Infrared Image Recognition of Power Equipment Based on Improved YOLOv5

1st Zhenzhou WANG
Information Science and Engineering College
HeiBei University Science And Technology
ShiJiazhuang, China
chinawzz@163.com

2nd Mingming LI
Information Science and Engineering College
HeiBei University Science And Technology
ShiJiazhuang, China
lm1380320@163.com

3rd Jingfang SU
Information Science and Engineering College
HeiBei University Science And Technology
ShiJiazhuang, China
18272565@qq.com

4th Zijian LIU
Information Science and Engineering College
HeiBei University Science And Technology
ShiJiazhuang, China
2336617735@qq.com

***Abstract*—Addressing the challenge of achieving both high accuracy and real-time processing in infrared target detection tasks within complex electrical scenarios, a lightweight YOLOv5s-based algorithm is designed for the detection of infrared electrical equipment. Initially, MobileNetV3 is employed to replace the backbone network of YOLOv5s, accelerating detection and facilitating network light-weighting. Subsequently, in the Neck part, the GSConv convolutional module is designed to fuse multi-channel feature information, and the EMA attention mechanism is incorporated, enhancing the recognition capability of electrical equipment while maintaining a lightweight model. Finally, the MpDIou loss function is utilized in place of the original loss function, improving the network's prediction accuracy. The improved model achieves a mAP of 94.58% on a self-constructed dataset, reduces GFLOPs to 3.0, and reaches a detection speed of 96FPS, indicating a 1.29% increase in mAP, an 81.7% reduction in GFLOPs, and a 37.9% improvement in detection speed compared to the original YOLOv5s model. Experimental results demonstrate that the enhanced YOLOv5s model effectively improves the accuracy and speed of infrared electrical equipment recognition while being lightweight, making it easy to deploy and meeting the demands of practical inspection scenarios..**

Keywords—Power equipment; Infrared images; Lightweight network; YOLOv5s; GSConv; Attention mechanism

I. INTRODUCTION

As substations continue to develop and be constructed, infrared imaging technology has been widely applied in the inspection of electrical equipment to improve detection efficiency and reliability, and to reduce the risk of potential faults[1]. Compared to visible light images, infrared images have the advantage of stronger penetration and are less affected by external weather conditions and lighting, allowing for the timely acquisition of the status information of electrical equipment[2-3]. This capability effectively prevents electrical faults caused by equipment overheating. Previous methods of infrared image recognition focused primarily on image processing, utilizing texture features, statistical features, and temperature distribution characteristics to identify electrical equipment within infrared images[4-6]. Most of these machine learning methods[7-9] can detect electrical equipment, but they do not balance accuracy and speed well, resulting in low detection precision or high computational requirements, which are not conducive to online inspection. Deep learning image recognition algorithms, however, can assist in the recognition of a large volume of images, and many scholars have conducted extensive work on combining deep learning with infrared image recognition[10-14]. Liu Yangfan et al. [15]proposed an improved YOLOv4 method for detecting weak infrared targets in space, meeting the needs of space infrared weak target detection tasks, yet the issue of low recall rate remains unresolved. Wu et al. [16] presented the problems of target occlusion and inaccurate classification in actual collected infrared images of electrical equipment, proposing the TA-YOLO algorithm to significantly improve the detection accuracy and speed of infrared targets, though the algorithm's detection effectiveness for small targets was not ideal. Jian et al. [17] added two new types of anchor boxes in the Faster R-CNN model to improve the detection accuracy of slender substation equipment. However, the introduction of new anchor boxes and other improvements may still result in high model complexity, making deployment and operation in resource-constrained environments challenging. Consequently, algorithms often need to be lightweight to meet the accuracy and speed requirements of detection models on embedded devices. Huang et al. [18] presented an improved lightweight network, ResFuse-YOLO_Tiny, for efficient detection of small infrared targets. By employing techniques such as expanded receptive fields, feature reuse, and attention mechanisms, it significantly improved detection accuracy in resource-limited scenarios, but deploying in edge device scenarios may still encounter computational and storage resource limitations.

Addressing these challenges, this article introduces an advanced lightweight infrared detection method based on YOLOv5s, designed to achieve model simplification while ensuring high accuracy. This method utilizes MobileNetV3 as the backbone network to hasten detection and further streamline the network; introduces a GSConv convolution module in the Neck section for the fusion of multi-channel feature information, alongside the integration of the EMA attention mechanism; and adopts the MpDIou loss function for more precise prediction framing. Employing a custom dataset for the identification of four types of electrical equipment—insulator strings (IS), current transformers (CT), voltage transformers (VT), and arresters (AR)—this model promises accuracy and real-time performance in the detection of these key electrical components.

II. METHODS

YOLOv5 stands as a highly efficient real-time object detection algorithm, available in four variants: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Considering application scenarios, YOLOv5s, compared to the other three versions, features the simplest network structure, fastest operational speed, and lowest computational resource consumption, making it the most easily portable across

979-8-3503-4929-0/24/$31.00 ©2024 IEEE

platforms. Consequently, it enjoys widespread application across various domains. The algorithm is structured into three main components: the Backbone as the foundational network, the Neck for aggregating multi-layered feature information from images, and the Head for predicting the location and category of objects. Initially, the algorithm employs the backbone network to extract features from the input image; subsequently, it leverages the neck network to amalgamate multi-layered feature information; finally, through different scaled detection heads, it predicts the objects' location and category, further refining the prediction by employing the Non-Maximum Suppression (NMS) algorithm to select the optimal prediction boxes.

The primary research focus of this paper is on improving the YOLOv5s model, which has the shallowest depth and width among the models. The overall structure is shown in Fig. 1. Firstly, this paper replaces the backbone network of the original YOLOv5s with MobileNetV3 to accelerate detection and achieve network lightweighting; secondly, in the Neck section, the GSConv convolution module is designed to fuse multi-channel feature information and incorporate the EMA attention mechanism. This approach not only maintains the model's lightweight but also better integrates multi-scale features by adaptively increasing the weight of key features to enhance the recognition capability of electrical equipment; finally, we use the MpDIou loss function to replace the original CIoU loss function, making the final prediction frame closer to the real frame and improving the network's prediction accuracy.

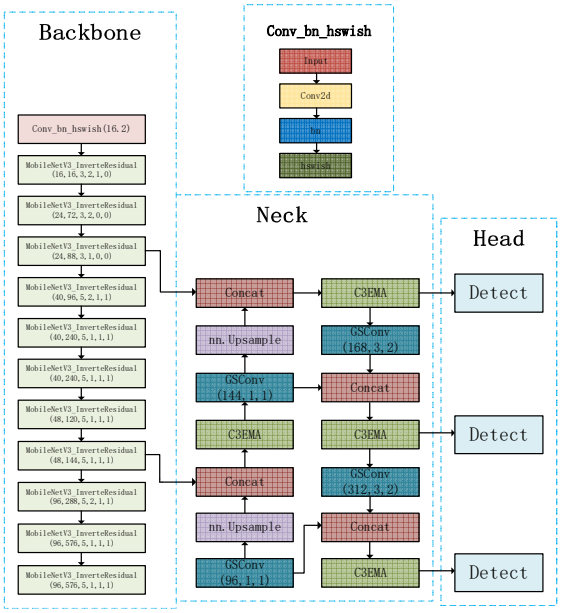


Fig. 1. The structure of the improved YOLOv5s.

III. Improvement of YOLOv5s Algorithm

A. Replacing the backbone network

This paper replaces the original YOLOv5s backbone network with MobileNetV3 to speed up detection and network lightweighting. In 2019, Howard and others introduced the MobileNetV3 network, which inherits the depth-wise separable convolution from MobileNetV1 and the inverted residual structure with linear bottlenecks from MobileNetV2. It also utilizes the NetAdapt algorithm to search and optimize the number and size of convolution kernels and channels, and incorporates the SE (Squeeze-and-Excitation) lightweight attention mechanism into network construction. In this paper, the MobileNetV3-Small structure is adopted as the improved backbone network for YOLOv5s, and the structure of MobileNetV3-Small is shown in Fig 2.

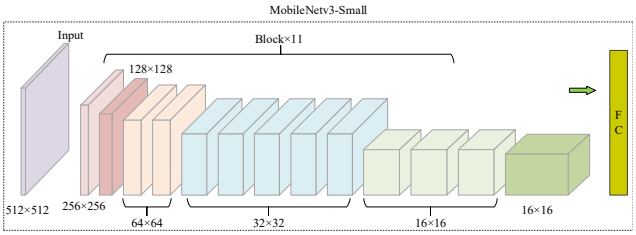
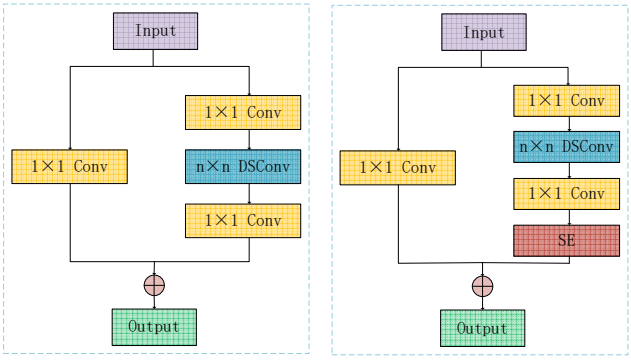


Fig. 2. Structure of MobileNetV3-Small module.

The MobileNetV3-Small architecture comprises 11 blocks, each featuring one of two unique configurations, as shown in Fig 3. Observing Fig 3a reveals that the process begins with a 1x1 convolution, aimed at reducing the feature map's channel count. This step is followed by a depthwise separable convolution aimed at decreasing computational demands, with n being either 3 or 5. Unlike standard convolution where each kernel processes every channel of the feature map, depthwise separable convolution assigns individual kernels to individual channels, thereby significantly lowering computational expenses. Post depthwise separable convolution, a 1x1 convolution is employed to define the channel count and consolidate the features. The concluding step involves summing the outputs from both paths to generate the cumulative feature map. This specific configuration, depicted in Fig. 3a, is utilized in the second and third blocks of the MobileNetV3-Small model. The rest of the blocks adhere to the Structure2, as seen in Fig. 3b. Upon comparison of Fig. 3a and 3b, it's apparent that the latter integrates the SE module , which functions to augment the semantic relevance of the target region.



（a）Structure1 (b) Structure2

Fig. 3. Two different structures in the MobileNetV3 module.

B. Neck improvement

Building on the lightweighting of the YOLOv5s backbone network, while significantly enhancing detection speed, a certain degree of accuracy in recognizing electrical equipment has been compromised. To address this, the study employs the efficient GSConv to extract more diverse channel information across scales, effectively boosting the Neck's capability to fuse multi-level semantic features. The structure of the GSConv module is depicted in Fig 4. The GSConv module enables better integration and utilization of multi-channel feature information while minimizing the

consumption of computational resources. This allows the detection network to enhance its perception of features across different scales and abstraction levels at a lower computational cost, thereby increasing the accuracy of object detection.

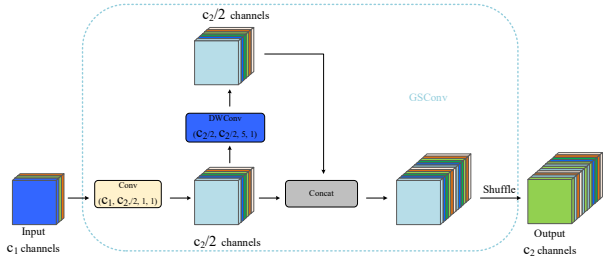


Fig. 4. GSConv Module Structure

To tackle scenarios commonly surrounded by complex background objects such as support steel frames and heating buildings around electrical equipment, and to enhance model performance and processing speed, the Efficient Multi-scale Attention (EMA) module is introduced. The EMA module is a parallel attention mechanism designed for computer vision tasks, aimed at boosting model performance and processing speed. In contrast to the sequential processing mode of traditional Convolutional Neural Networks (CNNs), it processes input data simultaneously through its parallel structure. This parallel processing approach not only accelerates the model training process for large datasets, effectively reducing the required amount of parameters and computational costs but also enhances the model's recognition accuracy by processing features of different scales in parallel. The structure of EMA is illustrated in Fig 5, where "g" represents the divided groups, "X Avg Pool" stands for 1D horizontal global pooling, and "Y Avg Pool" represents 1D vertical global pooling, respectively.

By integrating the GSConv and EMA (Efficient Multi-scale Attention) modules in the design of the Neck section, not only is the ability of the model to process multi-scale features in complex backgrounds enhanced while maintaining a lightweight network structure, but also the adaptive adjustment of feature weights further improves the recognition capabilities for electrical equipment.

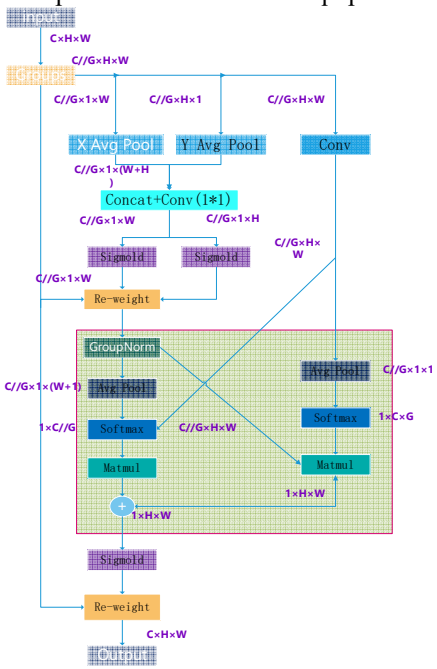


Fig. 5. EMA Network Structure

C. Replace loss function

Infrared images of electrical equipment are often captured under conditions of low contrast and noise. Additionally, due to the similarity in contour features of equipment such as current transformers and surge protectors, enhancing detection performance is particularly important. The regression loss function used in YOLOv5 is CIoU, which does not consider the orientation between the true and predicted boxes, leading to slower convergence and inferior inference performance. Therefore, this paper substitutes the loss function with MpDIou, whose primary function is to measure the distance between the network's predictions and the required information. This change is expected to improve the accuracy and robustness of the model for the task of recognizing electrical equipment in infrared images.

Mpdiou,which incorporates all relevant factors from existing loss functions, serves as a metric for comparing the similarity of bounding boxes, based on the minimum distance between points. Mpdiou streamlines the process of comparing two bounding boxes, applicable to both overlapping and non-overlapping scenarios in bounding box regression. Consequently, Mpdiou offers a viable substitute for the intersection-over-union (IoU) metric across various performance evaluations in 2D/3D computer vision tasks. Additionally, it simplifies calculations by minimizing the distances between the top-left and bottom-right points of the predicted and actual bounding boxes. The computation of Mpdiou is as follows.

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \quad (1)$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \quad (2)$$

$$\text{MpDIou} = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (3)$$

In Eqs. (1)~(3) d1, d2 denote the intersection and minimum point distance,the arbitrary shapes A and B are subsets within a larger set S that is defined in an n-dimensional real space, denoted as Rn: A, B ⊆ S ∈ Rn, w and h are the width and height of the input image, respectively and MpDIou is the output.Let$(x_1^A, y_1^A), (x_2^A, y_2^A)$denote the coordinates of the upper left and lower right points of A. Let $(x_1^B, y_1^B), (x_2^B, y_2^B)$ denote the coordinates of the upper left and lower right points of B, respectively.

IV. Experiment and Result Analysis

A. Production of datasets

The absence of public infrared datasets for substation equipment necessitated the manual acquisition of a diverse set of infrared power equipment images, using a handheld thermal imager at a 220kV substation in Handan city. The dataset encompasses electrical components such as Current Transformers (CT), Voltage Transformers (VT), Arresters (AR), and Insulator Strings (IS).A dataset of 1086 images was compiled by photographing the substation at various times, angles, and distances.To mitigate model overfitting and enhance generalization, data augmentation techniques such as mosaic, flipping, scaling, translating, and noise addition were implemented. Images were annotated with the LabelImg tool. The dataset was split into training, validation, and test subsets in a 7:2:1 ratio. This split resulted in 3123 training images, 783 validation images, and 1667 test images. Sample images are presented in Fig. 6.

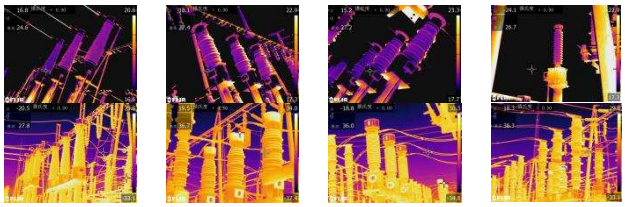


Fig. 6. Partial Samples

B. Environment configuration and evaluation indicators

The experiments were conducted on the Hengyuan Zhixiang cloud computing center platform, with all deep learning models implemented using the PyTorch framework. The training and testing of the experimental data were carried out on servers equipped with NVIDIA GeForce RTX 2060 SUPER (GPU), 32G of computer memory, and Compute Unified Device Architecture (CUDA) version 11.7.0. During model training, the number of epochs was set to 300, with a batch size parameter of 16.

This study utilizes metrics such as Average Precision (AP), mean Average Precision (mAP), Frames Per Second (FPS), and Giga Floating Point Operations per Second (GFLOP) to evaluate model performance. It also refers to precision (P) and recall (R), with their expressions shown in equations (4) and (5), respectively. AP is the area under the P-R curve, as depicted in equation (6). mAP represents the average of AP across all categories, as shown in equation (7), where n is the number of categories.GFLOP is used to measure the complexity of a model or algorithm. Generally, the smaller the GFLOP, the less computational power the model requires, the lower the hardware specifications needed, making it easier to deploy on lower-end devices. FPS indicates the speed at which the model processes image frames per unit of time.

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

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

Where in the formulas: TP represents true positives, FP represents false positives, and FN represents false negatives.

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{6}$$

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n} \tag{7}$$

C. Result analysis

In terms of deep learning, the magnitude of the loss function reflects the discrepancy between the final prediction of the object detection model and the actual values. It can be used to assess the quality of the training process, the degree of model convergence, and whether overfitting is present. To verify whether the improved YOLOv5s model enhances the performance of the network model, this study conducts epoch training on both the original and the improved models using the same dataset. The experiment data are then compared and analyzed. The curve showing the change of train/batch loss over the number of iterations is presented in Fig. 7. By analyzing the curve in Figure 7, the results of the model comparison can be realized.

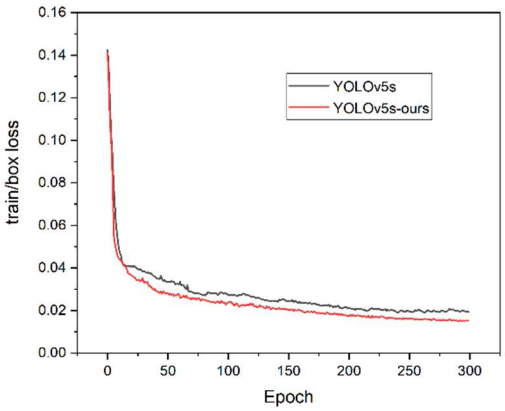


Fig. 7. Train/bos loss variation curve

As observed from Fig. 7, both models exhibit a rapid decline in the initial phase of training, eventually stabilizing, but the loss function of the improved YOLOv5 decreases faster than that of the original YOLOv5 model. The EIou Loss curve of the network models before and after improvement tends to stabilize at 250 epochs, with the model converging once the loss value reaches 0.018. With the same number of training epochs, the improved model exhibits a lower loss function value, indicating a stronger ability to learn features with less loss of detail.

During the training process, the mAP metric can often reflect the effectiveness of the object detection model's training. In this study, the value of mAP@0.5 is used to evaluate model performance, with performance changes depicted in the following Fig. 8, where the blue and red lines represent YOLOv5s and YOLOv5s-ours, respectively.

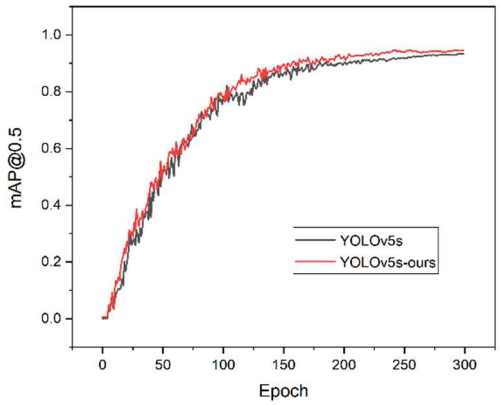


Fig. 8. mAP@0.5 variation curve

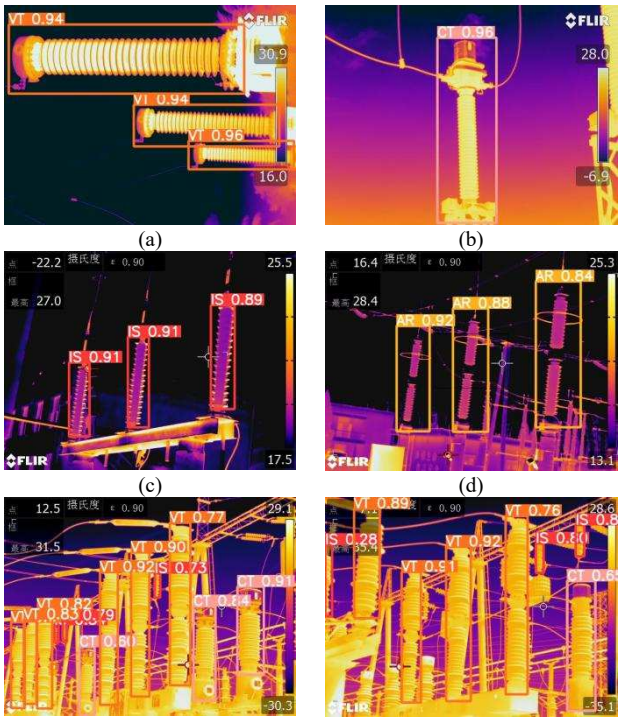
As illustrated in Fig. 8, the mAP values of both the original YOLOv5s and the improved YOLOv5s model exhibit varying degrees of fluctuation, but they generally rise steadily and start to converge around 270 epochs. After reaching a relative stability, the mAP values of the improved YOLOv5s model consistently exceed those of the original YOLOv5s model, indicating that the improved model has a slight increase in mAP value compared to the original YOLOv5s model, with an increase of 1.29%. Next, a detailed analysis of the detection outcomes for the four types of electrical equipment, namely current transformers, voltage transformers, surge arresters, and insulator strings, is conducted. The accuracy rates before and after improvement

for each category of electrical equipment are presented in Table 1.

TABLE 1. ACCURACY BEFORE AND AFTER IMPROVEMENT

Equipment type	Accuracy before improvement/%	Accuracy after improvement/%
CT	91.34	92.82
VT	91.93	93.23
AR	94.65	95.84
IS	96.84	97.82
Amount to	93.56	94.85

Fig. 9 displays the detection results of the proposed method on a subset of the test set. Specifically, Fig. 9(a), 9(b), 9(c), and 9(d) individually showcase the detection results for Voltage Transformers (VT), Current Transformers (CT), Insulator Strings (IS), and Surge Arresters (AR), respectively. Fig. 9(e) and 9(f) present the detection outcomes within the complex background of a substation. It can be observed that utilizing the improved algorithm not only achieves model lightweighting but also enables electrical equipment to be detected accurately with high confidence levels.



(a) (b) (c) (d) (e) (f)

Fig. 9. Model detection effect

D. Ablation experiment

To evaluate the impact of various module modifications on the optimization of algorithm performance in detecting power equipment in infrared images, a series of ablation experiments are conducted using YOLOv5s as the baseline. These experiments aim to meticulously explore how each modification enhances detection effectiveness. To ensure the consistency and comparability of the experiments, all are carried out under a fixed experimental environment and parameter configuration.Symbol ① denotes the use of the MobileNetv3-Small network to replace the YOLOv5s backbone extraction network; ② represents the combination of GSConv and EMA attention modules in the Neck section for improvement; ③ signifies the replacement of the CIoU loss function with MpDIou, with the experimental results presented in Table 2.

TABLE 2 RESULTS OF ABLATION EXPERIMENT

①	②	③	mAP/%	FPS	GFLOP
×	×	×	93.56	69.6	16.4
√	×	×	92.83	102	3.2
×	√	×	94.48	81.0	15.6
×	×	√	93.97	72.5	16.7
√	√	√	94.85	96	3.0

Table 2 indicates that substituting the backbone extraction network with the MobileNetv3-Small network resulted in a slight decrease in mAP. However, the detection speed of the network improved by 46.5%, and the GFLOPs decreased by 80.4%. This suggests that using the MobileNetv3-Small to replace the backbone network can significantly reduce the model's parameters, thus achieving model lightweighting. After improvements were made in the Neck section using a combination of GSConv and EMA attention modules, there was a slight increase in detection speed and a 0.92% increase in mAP, along with a 4.8% reduction in GFLOPs. The substitution of the loss function with MpDIou resulted in a 0.41% increase in mAP, achieving the anticipated results. With the implementation of these three improvement strategies to the YOLOv5 model, a balance was attained between the model's detection accuracy and speed. The improved model's mAP reached 94.85%, which is a 1.29% improvement from before the improvements, and the detection speed reached 96FPS, marking a 37.9% increase from prior, with GFLOPs being reduced by 81.7%. Consequently, the improvement strategies employed in the experiments are demonstrably viable.

E. Comparing experiments

To assess the efficacy of the improved model, this paper conducts comparative experiments with commonly used lightweight networks, including ShuffleNet and GhostNet, in addition to the MobileNetv3 replacement. According to the results in Table 3, YOLOv5s with MobileNetV3, despite having the lowest mAP among the models tested, has the smallest number of parameters, the lowest GFLOPS, and the fastest image processing speed. In summary, MobileNetV3 offers a solution that significantly enhances efficiency within an acceptable accuracy range, particularly in terms of parameter count and computational efficiency, making it highly suitable for fast and lightweight real-time detection applications.

TABLE 3 COMPARISON OF LIGHTWEIGHT MODELS

Model	Parameter quantity	GFLOPS	FPS	mAP/%
YOLOv5s	7035809	16.2	69.6	93.56
YOLOv5s-MobileNetV3	1494522	3.2	102	92.83
YOLOv5s-ShuffleNet	3802589	7.9	88.3	93.12
YOLOv5s-GhostNet	2378104	5.3	93	92.94

To validate the superiority of the improved YOLOv5s algorithm, a series of comparative experiments were conducted. Four other common object detection algorithms were selected for comparison, including Faster R-CNN, SSD, YOLOv4, and the original YOLOv5s. These algorithms were trained and tested on the same experimental platform, with various evaluation metrics presented in Table 4.

As discernible from Table 4, it is evident that different models exhibit varied levels of proficiency in identifying different categories of electrical equipment. While Faster R-CNN demonstrated commendable mAP accuracy on the custom dataset, the model's considerable weight makes it less deployable. In contrast, the modified YOLOv5 not only boasts a smaller model size but also surpasses other models in mAP accuracy. Notably, the AP results for current transformers and voltage transformers have improved significantly over the original YOLOv5s, underscoring the proposed method's enhanced effect in feature extraction and its efficacious detection of electrical equipment with similar contours.

TABLE 4 COMPARATIVE EXPERIMENTS OF DIFFERENT MODELS

Model	mAP/ %	Size/M	AP/%			
			CT	*VT*	*AR*	*IS*
Faster-RCNN	93.24	246.7	90.96	91.87	93.68	96.46
SSD	91.91	86.4	91.45	88.78	94.05	93.34
YOLOv4	92.29	62.9	90.25	90.97	93.39	94.56
YOLOv5s	93.56	22.4	91.34	91.93	94.65	96.84
Our-YOLOv5s	94.85	4.8	92.82	93.23	95.84	97.82

CONCLUSION

This study proposes an infrared recognition method for substation equipment based on the improved YOLOv5s, achieving detection of a variety of substation equipment such as current transformers, voltage transformers, lightning arresters, and insulator strings. Under the premise of ensuring high detection accuracy in complex power scenarios, this study aims to minimize the model size, increase the detection speed, and achieve model lightweight. The improved model has increased the mAP by 1.29% on a self-built infrared dataset of power equipment, decreased GFLOP by 81.7%, and increased detection speed by 37.9% compared to the original YOLOv5s model. However, there are still some missed detections in complex scenarios such as overlapping targets and partial obstructions of electrical equipment. Future work will explore how to further improve the model or enhance the detection accuracy of infrared power equipment in similar scenarios through data augmentation. The improved YOLOv5 model can be deployed on inspection robots for real-time processing of infrared images of substation equipment, significantly enhancing detection accuracy and efficiency. This not only reduces reliance on manual inspections but also has practical significance for diagnosing faults in electrical equipment.

REFERENCES

[1] Liu T, Li G, Gao Y. Fault diagnosis method of substation equipment based on You Only Look Once algorithm and infrared imaging[J]. Energy Reports, 2022(8): 171-180.

[2] Li Yunhong, Liu Yudong, Su Xueping. A review of research on infrared and visible light image registration technology [J] Infrared Technology, 2022, 44(07): 641-651

[3] Yanchun Y,Jianwu D,et al.Infrared and visible image fusion based on GEMD and improved PCNN[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(9): 2317-2329.

[4] .Xu Sheng,Wen Jiqun.Research on Maintenance and Prevention Tech nology for Power Equipment Faults [J] China Equipment Engineering ,2022,(20):87-89

[5] X. Zhang, Intelligent Detection Technology of Infrared Image of Substation Equipment Based on Deep Learning Algorithm, 2021 IEEE Sustainable Power and Energy Conference (iSPEC), pp. 3855-3860, 2021.

[6] Li S, Li J. Condition monitoring and diagnosis of power equipment: review and prospective[J]. High Voltage, 2017, 2(2): 82-91.

[7] Toya, Wang Wei, Mao Huamin, Cheng Hongbo, et al.Machine learning based infrared intelligent diagnosis method for power equipment faults [J] Journal of Henan University of Technology, 2022, 41 (05): 121-126

[8] Dang Xiaojing, Liu Shungui, Zhu Guangnan, et al. Infrared image recognition algorithm for electrical equipment based on feature extraction [J].Journal of Shenyang University of Technology,2023,45(03): 264-269

[9] Han S, Yang F, Yang G, et al. Electrical equipment identification in infrared images based on ROI-selected CNN method[J]. Electric Power Systems Research, 2020, 188: 106534.

[10] Wang Yu, Chen Xiuxin, Yuan Hejin. Multi target recognition of substation infrared images based on improved Faster RCNN [J]. Journal of Sensing Technology,2021,34(4):522-530

[11] Han S, Yang F, Jiang H, et al. A smart thermography camera and application in the diagnosis of electrical equipment[J]. IEEE Transactions on Instrumentation and Measurement, 2021,70:1-8.

[12] ZHENG Hanbo, CUI Yaohui, YANG Wenqiang, et al.An infrared image detection method of substation equipment combining iresgroup structure and CenterNet[J]. IEEE Transactions on Power Delivery, 2022, 37(6): 4757-4765.

[13] Wang B, Dong M, Ren M, et al. Automatic fault diagnosis of infrared insulator images based on image instance segmentation and temperature analysis[J]. IEEE Transactions on Instrumentation and Measurement,2021, 69(8):5345-5355.

[14] ZHENG Han, SUN Yonghui, LIU Xinghua, et al. Infrared image detection of substation insulators using an improved fusion single shot multibox detector[J]. IEEE Transactions on Power Delivery, 2021, 36(6): 3351-3359.

[15] Liu Yangfan, Cao Lihua, Li Ning, et al. Space infrared weak target detection based on YOLOv4 [J]. Liquid Crystal and Display, 2021,36(4):615-623

[16] Wu Huihai, Shen Wenzhong. TA-YOLO based infrared image detection method for power equipment [J]. Information Technology and Informatization, 2022(3):17-20

[17] OU Jianhua, WANG Jianguo, XUE Jian, et al. Infrared image target detection of substation electrical equipment using an improved faster R-CNN[J]. IEEE Transactions on Power Delivery, 2023, 38(1): 387-396

[18] HUANG Song, SHANG Bowen, SONG Yanlou, et al. Research on real-time disconnector state evaluation method based on multi-source images[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:1-15