ICIC Express Letters
Part B: Applications
Volume **14**, Number **9**, September **2023**

ICIC International ©2023 ISSN 2185-2766
pp. **927–933**

POINTER METER DETECTION BASED ON IMPROVED YOLOV5S

Changquan Song, Weicheng Xie*, Yefei Zhang, Xin Zhou
and Chunlin Yang

School of Electrical Engineering and Electronic Information
Xihua University
No. 999, Jinzhou Road, Jinniu District, Chengdu 610039, P. R. China
{ 198746715; 1344461017; 1151397315; 2457804510 }@qq.com
*Corresponding author: scxweicheng@mail.xhu.edu.cn

Received December 2022; accepted February 2023

Abstract. *Aiming at the problem that it is difficult to locate the meter in the image of the pointer meter with complex background, an improved pointer meter detection model on the basis of YOLOv5s is proposed. First, the ECA (efficient channel attention) module is added at the end of the backbone network of YOLOv5s to highlight the important features of the pointer meter while suppressing the general features, strengthening the network's ability to discriminate the pointer meter. Then the Complete-IOU Loss in the original network is replaced by Focal-EIOU Loss, which improves the regression accuracy of the bounding box. Finally, training and testing on the data-enhanced dataset, the results show that the mAP of the proposed model reaches 97.04%, Frames per second reaches 74.57. The model effectively improves the accuracy of pointer meter detection in complex backgrounds and can meet the requirements of real-time detection in substations.*
Keywords: Pointer-type meters, Complex background, ECA, Focal-EIOU Loss, YOLOv5s

1. **Introduction.** Most of the equipment in the substation is distributed outdoors, which is greatly affected by the environment. The pointer meters are widely used to measure the working condition of the substation equipment. Intelligent identification systems based on vision are widely used in substations, dramatically improving the efficiency of staff. However, most of the pointer meters are installed in complex environments surrounded by various pipelines and equipments, and there is often interference from the background when getting the meter images. To recognize the reading of the pointer meter image, we have to locate the meter from the meter image first, and then perform the reading recognition. Therefore, it is very necessary to accurately locate the meter from the pointer instrument image with the complex background.

Target detection algorithms based on deep learning are widely used because of their great feature extraction ability, high speed and high accuracy. Among them, the regression-based target detection algorithm, also known as the one-stage target detection algorithm, such as the YOLO series, has extremely fast detection speed and high detection accuracy. Therefore, it is favored by most researchers and applied in practical engineering after being improved.

Gu et al. [1] added a deformable convolution module [2] to the SSD (singleshot multibox detector) network structure. Although the detection speed of the network is improved, there is a problem of insufficient detection ability in the detection of small objects; Zhang et al. [3] improved Faster R-CNN, used VGG-16 as a feature extraction network, and added feature matching templates to improve detection accuracy; Liu et al. [4] added the number of anchors on the basis of Faster R-CNN, thereby improving the detection accuracy of the network. These two methods based on Faster R-CNN cannot meet the

DOI: 10.24507/icicelb.14.09.927

real-time requirements of substations in terms of detection speed and require high system performance.

Yin et al. [5] proposed Faster-YOLO, a joint network of deep random kernel convolutional extreme learning machine (DRKCELM) and double hidden layer extreme learning machine auto-encoder (DLELM-AE) as a feature extractor, which reduces parameter settings and improves training speed; Yuan et al. [6] improved the feature extraction network Darknet-53 in YOLOv3 [7] by replacing the original residual module with the ResneXt [8] residual module and introducing dense connections to optimize the network structure and improve the efficiency of feature extraction. Although these two improved methods have a certain improvement in detection speed, they still cannot meet the requirements of real-time and some accuracy is lost. Guo et al. [9] optimized YOLOv4 by PID (proportional-integral-derivative) and applied the optimized framework to road damage detection. However, the performance of the improved YOLOv4 was barely satisfactory.

The latest version of the YOLO series detection network YOLOv5 has reached a new height in detection speed, and the YOLOv5s model is only a dozen MB, which is easy to deploy.

Therefore, it is improved on the basis of YOLOv5s. First, the attention module of ECA [10] is introduced, and the local cross-channel interaction without dimensionality reduction is realized by adaptively selecting the size of the one-dimensional convolution kernel, which improves the feature extraction capability of the network model. Secondly, the loss function CIOU Loss [11] in the original network is replaced by Focal-EIOU Loss which solves the problem of sample imbalance in bounding box regression and the fuzzy definition of CIOU aspect ratio. A pointer meter detection method based on improved YOLOv5 is proposed, which provides a new method for rapid and accurate detection of pointer meter in substations.

The rest of this paper is organized as follows. Section 2 introduces the network structure of YOLOv5. Section 3 describes the structure of the ECA attention module and how to add this attention module in YOLOv5. Section 4 introduces the loss function used in this paper, Section 5 shows the experimental results and discussion, and Section 6 gives the conclusion and future research directions.

2. **YOLOv5 Algorithm Principle.** The one-stage target detection algorithm YOLOv5 is improved on the basis of YOLOv4 that greatly increases the detection speed and accuracy. It consists of four parts: the input (Input), the backbone network (Backbone), the network layer (Neck) and the output (Prediction). The Input side has the functions of Mosaic data enhancement, adaptive anchor box calculation, and adaptive image scaling. The backbone network includes a slice structure (Focus), a convolution module, a bottleneck layer (C3), and a spatial pyramid pooling structure (SPP). The Neck layer is a feature fusion network, which consists of a feature pyramid (FPN) and a path aggregation network structure (PAN). The Prediction layer performs multi-scale target prediction. The whole block diagram of the YOLOv5 algorithm is shown in Figure 1.

3. **Introduction of ECA Attention Module.** The channel attention mechanism SENet module and the mixed domain attention mechanism CBAM module inevitably increase the computational burden in order to obtain better performance. ECANet is an ultra-light attention module for improving the performance of deep CNNs, overcoming the paradox of performance-complexity trade-off. The ECA module, although involving only the k ($k = 9$) parameter, brings significant gains in performance. The ECA module avoids the side effects of unnecessary dimensionality reduction on channel attention prediction, and realizes local cross-channel information interaction through one-dimensional convolution with adaptive convolution kernel size. Calculation formula of convolution kernel size is

$$k = \left| \frac{\log_2 C + b}{\gamma} \right|_{odd} \tag{1}$$

Among them, k is the size of the convolution kernel, C is the number of channels, γ and b are 2 and 1 respectively, and *odd* means that k can only be an odd number.

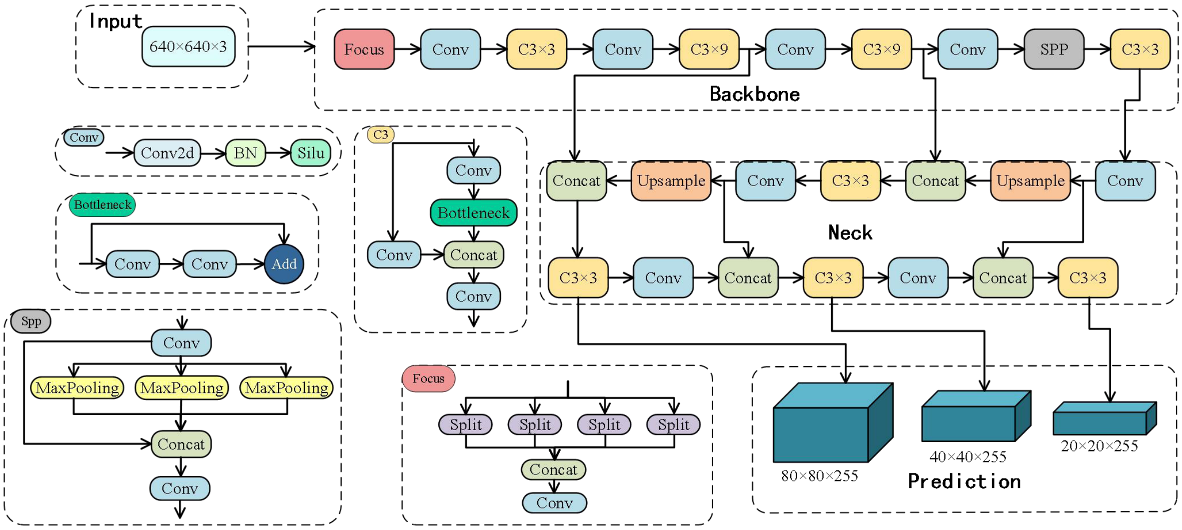

FIGURE 1. Whole block diagram of YOLOv5

The ECA module structure is shown in Figure 2, where H and W are the height and width of the features, and C is the number of channels. In the experiment, an ECA attention module is added at the end of the last C3 layer in the backbone network, and the structure of the backbone network after adding is shown in Figure 3, where C3 is a continuous three conv structure.

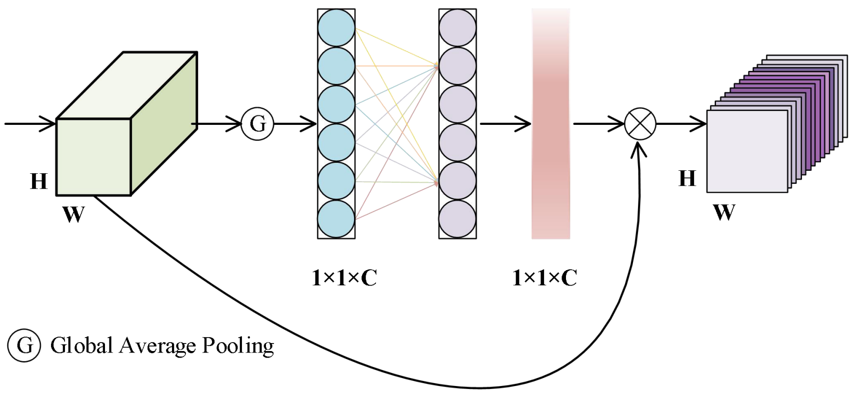

FIGURE 2. ECA structure

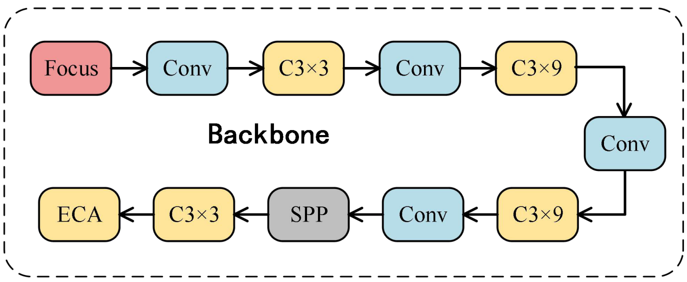

FIGURE 3. Improved backbone network structure

4. **Replacing the Loss Function.** The loss function of the original YOLOv5 is CIOU Loss. Although it considers the overlapping area, aspect ratio and center point distance of the bounding box regression, it reflects the difference in aspect ratio, not the real difference between the width and height and their respective confidences, sometimes appear to prevent the model from effectively optimizing for similarity. The Focal-EIOU Loss [12] solves this problem, while adding Focal Loss [13] to focus on high-quality anchor boxes. The Focal-EIOU Loss is expressed as follows:

$$L_{\textit{Focal-EIOU}} = IOU^{\gamma} L_{EIOU} \tag{2}$$

where γ is the parameter controling the level of outlier inhibition, and L_{EIOU} is expressed as follows:

$$L_{EIOU} = 1 - IOU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \frac{\rho^2\left(w, w^{gt}\right)}{C_w^2} + \frac{\rho^2\left(h, h^{gt}\right)}{C_h^2} \tag{3}$$

Among them, b and b^{gt} represent the predicted frame and the real frame respectively, $\rho\left(b, b^{gt}\right)$ means the Euclidean distance between the two bounding boxes' centre points, c is the diagonal distance of the minimum closed area that includes both bounding boxes, and $\frac{\rho^2\left(w,w^{gt}\right)}{C_w^2} + \frac{\rho^2\left(h,h^{gt}\right)}{C_h^2}$ is the width and height loss, where C_w and C_h are the width and height of the smallest external rectangle that covers both bounding boxes.

5. **Experimental Results and Analysis.** The improved model and the original YOLOv5 network are trained and tested on the self-built data set, and the experimental results are analyzed and evaluated.

5.1. **Dataset expansion.** The dataset of pointer meters in the real substation environment uploaded by other users is downloaded from the Internet, and it only contains 500 images, which is far from enough for network training. Therefore, data expansion is carried out by means of image filling, mirroring, flipping, rotation, etc., as well as their arrangement and combination. The expanded data set contains 2989 images. We use LabelImg for manual labeling, mark those that belong to the pointer meter as the meter class, and those that do not belong to the none class, and normalize the data size to 640 $\times$ 640 $\times$ 3 and align them.

Image filling Rotating+mirroring+filling Filling+rotating

FIGURE 4. Data expansion methods

5.2. **Experimental setup.** The software environment of the experiment is 64-bit Windows 10 operating system; Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz; NVIDIA GeForce GTX 3090 Ti (24GB); memory 64GB; CUDA 11.5; OpenCV 4.5.1 development platform. The dataset is randomly divided into training set and test set with a ratio of 9 : 1.

According to the size of the GPU's video memory, set the number of pictures to be processed in each batch to 64. To prevent the loss during training from generating oscillations or exploding gradients, the initial learning rate was reset to 10^{-5} after inputting images. Combined with the above parameter settings, considering the final convergence of the model, set the number of training to 200 times. Finally, save the weight file of the trained network model, and the performance of the network model is evaluated using a test set.

5.3. **Network performance evaluation metrics.** The Precision (P), the Recall (R), the mean Average Precision (mAP), the Frames Per Second (FPS) and the model volume are selected as the main evaluation indicators. The specific formulae are as follows:

$$P = \frac{TP}{FP + TP} \tag{4}$$

$$R = \frac{TP}{FN + TP} \tag{5}$$

$$AP = \int_0^1 P(r)dr \tag{6}$$

$$mAP = \frac{1}{n}\sum_{m=1}^{n} AP_m \tag{7}$$

TP is the number of positive samples predicted correctly, FP is the number of negative samples predicted incorrectly, FN is the number of positive samples predicted incorrectly, and AP is the average precision. P represents the prediction accuracy in the positive sample results, R is the proportion of correct predictions to all actual positives, and mAP reflects the accuracy of the model. The higher the value of mAP, the higher the accuracy of the model. FPS represents the number of images processed per second. The bigger the FPS, the faster the model can detect.

5.4. **Ablation experiment.** To verify the effectiveness of the improved method on the YOLOv5s model, the improved model and the original YOLOv5s were used for ablation experiments on the augmented dataset. The original YOLOv5s network, the network after adding the ECA attention mechanism, and the network after replacing the loss function are used for model training. Under the condition that the IOU value is 0.6, the samples of the test set are used to predict each version of the model. The results of the experiment are shown in Table 1. It can be seen that after introducing the ECA attention module alone, the accuracy rate P is increased by 5.7%, and mAP@0.5 is increased by 3.17%. After replacing the loss function alone, the accuracy P increased by 5.9%, and mAP@0.5 increased by 4.01%. After introducing the ECA attention module and the replacement loss function at the same time, the precision rate P is increased by 8.4%, the recall rate is increased by 0.3%, and the mAP@0.5 is increased by 4.48%.

There is a multitude of equipment in the substation, and it is better to make a false detection than to miss it. Therefore, the higher the recall rate of the model is required, the better. The above experiments show that adding the ECA attention mechanism module

Table 1. Ablation experiment

	Module		Precision/%	Recall/%	mAP@0.5/%
	ECA	Focal-EIOU			
YOLOv5s	×	×	79.2	96.4	92.56
	√	×	84.9	95.8	95.73
	×	√	85.1	96.1	96.57
	√	√	**87.6**	**96.7**	**97.04**

and replacing the loss function with Focal-EIOU Loss in the YOLOv5s network model can improve the accuracy of the model while ensuring a high recall rate, and improve the accuracy of substation pointer meter detection.

5.5. **Comparison of different network models.** In trying to validate the effectiveness of the improved YOLOv5s model in this paper, it was compared with YOLOv4 and the two-stage network Faster R-CNN on the same dataset and the same configuration environment, and the comparison results are shown in Table 2. From the results in Table 2, we can see that the improved model is optimal compared to the other methods in terms of mAP and P, reaching 97.04% and 87.6% respectively, and is only 0.4% and 2.35 away from the optimal value in R and FPS, respectively.

TABLE 2. Comparison of different network models

Network models	Precision/%	Recall/%	mAP@0.5/%	FPS/(frame/s)
YOLOv5s	79.2	96.4	92.56	**76.92**
YOLOv4	77.5	**97.1**	95.73	71.42
Faster R-CNN	82.9	94.7	96.57	72.31
Improved YOLOv5s	**87.6**	96.7	**97.04**	74.57

The model in this paper improves detection accuracy while keeping a faster detection rate. It has obvious advantages over the other 3 models.

Finally, the improved model was used on the test set for detection and to show some of the data. As shown in Figure 5, it can be directly seen that the model can accurately detect pointer meters even in complex backgrounds.

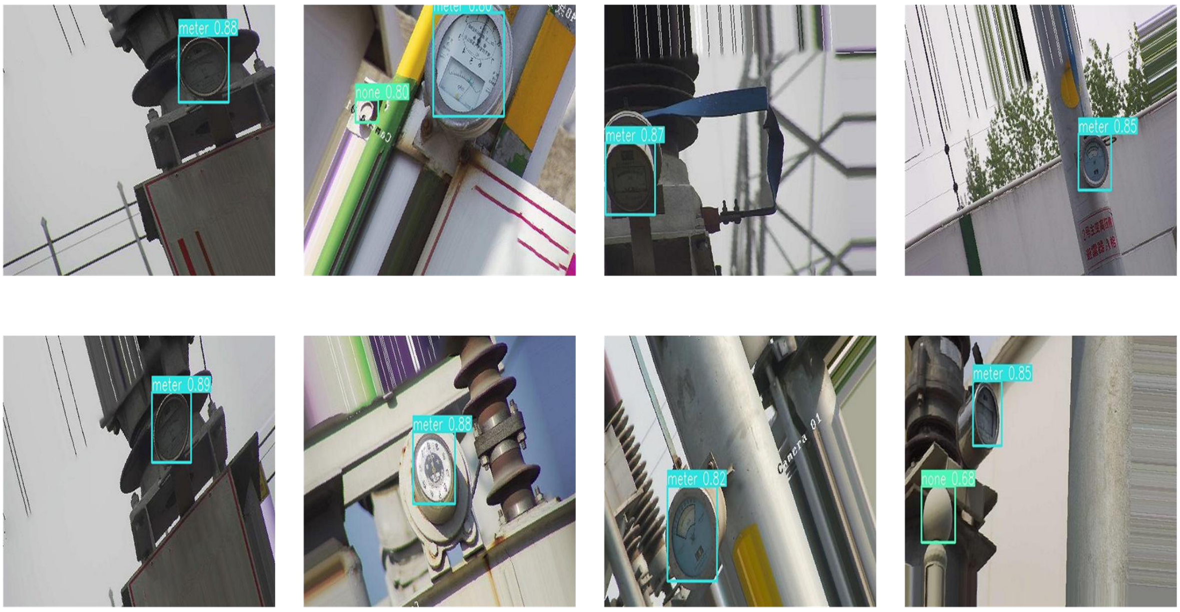

FIGURE 5. Detection results

6. **Conclusions.** This paper proposes a pointer meter detection method based on an improved YOLOv5 model to address the problem that pointer meters images photographed in complex environments of substations have complex backgrounds, from which it is difficult to locate pointer meters. The ECA attention module is added to the original YOLOv5 network structure to strengthen the feature extraction ability, and the CIOU Loss in the original network is replaced by the Focal-EIOU Loss to improve the bounding box regression accuracy. The substation pointer meter dataset was expanded by data enhancement

for the training and testing of the network model. Experimental results on the test set show that the mAP@0.5 of the improved network model reaches 97.04%, precision reaches 87.6%, and from the FPS of the model, it can be calculated that the detection time of a single pointer meter image is 0.013 s, which meets the actual needs of the substation.

This method replaces the manual visual inspection method in detecting the presence of pointer-type meters for substation inspection needs, which significantly improves the efficiency of substation inspection and is of great significance. In prospective work, we will investigate how to improve the robustness of the network to apply it to the positioning of various equipment in substations.

REFERENCES

[1] H. Gu, Y. Yang and Y. Qu, General target detection method based on improved SSD, *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Chongqing, China, pp.1787-1791, 2019.

[2] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu and Y. Wei, Deformable convolutional networks, *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp.764-773, 2017.

[3] X. Zhang, X. Dang, Q. Lv and S. Liu, A pointer meter recognition algorithm based on deep learning, *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, Shenzhen, China, pp.283-287, 2020.

[4] Y. Liu, J. Liu and Y. Ke, A detection and recognition system of pointer meters in substations based on computer vision, *Measurement*, vol.152, 107333, 2020.

[5] Y. Yin, H. Li and W. Fu, Faster-YOLO: An accurate and faster object detection method, *Digital Signal Processing*, vol.102, 102756, 2020.

[6] X. Yuan, X. Ma and S. Liu, Improved pedestrian and vehicle object detection algorithm in YOLOv3, *Science Technology and Engineering*, vol.21, pp.3192-3198, 2021.

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

[8] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, Aggregated residual transformations for deep neural networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp.5987-5995, 2017.

[9] L. Guo, R. Li and B. Jiang, A road surface damage detection method using YOLOv4 with PID optimizer, *International Journal of Innovative Computing, Information and Control*, vol.17, no.5, pp.1763-1774, 2021.

[10] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp.11531-11539, 2020.

[11] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye and D. Ren, Distance-IoU Loss: Faster and better learning for bounding box regression, *Proc. of the AAAI Conference on Artificial Intelligence*, pp.12993-13000, 2020.

[12] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang and T. Tan, Focal and efficient IOU loss for accurate bounding box regression, *arXiv.org*, arXiv: 2101.08158, 2021.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, Focal loss for dense object detection, *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp.2999-3007, 2017.