Fusing PSA to Improve YOLOv5s Detection algorithm for Electric Power Operation Wearable devices

Qiuming Liu[1,2(✉)], Wei Xu[1], Yang Zhou[3], Ruiqing Li[1], Dong Wu[3], Yong Luo[4], and Longping Chen[1]

[1] School of Software Engineering, Jiangxi University of Science and Technology, Nanchang 330013, China
liuqiuming@jxust.edu.cn,6720210702@mail.jxust.edu.cn
[2] Nanchang Key laboratory of Virtual Digital Factory and Cultural Communications, Nanchang 330013, China
[3] Information and Communication Branch, State Grid Jiangxi Electric Power Co, Nanchang 330095, China
6720210698@mail.jxust.edu.cn
[4] School of Software, Jiangxi Normal University, Nanchang 330022, China

Abstract. In order to determine whether the electric power workers wear safety equipment such as safety helmet, insulation boots, insulation gloves, insulation clothes, etc., to ensure the safety of the electric power construction site. We propose a electric power operation safety equipment detection algorithm incorporating PSA to improve YOLOv5s algorithm, using polarized self-attention mechanism to improve the feature extraction end of YOLOv5s algorithm, improving the channel resolution and spatial resolution of safety equipment images of electric power operation scenes, and preserving the information of key nodes of small targets that are obscured; GSConv is used to replace the ordinary convolution to reduce the complexity of the model, improve the calculation speed of the algorithm and improve the detection accuracy. The experimental results show that the average accuracy mean (IoU = 0.5) of the proposed algorithm reaches 0.961, which is 1.58% higher than that of the original network detection performance, and the model parameters are reduced from 7.03 to 5.48 millions. It effectively improves the detection speed and accuracy of the algorithm, and can effectively monitor whether the operator wears the safety equipment correctly when there are occlusions and missing safety equipment in the electric power operation scene, which has a excellent application effect.

Keywords: Improved YOLOv5s algorithm · Polarized self-attention · Power operation scenario · VoV-GSCSP · Safety wearable

C. Wu et al. (Eds.): MONAMI 2023, LNICST 559, pp. 121–135, 2024.
https://doi.org/10.1007/978-3-031-55471-1_10

1 Introduction

The safety of electric power operation is always one of the hot issues of social concern. From the investigation results of the causes of electric power operation accidents, most of the accidents are caused by the construction process is not standardized enough [1,2], and the workwear equipment is the basic equipment to ensure the safety of construction personnel. Staff wearing the operating equipment can not only effectively prevent the risk of electric shock during electrical operations and reduce personal injury caused by safety accidents, but also avoid serious safety accidents. Therefore, safety helmets, insulated boots, insulated gloves, insulated clothing and other safety equipment, etc. are the safety of electric power workers when operating [3].

For this reason, it is of great significance to test the wearing condition of the operating equipment for construction personnel in electric power operations. At present, many scholars have proposed methods for detecting construction personnel wearing equipment for electrical operations, such as the helmet wearing detection method proposed by Xin Liu et al. [4], which is to input construction scene images into a convolutional neural network, use this network to continuously iterate to extract construction personnel helmet features, and output helmet wearing classification results. However, this method is affected by the blurred images of the construction scenes and the existence of obscuring situations, and its monitoring results are not accurate enough; Zhang Mingyuan et al [5] explored the automatic real-time detection of construction workers wearing helmets at construction sites and proposed a Tensorflow-based framework, which uses a region based convolutional neural network (Faster R-CNN) method to monitor workers' helmet wearing condition in real time, but the evaluation indexes selected by the method are all expert voting results, which are subjective and lead to its poor final detection results; Wang Yusheng et al. [6] proposed a helmet wearing detection method based on human pose estimation (HPE) algorithm for the problem of difficult detection and low accuracy under complex posture of construction personnel.

Many scholars in China and abroad have used other deep learning and YOLO algorithms for the detection of electric power behavior. Such as, Qu Wenqian et al [7] proposed a YOLOv3-based helmet wearing detection method for electric power grid operation sites in order to effectively monitor the irregular helmet wearing behavior of electric power grid operators, and constructed image samples under three cases of correct, incorrect and un-wearing helmets, with 92.59% detection accuracy; Wang Jian et al. [8] proposed a separable convolution method to improve the YOLOv5 algorithm in order to identify the helmet wearing situation of the staff at the electric construction site, using the mosaic data enhancement method to improve the clarity of the visual image of the helmet and introducing the self-attention mechanism, which can effectively identify the helmet wearing situation; Fu Desu et al. [9] proposed an improved detection algorithm based on YOLOv5 for electric power operation helmet and insulated gloves wearing, adding a small target extraction mechanism and incorporating a weighted bidirectional feature pyramid network structure, the accuracy value of the detection of the target reached 93.3%.

In summary, although the YOLO series has made a breakthrough in target detection, its wearing equipment detection in a single case can achieve high detection accuracy, electric power wearing equipment including helmets, insulated boots, insulated gloves, insulated clothing, etc., obviously can not meet the requirements of the actual electric power operators multi-attribute wearing equipment. In order to ensure the life safety of electric power operators, the actual electric power operation also needs to detect and give feedback in real time to the danger and violation of wearing equipment in the complex environment of electric power operators, so the YOLO algorithm has high requirements for speed.

Therefore, in order to solve the requirement of single identification attribute of electric power operation wearer and the need for multiple wearer attributes to be detected simultaneously, and also the need for real-time detection and feedback of irregular and dangerous wearer behaviors in the shortest possible time, this paper proposes a electric power operation safety equipment detection algorithm incorporating PSA (Polarized self-attention mechanism) improvement of YOLOv5s algorithm, using polarized self-attention mechanism to improve the feature extraction end of YOLOv5s algorithm, and using orthogonal approach to ensure the low number of parameters while improving the channel resolution and spatial resolution of safety equipment images of electric power operation scenes; The K-mean clustering algorithm is used to obtain the candidate frame settings at the output of YOLOv5s algorithm; GSConv is used to replace the normal convolution to reduce the complexity of the model, improve the computational speed of the algorithm and increase the detection accuracy. It is able to detect the attributes of multiple wearable devices for electric power operation at the same time, and the detection speed has been greatly improved, solving the problems of single identification and slow detection speed of the attributes of wearable devices for electric power operation.

2 YOLOv5 Algorithm

After YOLOv1 appeared in 2015, a series of algorithms such as YOLOv2 [10]-YOLOv5 [11] emerged in order to continue improving its performance. YOLOv5 is more flexible and faster than YOLOv4, and it can also improve its accuracy. YOLOv5 has four models, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, with gradually increasing parameter size and accuracy. Distinguished by the number of Bottleneck, a control factor similar to that of EfficientNet [12] is used to achieve the variation of the version, enabling the selection of a suitable size model according to the application scenario. In this paper, version 6.0 of the YOLOv5s model is used, as shown in Fig. 1. The network structure of YOLOv5s includes 4 parts: Input side, Backbone model, Neck model, and Head side.

3 Improved YOLOv5s Algorithm

Although YOLOv5s is a fast and accurate algorithm, in the target detection task of the wearable device in this paper, the background image accounts for a

relatively large portion of the image, and semantic segmentation of the image is needed to retain key points of small target information, such as insulated boots and insulated gloves in this paper. Therefore, it is necessary to maintain high resolution as much as possible, but also consider the computational volume, as well as try to connect global information to get the results, which is what YOLOv5s lacks at present. In the task of electric power operation equipment wear detection, real-time detection and feedback of detection results are needed to further improve the detection accuracy and speed, so it is necessary to lighten the YOLOv5s model while improving its accuracy. Based on the above factors, this paper proposes a fusion of PSA [13] to improve YOLOv5s algorithm for electric power operation wearable detection, and its main improvement points are as follows.

- Using the K-mean clustering algorithm mechanism to generate candidate boxes that match the data image, the boxes are divided into three categories of different size levels according to the size of the data image wearing device, and there are three types of boxes in each category to match the size of the power operation wearing device, which helps to improve the detection accuracy.
- The Polarized Self-Attention mechanism (PSA) is introduced with the aim of ensuring high channel resolution and high spatial resolution while effectively fusing global information to give different attention, thus improving the attention effect of the neural network on small target key node detection information; adding nonlinearity to the attention mechanism makes the fitted output more delicate and closer to the real output.
- The VoV-GSCSP method is introduced by replacing the normal convolution with the GSConv [14] structure, which lightens the model complexity while maintaining the accuracy and can improve the detection speed of wearable devices.

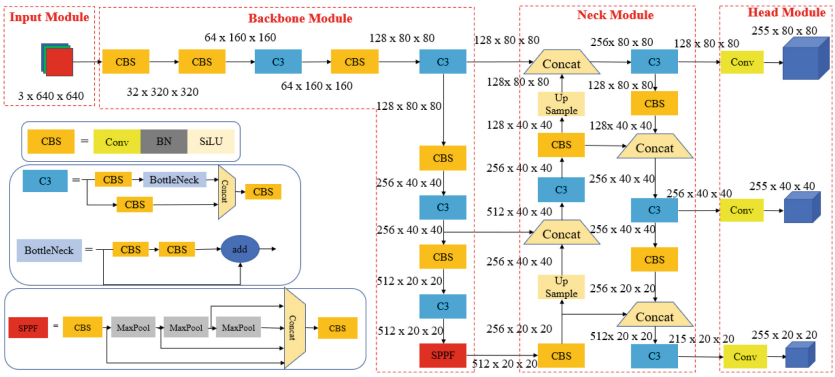


Fig. 1. YOLOv5s overall model structure diagram

3.1 Improvements of Backbone Module

Some of the key points in the image to be detected occupy fewer pixels but contain a lot of semantic information, and the prediction of their location information often has a large error, which can easily produce false detection or missed detection when detecting small targets such as insulated boots and insulated gloves in the image. Introducing attention mechanism in backbone net- works can better improve the detection of small targets. Attention mechanism is a widely used method in various target detection tasks. Attention mechanisms can be broadly classified into two categories according to the imposed dimensions: channel attention and spatial attention. For channel attention mechanisms, the representative mechanisms are effective channel attention (ECA) [15], and for spatial attention mechanisms, the representative mechanisms are squeeze and excitation (SE) [16]. With the proposal of spatial and channel attention mechanisms, it is natural that dual attention mechanisms combining both spatial and channel dimensions are also proposed, representing working convolutional block attention module (CBAM) [17], etc. The above self-attentive mechanisms usually use global pooling to write spatial information, but it is easy to cause the loss of position of the detected target. While PSA can maintain a relatively high resolution in channel and spatial dimensions (maintaining the dimensionality of C/2 in channel and [H,W] in space), this step can reduce the information loss caused by dimensionality reduction, thus improving the attention effect of the neural network on the detection information of small target critical nodes, and thus improving the accuracy of the YOLOv5s network on the detection of small target critical node information.

Therefore, the PSA module is introduced in the Backbone module. The PSA module is divided into two branches, one branch does the self-attention mechanism in channel dimension and the other branch does the self-attention mechanism in spatial dimension, and finally the results of these two branches are fused to get the output of the polarized self-attention structure. The PSA module architecture is shown in Fig. 2.

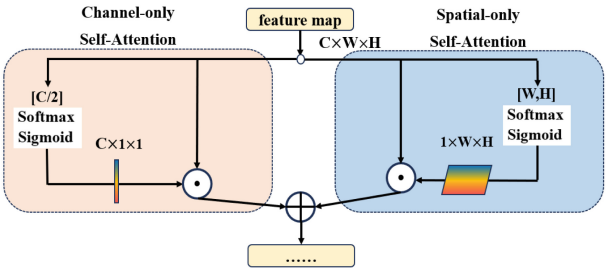


Fig. 2. PSA module architecture diagram

The PSA attention mechanism performs different degrees of compression in the spatial and feature map channel dimensions, respectively. The input fea-

ture map is first converted into two parts, the channel and spatial dimensions, using convolution, in which the information in the spatial dimension is completely compressed, while the information in the channel dimension remains at a relatively high level (i.e., C/2). Since the information in the spatial dimension is compressed, it needs to be augmented with information, so the authors augmented the information in the spatial dimension with Softmax and followed by convolution to raise the dimension of C/2 on the channel to C. Finally, the Sigmoid function is used to keep all the parameters between 0 and 1.

The formula for calculating the weights of channel branches is as follows:

$$\mathrm{A}^{ch}(\mathrm{X}) = \mathrm{F}_{SG}[\mathrm{W}_{z|\theta_1}(\sigma_1(\mathrm{W}_v(\mathrm{X})) \bullet \mathrm{F}_{\mathrm{SM}}(\sigma_2(\mathrm{W}_{\mathrm{q}}(\mathrm{X}))))] \tag{1}$$

where X represents the data in the feature map, $\mathrm{W_q}$, $\mathrm{W_v}$ and $\mathrm{W_z}$ are 1×1 convolution layers respectively, σ_1 and σ_2 are two tensor reshape operators, $\mathrm{F}_{SG}()$ is a SoftMax operator and "$\times$" is the matrix dotproduct operation $\mathrm{F}_{SM}(\mathrm{X}) = \sum_{j=1}^{\mathrm{N}_p} = \frac{e^{x_j}}{\sum_{m=1}^{\mathrm{N}_p} e^{x_m}} x_j$. The internal number of channels, between $\mathrm{W_v}|\mathrm{W_q}$ and $\mathrm{W_z}$, is C/2.

The formula for calculating the weights of spatial branches is as follows:

$$\mathrm{A}^{sp}(\mathrm{X}) = \mathrm{F}_{SG}[\sigma_3(\mathrm{F}_{\mathrm{SM}}(\sigma_1(\mathrm{F}_{GP}(\mathrm{W}_{\mathrm{q}}(\mathrm{X}))))) \bullet \sigma_2(\mathrm{W}_v(\mathrm{X}))] \tag{2}$$

where $\mathrm{W_q}$ and $\mathrm{W_v}$ are standard 1×1 convolution layers respectively, σ_1, σ_2 and σ_3 are three tensor reshape operators, and $\mathrm{F}_{SM}()$ is the SoftMax operator. $\mathrm{F}_{GP}()$ is a global pooling operator $\mathrm{F}_{SG}(\mathrm{X}) = \frac{1}{H \times W} \sum_W^H \sum_{j=1}^W X(:, i, j)$, and $\times$ is the matrix dot-product operation.

For the calculation of the results of the two branches, two fusions were used: parallel (Formula 3) and series(Formula 4) (attention on the channel first, then on the spatial one):

$$\mathrm{PSA}_p(\mathrm{X}) = \mathrm{Z}^{ch} + \mathrm{Z}^{sp} = \mathrm{A}^{ch}(\mathrm{X}) \odot^{ch} \mathrm{X} + \mathrm{A}^{sp}(\mathrm{X}) \odot^{sp} \mathrm{X} \tag{3}$$

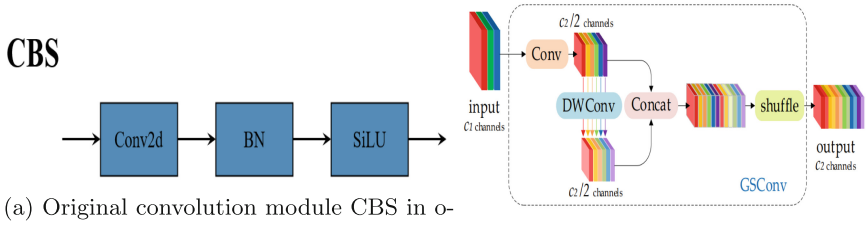
$$\mathrm{PSA}_s(\mathrm{X}) = \mathrm{Z}^{sp}(\mathrm{Z}^{ch}) = \mathrm{A}^{sp}(\mathrm{A}^{ch}(\mathrm{X}) \odot^{ch} \mathrm{X}) \odot^{sp} \mathrm{A}^{ch}(\mathrm{X}) \odot^{ch} \mathrm{X} \tag{4}$$

where "+" is the element-wise addition operator, $\odot^{ch}$ is a channel-wise multiplication operator, $\odot^{sp}$ is a spatial-wise multiplication operator.

3.2 Improvements of Neck Module

In the YOLOv5 version 6.0 algorithm, three functions are encapsulated in the basic convolution module (CBS), including convolution (Conv2d), BN (Batch Normalization) and SiLU function as the activation function, as shown in Fig. 3(a). Meanwhile autopad(k, p) implements the effect of automatic padding calculation. Overall CBS implements the input features through the convolution layer, activation function, and normalization layer to obtain the output layer.

Improvement of Convolutional Networks. Target detection based on YOLOv5 series is a difficult downstream task in computer vision. On the one hand, for the detection of power operation safety wearable equipment in this paper, it is difficult for large models to meet the requirements of real-time detection; on the other hand, the lightweight model constructed by a large number of deep separable convolutional layers cannot achieve sufficient accuracy for multi-attribute wearable devices. Therefore, this paper introduces the latest method GSConv instead of CBS in YOLOv5s Neck module to reduce the complexity of the model and maintain accuracy. GSConv can better balance the accuracy and speed of the model, as shown in Fig. 3(b).



(a) Original convolution module CBS in original YOLOv5s

(b) Convolution module GSConv used to replace CBS

Fig. 3. The structures of the (a) CBS module and the (b) GSConv modules

To accelerate the computation of predictions, the feed images in the CNN must undergo almost a similar transformation process in Backbone: the spatial information is gradually transferred to the channels. And each spatial compression and channel expansion of the feature map results in partial loss of semantic information. Dense convolution computation maximally preserves the hidden connections between each channel, while sparse convolution completely cuts off these connections. GSConv, on the other hand, preserves these connections as much as possible. But if it is used at all stages of the model, the network layers of the model will be deeper, and the deep layers will exacerbate the resistance to the data flow and significantly increase the inference time. By the time these feature maps go to Neck, they have become slender (channel dimension reaches its maximum and width-height dimension reaches its minimum) and no longer need to be transformed. Therefore, a better choice is to use them only at Neck. At this stage, using GSConv to process concatenated feature maps is just right: less redundant repetitive information, no need for compression, and better results for the attention module.

As can be seen from Fig. 3(b), GSConv firstly downsamples the input feature map with a normal convolution to obtain feature map_1, then obtains feature map_2 after DWConv (DSC) deep convolution operation, and then stitches the results of feature map_1 and feature map_2 together; finally, it performs shuffle operation, i.e., randomly disrupts the channel arrangement, so that the corre-

sponding channel numbers of the previous two convolutions are combined to obtain the new feature map.

The purpose of the GSConv operation is to make the output of the DSC as close as possible to the CBS. The information generated by the CBS (dense convolution operation) is permeated into each part of the information generated by the DSC using shuffle. This method allows the information from the CBS to be completely mixed into the output of the DSC without increasing the number of parameters. This method makes the output of the convolutional computation as close as possible to the CBS, while reducing the computational cost.

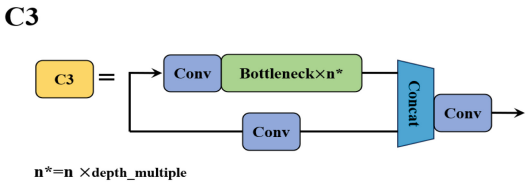


Fig. 4. C3 module structure diagram

VoV-GSCSP Replaces C3. The C3 module is transformed from the BottleneckCSP (bottleneck layer) module, whose structural roles are basically the same both for the CSP architecture, as shown in Fig. 4. This module is the main module for learning the residual features, and its structure is divided into two branches, one using the above specified multiple Bottleneck stacks and three standard convolutional layers, and the other only after a basic convolutional module, and finally the two branches are concat operation. Since C3 uses all base convolution, this necessarily results in a significant increase in the number of parameters.

To lighten the model, we chose to replace C3 with the latest VoV-GSCSP: First, the lightweight convolutional method GSConv is used instead of CBS. Its computational cost is about 60%–70% of that of CBS, and its contribution to the learning ability of wearable device detection in this paper is comparable to the latter. Then, GSbottleneck is continued to be introduced on top of GSConv. as shown in Fig. 5(a). Again, a one-time aggregation approach is used to design the cross-level partial network module VoV-GSCSP. The VoV-GSCSP module reduces the complexity of the computation and network structure, but maintains sufficient accuracy. Figure 5(b) shows the structure of VoV-GSCSP. It is noteworthy that if we use VoV-GSCSP instead of Neck's C3, where the C3 layer consists of standard convolution, the FLOPs (float operations per second) will be reduced by 15.72% on average compared to the latter, and the number of parametres decreases from 7.03M to 5.48 millions in the YOLOv5s model, which achieves the purpose of reducing the model and improving the detection speed of the model while ensuring the detection accuracy of the wearable device.

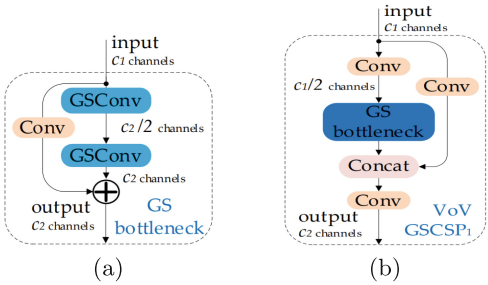


Fig. 5. The structures of the (a) GSbottleneck module and the (b) VoV-GSCSP modules

3.3 Improved YOLOv5s Network Structure

The PSA module was added to the Backbone module of the original YOLOv5s, and then the CBS of the Neck module is replaced with the latest GSConv, and C3 is replaced with VoV-GSCSP. Finally, an improved YOLOv5s model is generated, and its model structure is shown in Fig. 6. The model parameters are 5.48 millions, and the original YOLOv5s model parameters are 7.03 millions. The improved model is 30 % less than the original parameters, and the PAS module is added to the improved model, which can maintain a high resolution of the feature map in the space and channel dimensions, reduce the information loss caused by dimensionality reduction, and is more superior to the feature extraction of small targets such as insulating boots and insulating gloves. GSConv can lighten the convolution module. The VoV-GSCSP module with GSConv as the main body can ensure the detection accuracy and greatly reduce the parameters of the model. It is suitable for the requirements of fast detection speed and high precision for safe wearable devices in power operation.

4 Experimental Results and Analysis

The experimental computer operating system is Windows 10, CPU model is Intel(R) Xeon(R) Platinum8358P CPU@2.6 GHz, GPU model is GeForce RTX3090 with 24 GB memory size and 80 GB memory size. Python 3.8 and GPU acceleration using Cuda 11.8.

4.1 Data Set Production and Processing

In deep learning, the dataset largely influences the final experimental results. In this paper, the data is extracted from the simulated video frames of electric power operation scenes taken under surveillance as the dataset of the algorithm. Through YOLOv5s data cleaning, image enhancement (rotation, contrast change, flip, crop, zoom) and other methods, we finally obtained 2200 images of electric workers wearing protective equipment such as helmets, insulated boots,

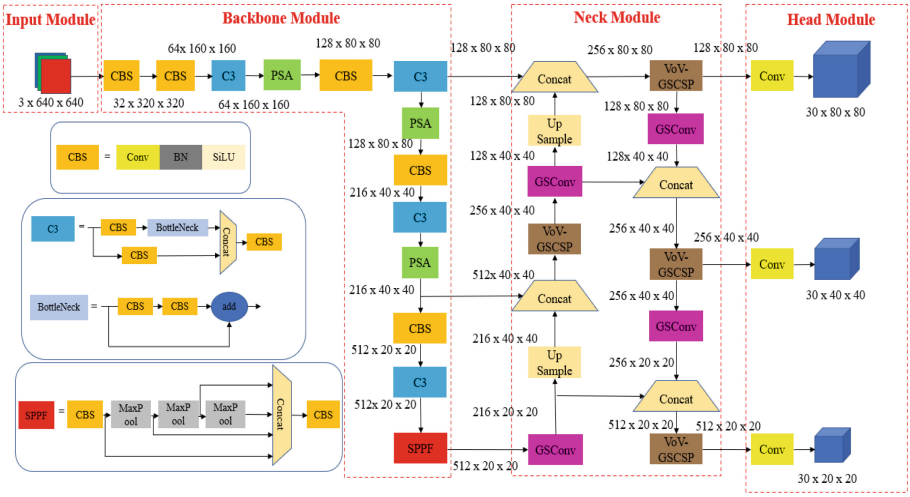


Fig. 6. YOLOv5s-VoVGSCSP-PSA overall model structure diagram

insulated gloves and insulated clothing during operation, including 1800 images in the training set, 250 images in the validation set and 150 images in the test set, and used the Labelimg tool to type The labels are divided into 5 types, namely "worker, shoes, glove, helmet, person". The target detection, the distribution of data samples and the candidate frame data after K-mean clustering are shown in Fig. 7.

From Fig. 8, it can be seen that most of the insulated gloves and insulated boots in the detection images belong to small targets. From the label distribution in Fig. 4(a), we can know that there are about 4,400 insulated gloves and 2,100 insulated boots. From the object size distribution in Fig. 4(b), it is known that the number of small target detection categories accounts for a large percentage.

4.2 Training Network

The hyperparameters of the model are configured before the network training. Initial learning rate lr0 = 0.01, final learning rate lr0 × lrf = 0.001, weight decay coefficient weight_ decay = 0.000 5, learning rate momentum momentum = 0.937, preheat initialization momentum is 0.8, its bias learning rate is 0.1. The effect of weight decay coefficient is to add a regularization term after the loss function with the purpose of reduce the problem of model overfitting; The learning rate momentum is mainly used to initialize the weights of the network when training the network. Using the adam [18] adaptive optimizer, the training period is set to 75 rounds, and then the K-means algorithm is used to re-cluster, and then the rectangular filling training is used to accelerate the model inference process. In the training process, the picture with bad training results in the previous round is also used. The training method of increasing the weight in the next

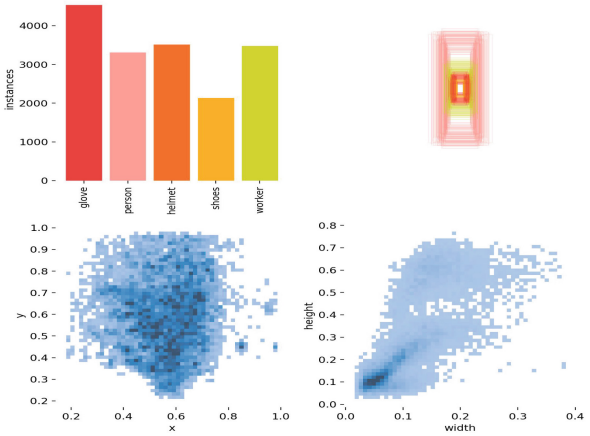


Fig. 7. Distribution of labels and data samples in target detection

round, and adjusting the cosine annealing function value in the hyperparameter and changing the picture shear ratio, flip direction, rotation angle, scaling size, learning rate momentum value, Mixup coefficient, etc. Finally, after repeated training, the best improved YOLOv5s network model is obtained. The improved model training process is shown in Fig. 8. The upper part of the picture is the training IOU loss, confidence loss and classification loss, accuracy and recall rate; the lower part of the picture is the verified IOU loss, confidence loss and classification loss, mAP value and mAP0.5: 0.95 value. It can be seen that the training process is stable and the model fitting ability is excellent.

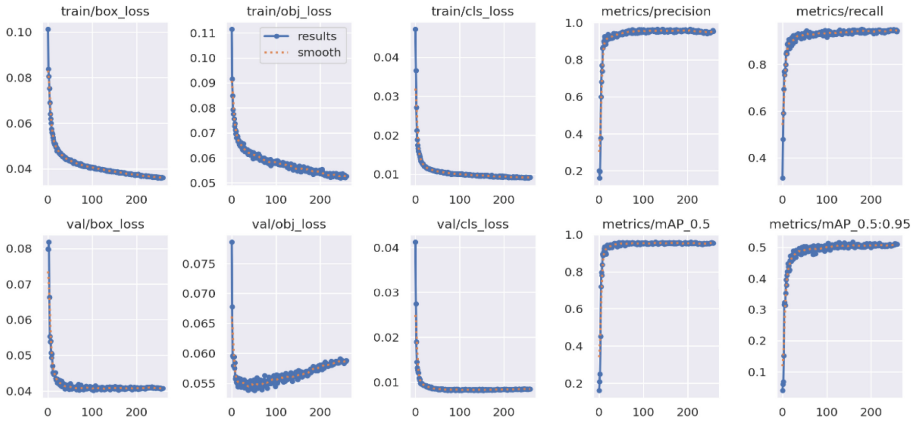


Fig. 8. Model Fitting Process

4.3 Comparison of Model Evaluation Between the Improved Algorithm and the Original Algorithm

Model Evaluation Metrics. The performance of the model needs a good evaluation. In order to prevent the uneven distribution of sample targets, which leads to failure to reflect the actual performance of the model, precision (P), recall (R), average precision (AP) and mean average precision (mAP) are adopted as evaluation indexes. Among them, the precision is mainly for the level of prediction results, while the recall is mainly for the sample itself, and the related formula is as follows:

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

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$\mathrm{AP} = \int_0^1 \mathrm{P(R)dR} \tag{7}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{\mathrm{m}} \mathrm{AP}_i}{\mathrm{m}} \tag{8}$$

where TP(true positives) denotes positive samples predicted by the model as positive classes; FP(false positives) denotes negative samples predicted by the model as positive classes; FN(false negatives) denotes positive samples predicted by the model as negative classes. AP is the area enclosed by the P-R curve (P is the vertical coordinate and R is the horizontal coordinate), mAP is the mean of the average precision AP of all categories, where m is the number of detected categories.

Second, the number of model parameters is also an important indicator to evaluate the performance of a model. Generally speaking, the smaller the number of model parameters, the lighter the model is and the smaller the amount of computation required, measured by FLOPs, the number of floating point operations that can be done per second, and the smaller the hardware requirements of the device, the faster the detection speed of the wearable device in this paper.

The Detection Performance Comparison Between the Improved Model and the Original Model. After repeated experiments, the overall performance comparison results of target detection before and after the improved YOLOv5s algorithm are shown in Table 1. It can be seen from Table 1 that the improved YOLOv5s has higher accuracy, recall rate and mAP value than the original YOLOv5s for all detections. The experimental results show that the average accuracy is increased by 1.58%, up to 96.1%. Moreover, the model parameters are reduced from 7.03 to 5.48 millions, the model size is also reduced from 13.7 MB to 10.7 MB, and the amount of calculation required per second is also reduced from 15.8 GFlops to 12.5 GFLOPs. The lightweight model ensures the accuracy of detection. On the whole, it not only improves the accuracy of

detection, but also meets the requirements of real-time detection of power operation wearable devices in reality.

Table 1. Overall performance comparison results of target detection before and after YOLOv5s algorithm improvement

Model	Parameter	Model Size	P	R	mAP0.5	GFLOPs
YOLOv5s	7037095	13.7 MB	93.1%	94.43%	94.43%	15.8G
YOLOv5s-VoVGSCSP-PSA	5485661	10.7 MB	95.07%	96.01%	96.01%	12.5G

Ablation Experiments. In order to compare the advantages and disadvantages of the models more comprehensively, ablation experiments were also conducted to compare the effects of the added modules on the original models before and after the modules were added to YOLOv5s 6.0. The results of the ablation experiments are shown in Table 2. From Table 2, it can be seen that the PSA module can effectively increase the accuracy of the YOLOv5s model when adding a separate attention module. When adding VoV-GSCSP to the existing attention module, the experimental results of YOLOv5s+VoVGSCSP+PSA model are the best, i.e., the improved model is the highest in detection accuracy, recall, and mean accuracy mean.

Table 2. Results of ablation experiments

Model	no-K_mean	K_mean	PSA	VoVGSCSP	P/%	R/%	mAP0.5/%
YOLOv5s	✓				93.1	93.86	94.43
YOLOv5s-K		✓			94.19	93.77	95.04
YOLOv5s-PSA		✓	✓		95.25	94.24	95.37
YOLOv5s-VoVGSCSP		✓		✓	94.12	94.47	95.52
YOLOv5s-VoVGSCSP-PSA		✓	✓	✓	96.51	95.07	96.01

The Comparison of the Detection Effect Between the Improved Model and the Original Model. The differences of YOLOv5s under different detection conditions before and after the improvement are compared, some of which are shown in Fig. 9. It can be seen from Fig. 9 that the detected object may be affected by occlusion, position, distance and other factors during the detection process, resulting in missed detection or false detection. For example, the detection confidence of the improved YOLOv5s is much larger than that of the original YOLOv5s in the lower left corner insulation boots. As a whole, it can be seen that the improved YOLOv5s model has obvious advantages. By adding PSA

and VoV-GSCSP modules, the network training enhances the feature extraction of key nodes of the object, so that the generated model is better and easier to detect. It not only reduces the missed detection rate and false detection rate of target detection, but also improves the accuracy and speed of detection while lightweighting the model. It has certain advancement in the target detection algorithm of power operation wearable equipment.

(a) Original (b) YOLOv5s (c) Improved YOLOv5s

Fig. 9. Comparison of detection results

5 Conclusion

Aiming at the problems of high missed detection rate, low recognition accuracy and slow detection speed of the traditional detection algorithm for power operation safety wearable devices, an improved YOLOv5s algorithm is proposed. By adding VoV-GSCSP structure and PSA module to lightweight the model, the accuracy and speed of detection are improved. In particular, the detection of partial occlusion of wearable devices has been greatly improved compared with traditional algorithms. The VoV-GSCSP structure added in provides a good help for the model to be lightweight while ensuring the detection accuracy, and the PSA module enables the network model to better extract the feature of the key nodes of the occluded small target, thereby improving the ability to detect the occluded small target. The experimental results show that the improved YOLOv5s network has higher final accuracy and recall rate and better detection effect on the basis of satisfying real-time detection, which greatly improves the robustness and generalization of the algorithm.

Acknowledgment. This work was supported in part by National Natural Science Foundation of China (No. 62067003), Culture and Art Science Planning Project of Jiangxi Province (No. YG2018042), Humanities and Social Science Project of Jiangxi Province (No. JC18224).

References

1. Zhang, Y., Wu, K., et al.: Research based on the improved yolov3 helmet detection method. Comput. Simul. (2021)
2. Liu, Z., Wang, X., et al.: An improved method for infrared image target detection based on yolo algorithm. Laser Infrared **50**(12), 9 (2020)
3. Han, K., Li, S., et al.: Yolov3-based helmet wearing status detection in construction scenarios. J. Rail. Sci. Eng. **018**(001), 268–276 (2021)
4. Liu Xin, Z.C.: Mining helmet wearing detection based on convolutional neural network. Electron. Technol. Appl. **46**(9), 6 (2020)
5. Zhang, M., Cao, Z., et al.: Deep learning based construction worker helmet wearing recognition research. J. Saf. Environ. (2), 7 (2019)
6. Wang, Y., Gu, Y., et al.: Research on helmet wearing detection method based on pose estimation. Comput. Appl. Res. (2021)
7. Qu, W.Q., Qiu, Z.B., et al.: Yolov3-based helmet wear detection for power grid operators. China Saf. Prod. Sci. Technol. **18**(2), 6 (2022)
8. Jian, W.: Visual image detection method of helmet for power construction scene based on improved yolov5 algorithm. Rob. Appl. **2**, 22–26 (2023)
9. Fu, D., Su, G., et al.: Improved yolov5 algorithm based on critical equipment detection for electrical workers' operational safety. J. Hubei Univ. National. Nat. Sci. Ed. **40**(3), 320–327 (2022)
10. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: IEEE Conference on Computer Vision & Pattern Recognition, pp. 6517–6525 (2017)
11. G.J.: Yolov5 (2021)
12. Tan, M., et al.: Efficientnet: rethinking model scaling for convolutional neural networks (2019)
13. Liu, H., Liu, F., Fan, X., Huang, D.: Polarized self-attention: towards high-quality pixel-wise regression (2021)
14. Li, H., et al.: Slim-neck by gsconv: a better design paradigm of detector architectures for autonomous vehicles (2022)
15. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-NET: efficient channel attention for deep convolutional neural networks (2019)
16. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
17. Woo, S., Park, J., Lee, J.Y., Kweon, I.: CBAM: convolutional block attention module. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 3–19 (2018)
18. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. Comput. Sci. (2014)