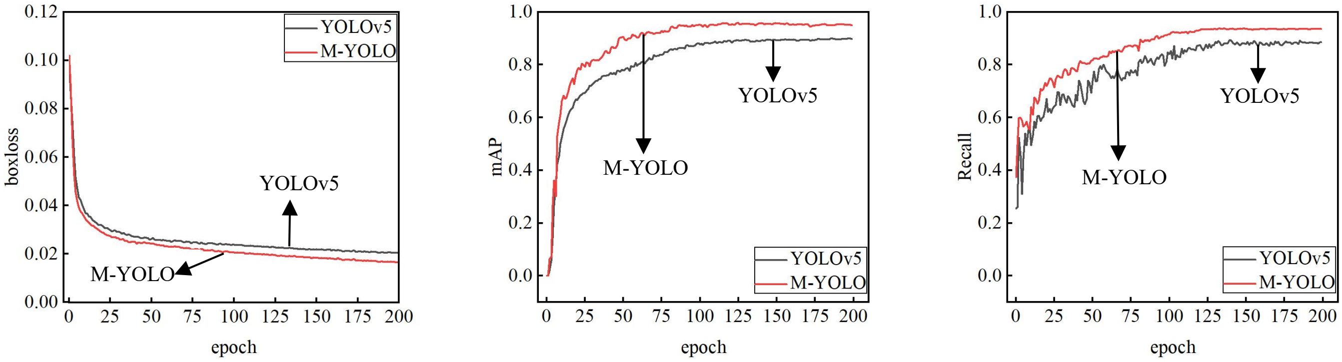
In the formula, C is the number of pointer instrument categories under the background of a substation in this paper.

V. Experimental results analysis

We conduct experiments under the data set of six types of instruments in this paper to show the effect of M-YOLO in instrument detection in substations. First, we compare the improved model with the original model. Additionally, we conduct an ablation analysis on various components of the improved method and visually represent the extent of the attention mechanism towards the target area. Then, this method is compared with other methods. Finally, the instrument detection results are visualized, and three sets of comparison diagrams further verify the effectiveness of the proposed method.

A. Model performance comparison experiment

Figure 9 shows the contrast of performance indicators between the YOLOv5 model and the M-YOLO model, including loss value, mAP, and Recall rate.



(a) boxloss (b) mAP (c) Recall

Fig. 9. Performance evaluation

Figure 9(a) illustrates that as the training progresses, the loss decreases and approaches zero for both models. However, the improved YOLOv5 model shows a faster convergence, with the loss value reaching closer to zero. This indicates that the proposed improvements lead to faster convergence during training. To evaluate the detection performance, the mAP metric is used, which measures the precision and recall of the instrument detection algorithm. Recall, as an indicator that can specifically represent the amount of instruments correctly detected by the algorithm, can be used to balance the two models. We compare the mAP and Recall values of the validation set of the two models during training. Figure 9(b) demonstrates that after 200 training iterations, the mAP of M-YOLO reached 0.959, while the original YOLOv5 model achieved only 0.897. This indicates an improvement of 6.2% in mAP with the proposed modifications. Similarly, Figure 9(c) shows that after 200 training iterations, the Recall of the improved model reaches 0.937, compared to 0.886 for the original YOLOv5 model. This represents a 5.1% increase in Recall with the improved model. These results demonstrate that the M-YOLO outperforms YOLOv5 in instrument detection, particularly in complex background scenarios. The model achieves faster convergence and higher detection accuracy, as evidenced by lower loss values, higher mAP scores, and improved Recall rates.

B. RepVGG improved contrast experiment

Based on the findings of Li [16], this paper proposes modifications to the activation function of RepVGG. These modifications are designed to facilitate model depth and enhance overall model accuracy. The following compares the mAP and Recall before and after the improvement to reflect the improved effect.

TABLE II
UNITS FOR MAGNETIC PROPERTIES

Method	mAP	Recall
Li [16]	0.903	0.893
RepVGG-S	**0.912**	**0.905**

As shown in Table II, after modifying the RepVGG activation function, mAP and Recall increased by 0.9 % and 1.2 % respectively under the data set of this paper. It shows that the accuracy of the model that modifies the activation function is higher under the same data set and device.

C. SAC3 Visualization Experiments

We improved the SAM attention mechanism, proposed the SAC3 attention module, compared the simulation results of SAM, CBAM, and SAC3, and compared the detection results. The results are shown in Figure 10.

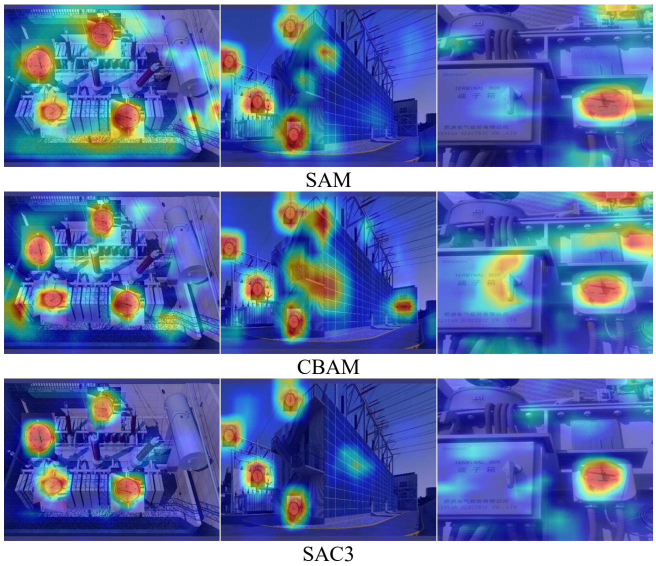


Fig. 10. Visualization results

In the first picture, SAM and CBAM show all the instruments, but they also wrongly focus on the substation background. In the second picture, CBAM wrongly focuses on the square background. In the third picture, both SAM and CBAM incorrectly focus on the error information above the picture. A comprehensive comparison of three different detection scenarios, the improved SAC3 pays more attention to the instrument information and ignores irrelevant information. SAC3 increases the processing of Conv3d so that the spatial information of each channel gets more attention. which is conducive to extracting more instrument information in a complex background. This improved attention mechanism helps to focus attention on important features while considering less important features, so that feature extraction is more comprehensive.

D. Ablation experiment

In general, in the parameter setting of model training, to better extract features and accurately detect targets, the confidence level is usually set very small. Therefore, the ablation experiment is executed under the condition of a confidence level of 0.6. After adding SAC3, RepBlock, and improved BiFPN to RFB, the models were named A, B, and C respectively, and compared with YOLOv5 to test the mAP of the model. The experimental results are recorded in Table III.

As shown in Table III. After adding SAC3 attention, the mAP value is increased by 2.1 %. After C3 is replaced by RepBlock, the mAP value is enhanced by 1.7 %. After adding the RFB module and optimizing the feature pyramid, the mAP value is enhanced by 2.6 %. After the combination of the three methods, YOLOv5 is added, which is the model of this paper, and the mAP value is enhanced by 6.2 %.

TABLE III
ABLATION EXPERIMENT

Method	SAC3	RepBlock	RFB＋RG-BiFPN	mAP
YOLOv5				0.897
A	√			0.918
B		√		0.914
C			√	0.923
M-YOLO	√	√	√	**0.959**

In summary, the improved methods in this paper can improve the detection accuracy, and the combination of the three has the best improvement effect. The M-YOLO model structure is more reasonable for the instrument detection of various complex backgrounds in substations and is more conducive to later work.

E. Contrast experiment with common models

In addition, under the data set of this paper, the mAP, Recall, and FPS of YOLOv3 [23], SSD [24], YOLOv7 [25], YOLOv8, Liang [12], Tao et al [13]and YOLOv5 in the same test environment were compared using the same device. Details are shown in Table IV.

TABLE IV
PERFORMANCE COMPARISON WITH COMMONLY USED MODELS

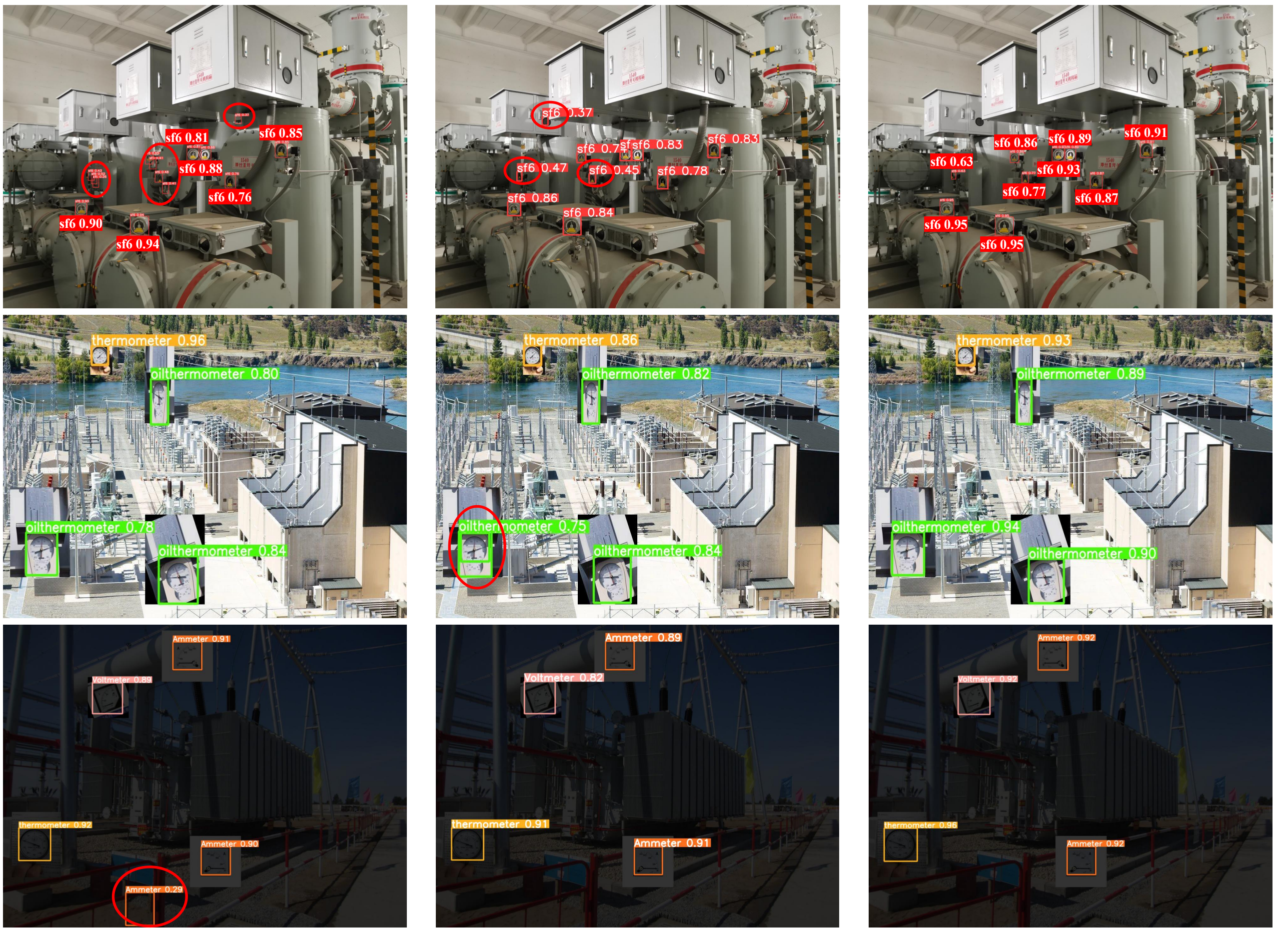
Method	mAP	Recall	FPS
YOLOv3[23]	0.879	0.809	26
SSD[24]	0.764	0.783	31
YOLOv7[25]	0.887	0.889	35
YOLOv8	0.896	0.860	48
Liang[12]	0.913	0.904	45
Tao et al[13]	0.921	0.911	43
YOLOv5	0.897	0.886	53
M-YOLO	**0.959**	**0.937**	**62**

As shown in Table IV, the M-YOLO model is compared with other commonly used detection algorithms. In the matter of mAP, Recall, and FPS, the detection effect of the instrument detection algorithm M-YOLO proposed in this paper is better than the mainstream one-stage algorithm and two-stage algorithm. The results show that the M-YOLO algorithm has better instrument detection performance in complex backgrounds.

F. Model detection effect comparison experiment

To have a more intuitive understanding of the improvement in algorithm performance, a selection of images was chosen to validate YOLOv5, YOLOv8, and M-YOLO. The instrument test results are shown in Fig.11.

In the first image, YOLOv5 had three false detections, while YOLOv8 incorrectly identified the circular background as an instrument and had two detections with confidence scores below 0.5. On the other hand, the proposed M-YOLO model accurately detected all instruments. In the second image, although YOLOv5 correctly detected the instruments, its confidence scores were generally lower than those of M-YOLO. Additionally, YOLOv8 had multiple detections for one type of instrument. In the third image with poor lighting conditions, YOLOv5 mistakenly detected a current meter, while YOLOv8 successfully detected all instruments but with lower confidence scores compared to M-YOLO. In conclusion, the YOLOv5 and YOLOv8 models exhibited lower accuracy in instrument detection and recognition in complex backgrounds, with issues of false detections and duplicate detections. On the other hand, the proposed M-YOLO model showcased remarkable performance in instrument detection, especially in challenging backgrounds, effectively meeting the practical demands.



YOLOv5 YOLOv8 M-YOLO

Fig. 11. Comparison of detection effect of complex background instrument

VI. Conclusion

This paper mainly aims at the poor detection effect of the pointer meter under the complex background of the substation, which brings difficulties to the subsequent meter reading. To improve the detection precision of YOLOv5 under complex background and decrease the case of false detection and missing detection. By adding improved spatial attention SAC3, the network pays more attention to effective features, and the backbone convolutional module is replaced by an improved RepBlock, which improves the feature information extraction ability of the pointer instrument of the backbone part and the loss of feature information is reduced. Finally, each convolution of BiFPN is replaced by the RepGhost convolution module, and the RFB module is introduced to improve the precision of the model. From the final experiment, the mAP of M-YOLO was increased by 6.2 %, and the FPS was 62. In the substation scenario, the detection speed and accuracy of the pointer instrument of the algorithm in this paper meet the requirements. However, improved instrument detection algorithms have lower detection accuracy in a single class of pointer instruments than in other classes. Therefore, subsequent texts will improve the data set and further optimize the algorithm to meet the detection of various instruments. Our improved attention mechanism relies heavily on visual cues. However, in scenarios where the instrument is partially obscured or blurred, the attention module may encounter challenges in generating precise attention maps, leading to suboptimal instrument detection. We are actively working on further improvements in this area to address the issues related to blur and occlusion and enhance the model's performance.

References

[1] Y. H. Song, Z. Z. Zhou, H. R. Zhang, H. H. Su, H. Zhang, and Q. Wang, “Instrument Identification Technology Based on Deep Learning,” *International Journal of Computational Intelligence and Applications*, vol. 20, no. 3, pp. 2150017, 2021.

[2] J. Li, “Research on reading recognition method of pointer instrument based on substation inspection robot,” M.S. thesis, Dept. Mechanical. Eng., Xinjiang University, Xinjiang, China, 2021.

[3] R. Sablatnig and W. G. Kropatsch, “Application constraints in the design of an automatic reading device for analog display instruments,” *In Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, pp. 205-212, 1994, Sarasota, FL, USA.

[4] G. Peng, B. Du, C. Cao, and D. He, “Pointer-type instrument positioning method of intelligent inspection system for substation,” *Journal of Electronic Imaging*, vol. 31, no. 1, pp. 013001-013001, 2022.

[5] J. L. Gao, L. Guo, Y. Y. Lv, Q. W. Wu, and D. Q. Mu, “Research on Algorithm of Pointer Instrument Recognition and Reading Based on the Location of the Rotation Center of the Pointer,” *2018 IEEE International Conference on Mechatronics and Automation*, pp. 1405-1410, 2018, Changchun, China.

[6] W. Ding and B. Zhang, “Detection and recognition method for pointer-type meter based on deep learning,” *In 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology*, pp. 1088-1091, 2021, Changsha, Hunan, China.

[7] F. Yang, Z. D. Ding, M. M. Xing, and B. Ding, “Overview of deep learning target detection algorithm improvement,” *Computer Engineering and Applications*, vol. 59, no. 11, pp. 1-15, 2023.

[8] H. C. Guo, “Applications of Machine Learning Technologies in Artificial Intelligence,” *Electronic Technology*, vol. 50, no. 10, pp. 294-296, 2021.

[9] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A Review of Yolo algorithm developments,” *Procedia Computer Science*, no. 199, pp. 1066-1073, 2022.

[10] S. Ren, K. He, R. Girshick, and J Sun, “Faster r-cnn: Towards realtime object detection with region proposal networks,” *Advances in neural information processing systems*, pp. 28, 2015.

[11] A. P. Jin, L. Yuan, D. Q. Zhou, and K. G. Yang, “A pointer meter reading recognition method based on YOLOv5 and U-net,” *Instrument Technique and Sensor*, no. 11, pp. 29-33, 2022.

[12] C. Liang, “Research on automatic recognition algorithm of substation pointer meters based on machine vision and deep learning,” M.S. thesis, Dept. Electrical. Eng., Shanghai Institute of Electrical Engineering, Shanghai, China, 2022.

[13] J Tao, W. W. Lin, L. Zeng, J. H. Zhang, Z. Y Zhao, Z. Y. Xu, and C. T Zhang, “Pointer meter reading recognition based on YOLOv4-tiny and Hourglass,” *Journal of Electronic Measurement and Instrumentation*, vol. 37, no. 5, pp. 1-10, 2023.

[14] X. Zhu, D. Cheng, Z. Zhang, S Lin, and J Dai, “An empirical study of spatial attention mechanisms in deep networks,” *In Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 6688-6697, 2019, Seoul, South Korea.

[15] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J Sun, “Repvgg：making vggstyle convents great again,” *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13733-13742, 2021, Virtual.

[16] X. J. Li, Y. M. Deng, Z. H. Chen, and X. He, “Improved YOLOv5 airport runway foreign object detection algorithm,” *Computer Engineering and Applications*, vol. 59, no. 2, pp. 202-211, 2023.

[17] S. Liu and D. Huang, “Receptive field block net for accurate and fast object detection,” *In Proceedings of the European Conference on Computer Vision*. pp. 385-400, 2018, Munich, Germany.

[18] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S Belongie, “Feature pyramid networks for object detection,” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117-2125, 2017, Honolulu, HI, USA.

[19] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8759-8768, 2018, Salt Lake City, UT, USA.

[20] M. Tan, R. Pang, and Q. V. Le, “Efficientdet: Scalable and efficient object detection,” *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10781-10790, 2020, Virtual.

[21] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C Xu, “Ghostnet: More features from cheap operations,” *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1580-1589, 2020, Virtual.

[22] C. Chen, Z. Guo, H. Zeng, P. Xiong, and J. Dong, “RepGhost: A Hardware-Efficient Ghost Module via Re-parameterization,” *arXiv e-prints*, arXiv: 2211.06088, 2022.

[23] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint* arXiv: 1804.02767, 2018.

[24] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” *In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam*, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, pp. 21-37, 2016.

[25] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464-7475, 2023, Vancouver, Canada.