*Modern Management based on Big Data III*
*A.J. Tallón-Ballesteros (Ed.)*

*doi:10.3233/FAIA220085*

# A Lightweight Meter Detection Method Based on Yolov5

Junren SHAO[a], Jianglong YU[a], Lei LONG [a], Tao YANG[a], Anyun SHI[a], Qingshi YANG [a], Ning WANG [b], Tianqing XIA [c] and Shuaitong ZHANG [d,1]
[a] *China Southern Power Grid Co., Ltd. Kunming Bureau of EHV Transmission Company, Kunming, China*
[b] *Operation and maintenance center of Information and communication, CSG EHV Power Transmission Company, Guangzhou, China*
[c] *CMCCSI Co., Ltd., Beijing, China*
[d] *North China Electric Power University, Baoding, China*

**Abstract.** With the deepening of industrial automation, a large number of edge intelligent devices are deployed in industrial meter detection. In view of the limited computing and storage capacity of these embedded devices, we propose a lightweight meter detection method. Our proposed method is based on the widely used Yolov5, the depthwise separable convolution and squeeze and excitation channel attention module are used to simplify the backbone and head of the network, and further prune the filters of convolution layers via geometric median. Finally, model parameters and floating-point operations are reduced to 0.250M and 0.687G on the premise of ensuring the effect of the meter detection.

**Keywords.** meter detection, lightweight, depthwise separable convolution, channel attention mechanism, filter pruning

## 1. Introduction

In the environment with electromagnetic interference, digital electronic meter are prone to failure. Considering the factors of anti-interference ability, production structure and cost, pointer meter is widely used in many fields of society. Additionally, due to the layout design of pipelines, many meters are in positions that are difficult to be observed by humans, and need to be recognized with the help of embedded devices such as UAV and robot. Laroca et al. applied convolutional neural network to automatic meter reading, but the method based on Fast-Yolo has a slight deficiency in accuracy[1]. He et al. used Mask-RCNN to detect the meters, but the model is complex and huge, and the inference is poor in real time[2]. Fang et al. used ResNet-18 as the backbone of Mask-RCNN to reduce the network volume and meter detection time[3]. Zhang et al. selected Yolo network for rough positioning, and then used SSD network for fine positioning of instruments[4]. Li et al. utilized Mobilenetv2 to simplify the network, but its complexity still needed to be processed by the server, and only one category of meter was tested[5]. These methods are limited to the backbone network proposed for other general datasets,

[1] Corresponding Author, Shuaitong Zhang, North China Electric Power University, Baoding, China; E-mail: zst9512@126.com.

and the model parameters and calculations are not optimized according to the difficulty of instrument data. Considering that most of CNN's architectures are application specific[6], we focus on redesigning the backbone structure and further verifying the reduced network capacity through channel pruning.

In this work, aiming at the problem of multi-class meter detection, we improve Yolov5[7-10] which has been proposed recently and proved to be generally effective in a variety of production environments, to ensure the accuracy and complexity of the model. Specifically, the main research contents are shown as follows. 1) We collect the meter dataset in the real production environment, and mark the bounding box position and category labels. 2) We improve the original Yolov5 structure by introducing depth separable convolution[11, 12] and squeeze and excitation channel attention module[13] to reconstruct the network backbone and head, which reduces the parameters and computation of the original network and enhances the feature extraction efficiency of network parameters. 3) For the improved Yolov5 model, we use the popular filter pruning via geometric median method[14] to automatically adjust the hyperparameters of the network channel according to the model training results, so as to obtain a more lightweight model and almost no influence on mean average precision.

## 2. Methodology

The research methods of this paper are mainly carried out in three parts. Firstly, the overall architecture of Yolov5 is analyzed and introduced, and then the structure of network parameter redundancy is located. Secondly, the backbone and head structure with large parameters are simplified through depth separable convolution and squeeze and excitation channel attention module. Thirdly, the number of redundant channels in the network is removed by filter pruning via geometric medium method, and the mean average precision of the model is restored by fine-tuning.

### *2.1. Yolov5 Network Architecture*

Yolov5 is a recent version of Yolo architecture series, which is widely used in various object detection scenes. Compared with the previous Yolov4, the volume of network is reduced by nearly 90% in the Yolov5. Therefore, it can be deployed to some devices with relatively limited resources. According to the difference between the number of feature extraction modules and internal convolution layer channels, Yolov5s, Yolov5m, Yolov5l and Yolov5x are derived by the complexity of the model from low to high.

Taking Yolov5s as an example, the specific structure is shown in Figure 1. The network architecture includes backbone and head parts. The backbone is mainly responsible for extracting multi-scale image features, and the head detects objects of various sizes on three type of feature maps (80×80, 40×40 and 20×20) to obtain information of bounding box (x: horizontal coordinates, y: vertical coordinates, w: width, h: height, conf: classification confidence, cls: classification category).

### *2.2. Improved Network Architecture on Backbone and Head*

The original Yolov5 is designed based on the large-scale object detection dataset—coco, which requires more model parameters to fit the data samples. Considering the meter detection problem we solved, there are few types and quantities to be distinguished, so we can use less convolution channels to extract features.

In addition, we also use depth separable convolution composed of depthwise convolution (DWConv) and pointwise convolution (PWConv) to compress the model parameters and calculation, and construct the channel attention mechanism (Squeeze and Excitation module, SE) to further enhance the efficiency of model feature extraction. According to the experiment and experience, the layer by layer output channels of the improved backbone are set as 32, 32, 64, 64, 128, 128, 256 and 256, and the processing channels in head are also reduced to 32, 64 and 128, as shown in Figure 2.

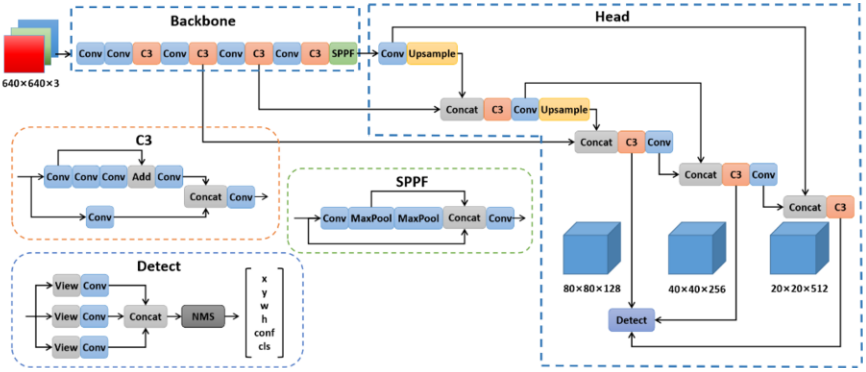

**Figure 1.** Original structure of Yolov5s.

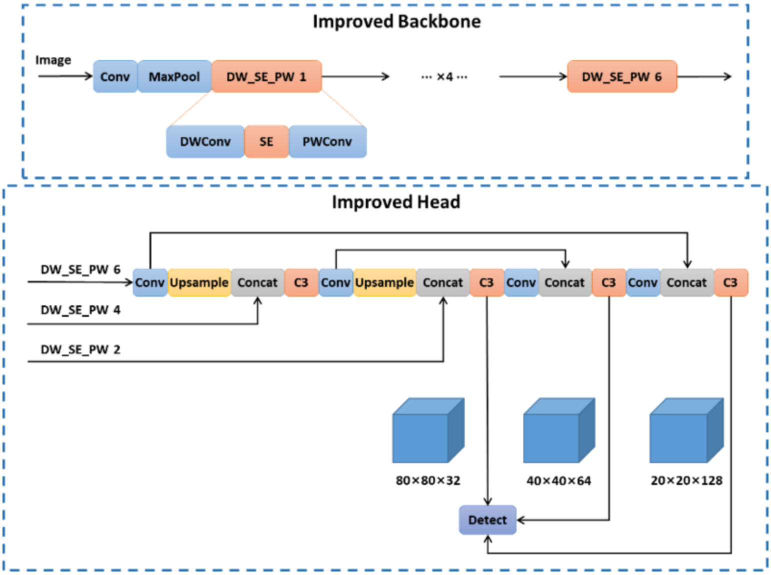

**Figure 2.** Improved structure of backbone and head.

### 2.3. Channel-pruned Network with FPGM

Some redundant parameters inevitably exist in the network structure designed by handcraft., so it is necessary to modify the network structure automatically in view of the training results to remove redundancy. Filter pruning via geometric median is a popular structured pruning method, which removes redundant channels in the network convolution layers based on geometric median.

In detail, for a certain convolution layer, the parameter weight tensor of each filter is arranged in descending order based on 1-norm, and then several channels with the minimum sum of Euclidean distance accumulation between each filter will be removed. The number of channels pruned in each layer is divided into seven equal sections by the ratio we set. As the number of channels in the network increases layer by layer, we increase the proportion of pruning gradually, the basic features learned from shallow convolution can be retained when pruning as many redundant channels as possible. For example, our model has 52 convolution layers divided into seven parts: 1-8, 9-15, 16-23, 24-30, 31-38, 39-45 and 46-52. The pruning rates of each segment are about 0.05, 0.1, 0.15, 0.2, 0.25, 0.3 and 0.35 respectively. Since the number of segments cannot be divided by 7 evenly, the proportion of each segment increases or decreases slightly to maintain an average of 20%.

## 3. Experimental results and analysis

### *3.1. Dataset*

Deep Learning algorithms can obtain a great degree of perfection in front of large datasets[15], so we collect numbers of meter data from the transmission lines of China Southern Power Grid. The data is manually cleaned by multiple professional, marked the rectangular frame coordinate value of the existing object, and labeled into six categories, including five categories: thermometer square, thermometer circle, lightning arrester, oil level gauge and SF6 meter. These images constitute our experimental dataset, of which 4364 are used for training and 181 for testing.

### *3.2. Related Details*

The main details of the experiment include the augmentation method for expanding the amount of data, the software and hardware configuration for training, and the complexity and accuracy metrics for evaluating the model.

Like Yolov5, we adopt mosaic method[16], which uses four pictures to splice through random scaling, random clipping and random arrangement, so as to enrich the dataset and greatly improve the training speed of the network.

All programs are based on pytorch1.7 framework and accelerated model training through NVIDIA Geforce RTX 2080s. During routine training, 300 epochs are trained by cosine annealing algorithm[17], and then 100 epochs for finetuning after pruning.

In terms of evaluating the performance of the model, on the one hand, the complexity of the model is measured by the amount of parameters and floating point operations (FLOPs), and on the other hand, the accuracy of the model is evaluated by precision, recall and mean average precision (mAP), as shown in formula 1-4.

$$AP = \sum_{i=1}^{n-1} (Recall_{i+1} - Recall_{i}) \cdot \max_{j \geq i}(\Pr ecision_{j}) \tag{1}$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{2}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{3}$$

$TP_i$, $FP_i$ and $FN_i$ respectively represent the number of ground truth IOU more than 0.5, less than 0.5, and the number that is wrongly identified as other categories in the top *i* bounding boxes of the model prediction on a single meter category. The IOU is the ratio of the intersection and union area of bounding box and ground truth, which is defined as follows:

$$IOU_j = \frac{S_{Boundbox}(k) \cap S_{Groundtruth}(k)}{S_{Boundbox}(k) \cup S_{Groundtruth}(k)} \quad (k \le i, k \in N^+) \tag{4}$$

### *3.3. Results Analysis*

Table 1 shows the detection results of the final model we proposed. It can be seen that our model can achieve a precision and recall close to 1 on five categories of test samples. It is worth noting that the simplified model can still accurately locate almost all the samples to be tested, which can meet the needs of our real production environment.

**Table 1.** Experimental result of the proposed method on different classes.

| Class | amount | Precision | Recall |
|---|---|---|---|
| thermometer square | 30 | 0.933 | 0.933 |
| thermometer circle | 32 | 1.000 | 0.989 |
| lightning arrester | 55 | 0.979 | 1.000 |
| oil level gauge | 19 | 0.986 | 1.000 |
| sf6 meter | 45 | 0.964 | 0.978 |
| all | 181 | 0.972 | 0.980 |

In order to intuitively show the effect of the model and explain the feature information learned by the model as much as possible, we visualized the detection results of samples and gradient-weighted class activation mapping (Grad-CAM)[18]. As shown in Figure 3, the proposed model accurately predicts the position and category of the meter, and the regions of interest concerned by the model are concentrated in highly discriminative positions.

Further, in order to show the lightweight degree of the proposed model, we enumerate the differences between the original Yolov5 and our improved and pruned model in terms of parameters and FLOPs. As shown in Table 2, the network parameters after improving the backbone and head structure are reduced by nearly 18 times, and the network complexity after pruning is further reduced by more than 20 times with almost no loss of mAP. It is worth noting that compared with Yolov5 architecture of other backbone networks (ResNet-18, VGG-16 and MobileNet), we achieved a balance between model complexity and accuracy by decomplacing conventional complex convolutional operations through depthwise separable convolution and removing unnecessary model weights through channel pruning. It can also be seen from Figure 4 that compared with the original high parameter network, our improved and pruned model is only slightly different in object location and confidence, which is enough to meet the detection requirements, and proves the effectiveness of the proposed method.

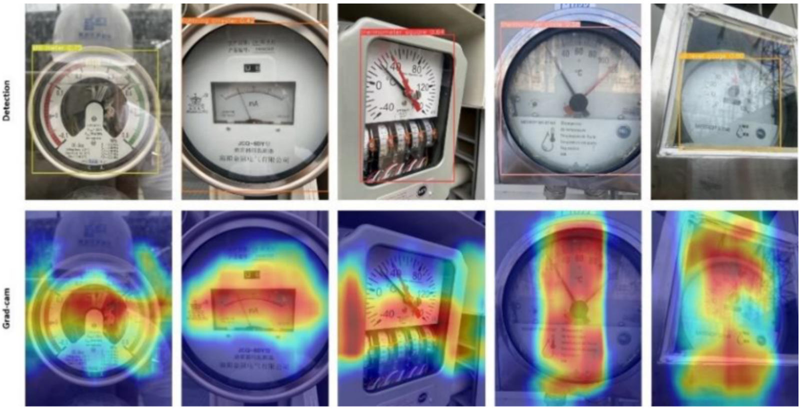

**Figure 3.** Detection and grad-cam results on our improved & pruned Yolov5.

**Table 2.** Comparison of model complexity and precision.

| Model | Parameters (M) | FLOPs (G) | mAP |
|---|---|---|---|
| Yolov5s | 7.033 | 15.980 | 0.993 |
| Yolov5s-ResNet18 | 11.402 | 30.238 | 0.996 |
| Yolov5s-VGG16 | 14.961 | 251.129 | 0.997 |
| Yolov5s-MobileNet | 3.490 | 3.488 | 0.991 |
| Improved Yolov5 | 0.393 | 0.926 | 0.989 |
| Improved & Pruned Yolov5 | 0.250 | 0.687 | 0.987 |

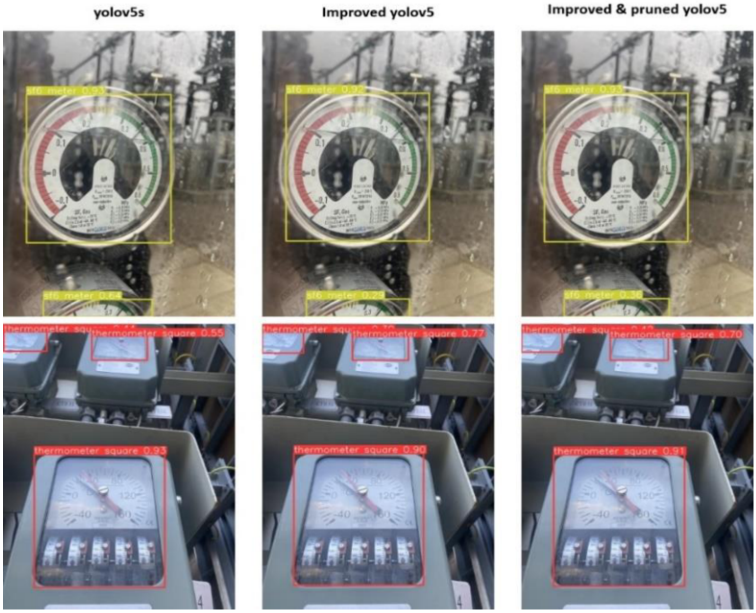

**Figure 4.** Comparison of detection effects of different models.

## 4. Conclusion

Aiming at the problem that resource constrained devices cannot efficiently detect meters in the real environment, we improve the structure of backbone and head based on Yolov5 architecture, greatly reduce the amount of network parameters and flops, and further remove the redundant channels in the model designed by handcraft through structured pruning method. Compared with the original network, our method reduces the complexity by more than 20 times, and can accurately locate the instruments. In the future, we will continue to collect more relevant data and study more advanced structures to comprehensively improve the detection quality and efficiency of the model.

## References

[1] Laroca R, et al. Convolutional neural networks for automatic meter reading. Journal of Electronic Imaging, 2019 Feb;28(1):13-23.

[2] He P, et al. A value recognition algorithm for pointer meter based on improved Mask-RCNN. 2019 9th International Conference on Information Science and Technology; 2019 Aug 2-5, Hulunbuir, China: IEEE. pp. 108-113.

[3] Fang Y, Dai Y, He G, et al. A mask RCNN based automatic reading method for pointer meter. 2019 Chinese Control Conference; 2019 Jul 27-30, Guangzhou, China: IEEE. pp. 8466-8471.

[4] Zhang X, Lu Y, Zhang X, et al. Research on Detection and Recognition of Pointer Instrument Based on Lightweight Network. Proceedings of the 2020 International Conference on Aviation Safety and Information Technology;2020 Oct 14-16, Weihai, China: ACM. pp. 301-306.

[5] Li Q, et al. GIS Room Autonomous Inspection System Based on Multi-rotor UAV. 2021 International Conference on Electrical Materials and Power Equipment; 2021 Apr 11-15; Chongqing, China: IEEE; pp. 1-4.

[6] Patel C, Bhatt D, Sharma U, et al. DBGC: Dimension-based generic convolution block for object recognition. Sensors, 2022, 22(5): 1780-1804.

[7] Yan B, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5. Remote Sensing, 2021 Apr;13(9): 1619-1641.

[8] Zhou F, et al. Safety helmet detection based on YOLOv5. 2021 IEEE International Conference on Power Electronics Computer Applications; 2021 Jan 22-24; Shenyang, China: IEEE. pp. 6-11.

[9] Zhao J, et al. A wheat spike detection method in UAV images based on improved YOLOv5. Remote Sensing, 2021 Aug;13(16), 3095-3110.

[10] Wen P, et al. Research on early fire detection of Yolo V5 based on multiple transfer learning. Fire Science and Technology; 2021 Jan;40(1): 109-112.

[11] F Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21-26; Honolulu, United States: IEEE. pp. 1800-1807.

[12] Howard, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. 2017; 1-9.

[13] Hu J, et al. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018 Jun 18-22; Salt Lake City, United States: IEEE; pp. 7132-7141.

[14] He Y, et al. Filter pruning via geometric median for deep convolutional neural networks acceleration. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019 Jun 15-21; Los Angeles CA, United States: IEEE; pp. 4340-4349.

[15] Vasoya S, Patel N, Ramoliya D, et al. Potentials of Machine Learning for Data analysis in IoT: A Detailed Survey. 2020 3rd International Conference on Intelligent Sustainable Systems; 2020 Dec 3-5: Thoothukudi, India: IEEE; pp. 291-296.

[16] Bochkovskiy A, et al.. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. 2020; 1-17.

[17] Loshchilov I, et al. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983. 2017; 1-16.

[18] Selvaraju R, et al. Grad-cam: Visual explanations from deep etworks via gradient-based localization. Proceedings of the IEEE international conference on computer vision; 2017 Oct 22-29; Venice, Italy: IEEE; pp. 618-626.