Mathematical Biosciences and Engineering

MBE, 20(8): 15476–15495.
DOI: 10.3934/mbe.2023691
Received: 18 May 2023
Revised: 17 July 2023
Accepted: 20 July 2023
Published: 25 July 2023

http://www.aimspress.com/journal/MBE

Research article

HDS-YOLOv5: An improved safety harness hook detection algorithm based on YOLOv5s

Mingju Chen[1,3], Zhongxiao Lan[2,*], Zhengxu Duan[2], Sihang Yi[2] and Qin Su[2]

[1] Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin 644002, China
[2] School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644002, China
[3] Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Sichuan University of Science & Engineering, Yibin 644002, China

* **Correspondence:** Email: 321085404231@stu.suse.edu.cn.

Abstract: Improperly using safety harness hooks is a major factor of safety hazards during power maintenance operation. The machine vision-based traditional detection methods have low accuracy and limited real-time effectiveness. In order to quickly discern the status of hooks and reduce safety incidents in the complicated operation environments, three improvements are incorporated in YOLOv5s to construct the novel HDS-YOLOv5 network. First, HOOK-SPPF (spatial pyramid pooling fast) feature extraction module replaces the SPPF backbone network. It can enhance the network's feature extraction capability with less feature loss and extract more distinctive hook features from complex backgrounds. Second, a decoupled head module modified with confidence and regression frames is implemented to reduce negative conflicts between classification and regression, resulting in increased recognition accuracy and accelerated convergence. Lastly, the Scylla intersection over union (SIoU) is employed to optimize the loss function by utilizing the vector angle between the real and predicted frames, thereby improving the model's convergence. Experimental results demonstrate that the HDS-YOLOv5 algorithm achieves a 3% increase in mAP@0.5, reaching 91.2%. Additionally, the algorithm achieves a detection rate of 24.0 FPS (frames per second), demonstrating its superior performance compared to other models.

Keywords: power operation; safety harness hook; YOLOv5; decoupled head; loss function

1. Introduction

The electrical equipment such as power towers and substations is vulnerable to environmental factors like rain, snow and hail as well as equipment failures. To ensure the safety of transmission lines and the stability of the power grid, the power company regularly arranges inspections conducted by personnel. Typically, these inspections take place at high altitudes, making falls from height the most common safety hazard in the power industry. The proper use of safety harnesses is crucial in protecting the lives of personnel during these operations. However, workers are required to move up and down the tower, which means they need to unhook and reconnect the safety harness hook each time. Unfortunately, in an attempt to save time, some workers improperly hang the hook or even neglect to attach it at all during the power operations process.

The proper use of a safety harness is essential for protecting operators and preventing accidents. The non-standard use of hooks is the main cause of accidents, such as hanging on an unstable slope, at a sharp angle or with unclosed hooks. In the environment of electrical power operations, high-intensity work and a complex environment frequently make staff careless, increasing the probability of non-standard use of harness hooks and resulting in safety mishaps. To enhance field monitoring, machine vision technology has been implemented to detect helmets [1], faces [2], safety harnesses [3], transmission lines [4] and recognize anomalous operator behavior [5].

Deep learning has gained popularity as a method for enhancing computer vision performance. Currently, there are two main categories of deep recognition networks: two-stage algorithms such as R-CNN [6], Fast R-CNN [7], Faster R-CNN [8], Mask R-CNN [9], FPN [10] and SPPnet [11] first filter out candidate regions of potential targets from input images before using convolutional neural networks to accurately identify classification and bounding box prediction information; Another is the one-stage algorithm such as SSD [12], DSSD [13], EfficientDet [14], RetinaNet [15], YOLO [16–20] and YOLOX [21] that directly produce predictions without a prefiltering stage. The recently released YOLOv5 model applies methods of improvement such as adaptive anchor frame calculation, mosaic data enhancement and adaptive image scaling to significantly improve both processing speed and accuracy. One-stage algorithms usually have lower accuracy than two-stage algorithms. But the operation is faster and more real-time, making it more suitable for power operation site inspection.

Machine vision technology is based on deep learning technology, which has been applied to the safety control of electric power operation [22], obtaining better results. However, further research is needed to improve the accuracy of identifying safety harness wearing by electric power operation personnel. Fang et al. [23] developed a computer vision-based method utilizing two convolutional neural network (CNN) models to determine whether workers are wearing their harnesses while working at height. Although providing an advantage over manual examination of safety harnesses, this method is still inaccurate and relies on a large amount of data and computational resources. To increase accuracy, Fang et al. [24] proposed a harness detection algorithm based on YOLOv5 and OpenPose network. The dataset is created by using video streams of workers wearing safety harnesses and the networks are trained to detect safety harness. Li et al. [25] Proposed a CME-YOLOv5 network to reduce environ-mental disturbances and mutual occlusion as well as to facilitate the detection of small targets. Zhou et al. [26] proposed a method for detecting insulator defects using an improved YOLOv7 and a multi-UAV cooperative system. However, the efficacy of the detection is contingent upon the states of UAV and weather conditions. Due to its excellent detection performance, the YOLO series has also been widely applied in other fields [27,28]. Lawal [29] designed the YOLO-Tomato model

for detecting tomatoes in complex environmental conditions. Roy and Bhaduri [30] provided an efficient damage classification and localization model based on YOLOv5. Their model addresses the shortcomings of existing deep learning-based damage detection models by offering highly accurate localized bounding box predictions. Ref [31] proposed light-weight object detection method (Efficient-YOLOv5) for detecting safety harness wearing for general construction operations, although it has certain limitations in the electric power operations scene. Moreover, this method solely verifies the presence of safety belts on the workers and does not address the assessment of the hook's status thereby rendering the evaluation of the condition of the safety harness hook impossible.

Given the above issues, in order to address the recognizing the status of the safety harness hook in complex electric power operation scenes. It involves challenging backgrounds, such as farmland, grassland, trees, houses and other electric power facilities, which can cause interference. This research proposes an efficient one-stage deep learning network based on the YOLOv5 network. This paper designs a HOOK-SPPF module to enhance the backbone network and express the target features more accurately in complex backgrounds. Moreover, it adopts the decoupled head [21] for independent implementation of confidence and regression frames, consequently improving the detection accuracy and accelerating the network convergence. Furthermore, the SIoU loss function [32] is invoked to further accelerate model convergence and make the loss function smoother. Finally, extensive experiments are conducted on a homemade Hook dataset to evaluate and verify the performance of the proposed model.

2. YOLOv5s object detection algorithm model

The algorithms in the YOLO (you only look once) series hold a momentous position within the sphere of deep learning for target detection. These algorithms have undergone a steady stream of innovations and improvements from versions 1 to 5. YOLOv5 is the fifth version. It boasts higher precision and faster speed while maintaining a relatively small model size. The YOLOv5 network structure consists of three parts: backbone, neck and head. The backbone is responsible for extracting feature maps, while the head generates detection boxes and predicts classes. The backbone employs two modules, C3 and SPPF (In YOLOv7, the SPPF module is replaced with the SPPCSPC module), which effectively improve the quality and quantity of feature maps. The YOLOv5 utilizes the same coupled head as that of YOLOv3 as the default head, while YOLOX employs a decoupled head as the head. YOLOv5 also introduces some new features. For instance, in backbone, DropBlock regularization is used to enhance the model's robustness. Additionally, MixUp data augmentation method is added to improve model generalization and reduce the risk of overfitting.

The YOLOv5 model comprises five versions: YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, with each version progressively increasing in depth and width. The model details are shown in Table 1. This study specifically focuses on the task of recognizing the state of hanging safety harness hooks, requiring high real-time performance and accuracy. YOLOv5s is the smallest model within the YOLOv5 series, as opposed to larger models like YOLOv5l and YOLOv5x. Its network layers and parameters are relatively minimal, resulting in enhanced inference speed during the process. YOLOv5s exhibits higher detection efficiency and lower hardware requirements. It achieves faster detection speed while ensuring accurate results. As a result, this research focuses on improving and designing YOLOv5s.

Table 1. YOLOv5-6.0 Comparison of metrics by version.

Model	Size (pixels)	Speed V100 (ms)	Parameters (M)	Model Size (MB)
YOLOv5s	640	2.0	7.2	14
YOLOv5m	640	2.7	21.2	40.7
YOLOv5l	640	3.8	46.5	89.2
YOLOv5x	640	6.1	86.7	166

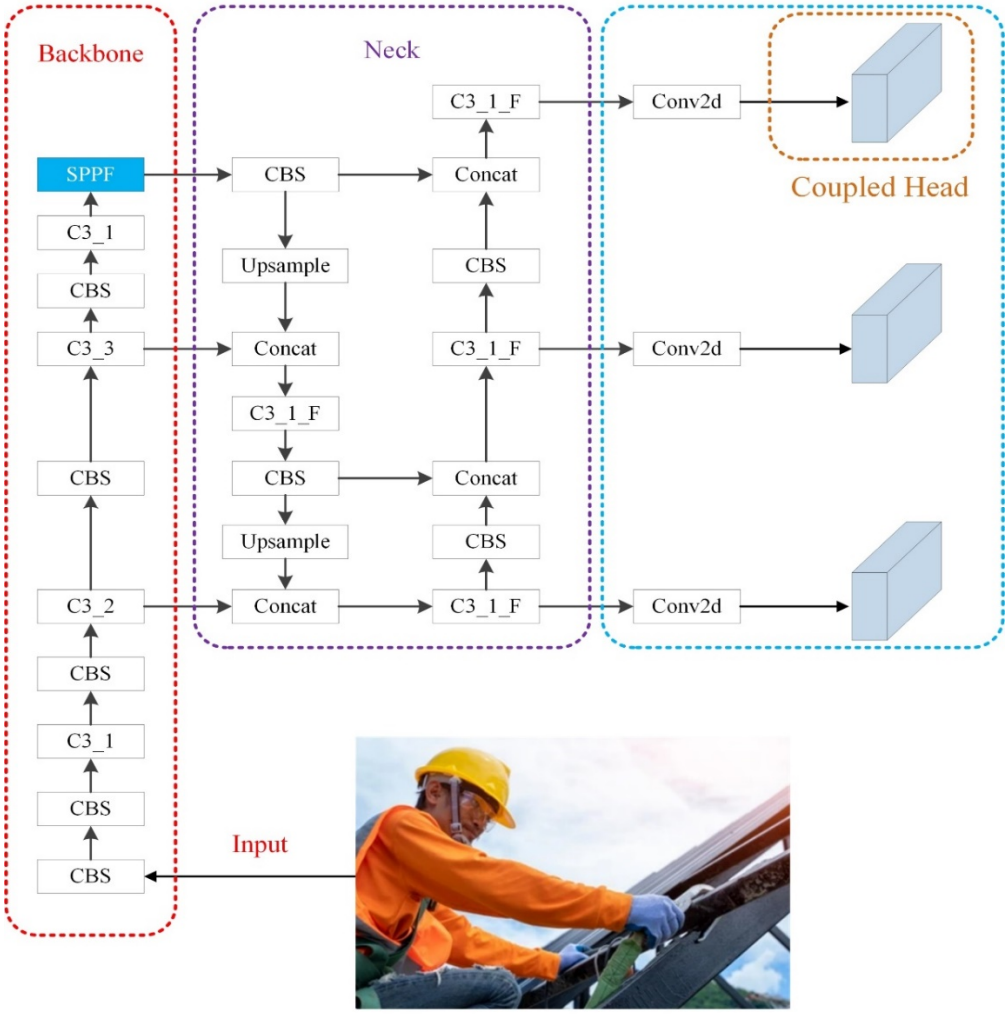


Figure 1. YOLOv5s 6.0 network structure.

3. HDS-YOLOv5 network model

This section goes into greater detail about the proposed improved method for YOLOv5s. including the HOOK-SPPF module, decoupled head and SIoU [32]. HDS-YOLOv5 network architecture ultimately appears as illustrated in Figure 2.

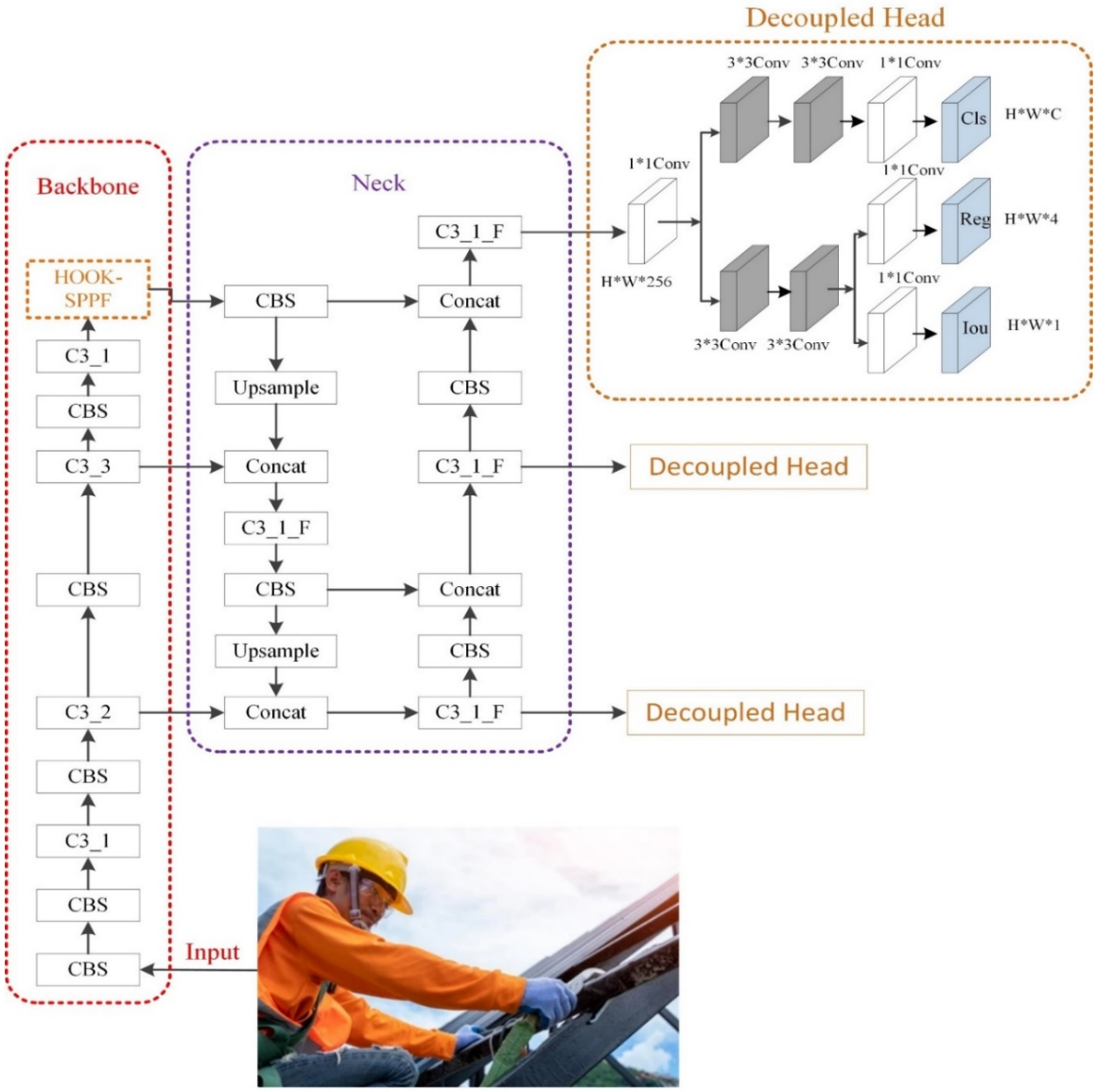


Figure 2. HDS-YOLOv5 network structure.

3.1. Building the HOOK-SPPF module

In the current hook target detection tasks, there are often problems with targets being confused with the background, difficulties in extracting small feature hooks and multiple overlapping dense categories of targets. As shown in Figure 3, the hook target is relatively small and the color of the hook is similar to that of the power tower. This results in high performance requirements for hook target detection models in practical applications. To solve this problem, this chapter proposes a HOOK-SPPF structure to further enhance the features extracted by the backbone network, enabling the network to improve the ability to recognize and classify safety harness hooks in complex environments.

The SPPF used in YOLOv5s is an enhanced version of the spatial pyramid pooling (SPP) [33] module, as shown in Figure 4. SPPF establishes connections between each pooling layer, preserving more feature information and enhancing the network's receptive field. This optimization retains the advantages of SPP while further improving the calculation speed.

Figure 3. The small and indistinguishable hook.

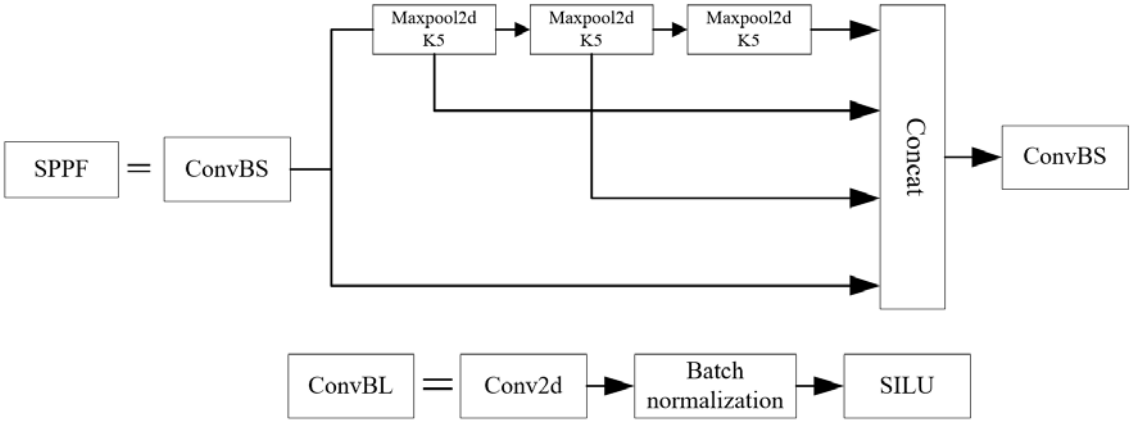


Figure 4. SPPF module structure.

HOOK-SPPF is a module that incorporates SPPF into cross stage partial network (CSPNet) [34], as illustrated in Figure 5. CSPNet optimizes the problem of duplicate gradients in the network by integrating the feature maps at the beginning and end of the network, achieving a 20% reduction in computational requirements while achieving the same or even higher accuracy. CSPNet divides the input feature map into two parts. These parts are then merged through the cross-stage hierarchy structure. By separating the gradient flow and propagating it through different network paths, the network achieves greater diversity in gradient combinations, thus enhancing both the speed and accuracy of network inference.

HOOK-SPPF is divided into two parts. The first part convolves, normalizes and activates the feature information extracted from the backbone network with the RELU activation function. This part plays a role in auxiliary optimization. It also retains the positional information contained in the input feature layer. The other part first convolves, normalizes and activates the feature information three times with the RELU activation function to extract deep image information. Subsequently, the SPPF structure increases the receptive field size and two convolutions are applied, followed by batch normalization and RELU activation functions to extract the features. At this point, the feature layer contains more semantic information. HOOK-SPPF stacks these two parts together, greatly improving

the network's ability to learn multiscale features while reducing the number of parameters and enhancing the accuracy of detection.

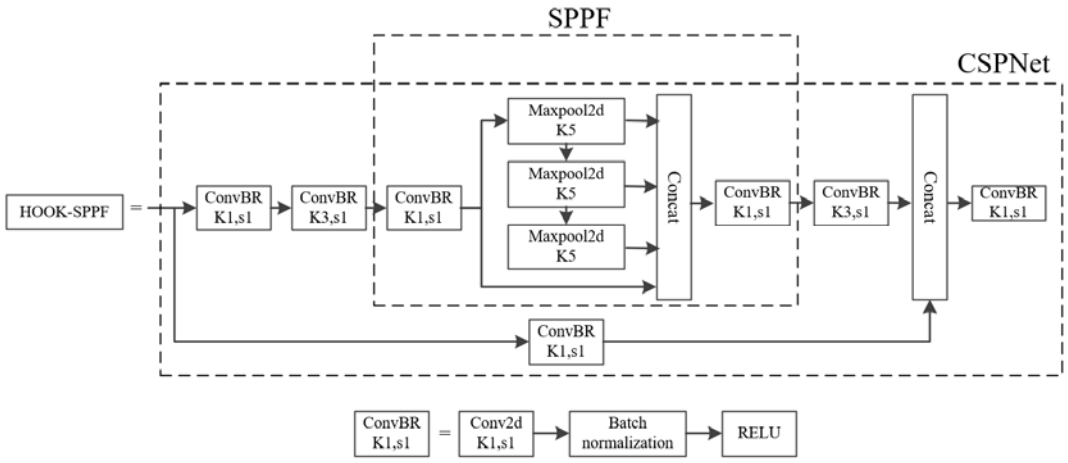


Figure 5. The HOOK-SPPF module structure.

The HOOK-SPPF structure inherits the advantages of the SPPF structure, which includes adaptive size output without distortion, lower model computational complexity and faster processing speed, while avoiding repeated image feature extraction. These advantages contribute to the HOOK-SPPF structure's effectiveness in detecting hooks.

3.2. Decoupled head

The YOLO family's backbones and feature pyramids have evolved, while keeping their detection heads coupled. YOLOv5s utilizes a coupled head with a 1 × 1 convolution to confidently finalize the classification and regression frame. Figure 6 demonstrates the implementation of the coupled head.

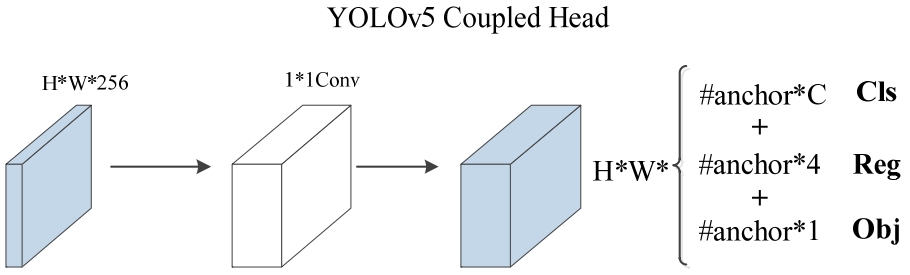


Figure 6. The coupled head structure.

The structure of the decoupled head is illustrated in Figure 7. For the given input feature layer, the decoupled head employs a 1x1 convolution to reduce its dimension. It further utilizes two sets of 3x3 convolutions in each classification and regression branch with parallel channels dedicated to object

classification and target frame coordinate regression tasks, respectively. This processing generates three outputs: Cls, Reg and Obj. Cls represents the category corresponding to the target frame, Reg represents the location information of the target frame and Obj indicates whether each feature point contains an object. All three output values are combined to generate the final prediction information. The decoupling operation separates the confidence degree and regression frame, slightly increasing the complexity of the process. However, it alleviates the negative impact caused by the conflict between the classification and regression tasks [35,36], ultimately improving the detection accuracy of the network and accelerating network training convergence.

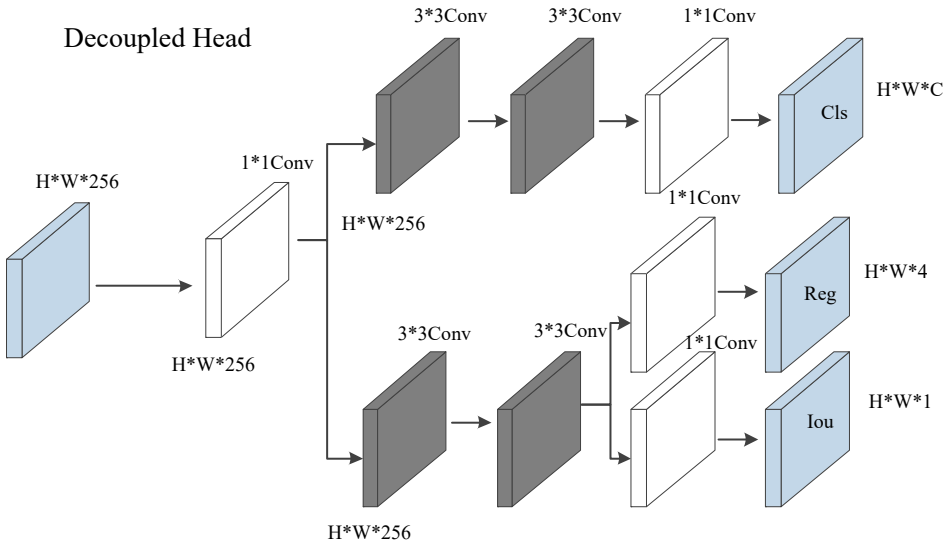


Figure 7. The decoupled head structure.

3.3. Loss function

The loss function of YOLOv5s measures the distance between the predicted information and the expected information (label) of the Neural Network. The closer the predicted information aligns with the expected information, the smaller the value of the loss function becomes. This loss function consists of three components: rectangular box loss (loss_rect), confidence loss (loss_obj) and classification loss (loss_cls). The overall loss is calculated as the weighted sum of these three components with the flexibility to adjust the emphasis on each loss by modifying the weights. The YOLOv5s loss function can be expressed using the following formula:

$$Loss = a * loss\_obj + b * loss\_rect + c * loss\_cls \tag{1}$$

YOLOv5s uses complete intersection over union (CIoU) loss [37] to calculate the rectangular box loss (loss_rect), confidence loss and classification loss with BCE loss and CIoU loss is calculated as:

$$Loss_{CIOU} = 1 - IOU + \frac{\rho^2}{c^2} + \alpha v \tag{2}$$

$$\alpha = \frac{v}{1 - IOU + v} \tag{3}$$

$$v = \frac{4}{\pi}\left(\arctan\frac{w_l}{h_l} - \arctan\frac{w_p}{h_p}\right)^2 \tag{4}$$

where ρ is the distance between the center points of prediction bounding box A and real bounding box B. c represents the diagonal length of the minimum bounding rectangle of box A and box B. v and represents the aspect ratio similarity of box A and box B. α is the influence factor of v.

In Eq (4), The value range of the arctan function is $0\text{-}\frac{\pi}{2}$; then the value range is 0–1. When the width-to-height ratio of prediction bounding box A and real bounding box B are equal, $v = 0$. At this time, the influence factor v of α is also equal to 0. The αv in Eq (2) does not work. In this case, the CIoU loss function does not get a stable expression.

In this regard, the SIoU loss function is chosen to replace the original CIoU loss function. The vector angle between the true and predicted frames is further considered to redefine the associated loss function, which contains four components: angle cost, distance cost, shape cost and IoU cost. The SIoU schematic is illustrated in Figure 8.

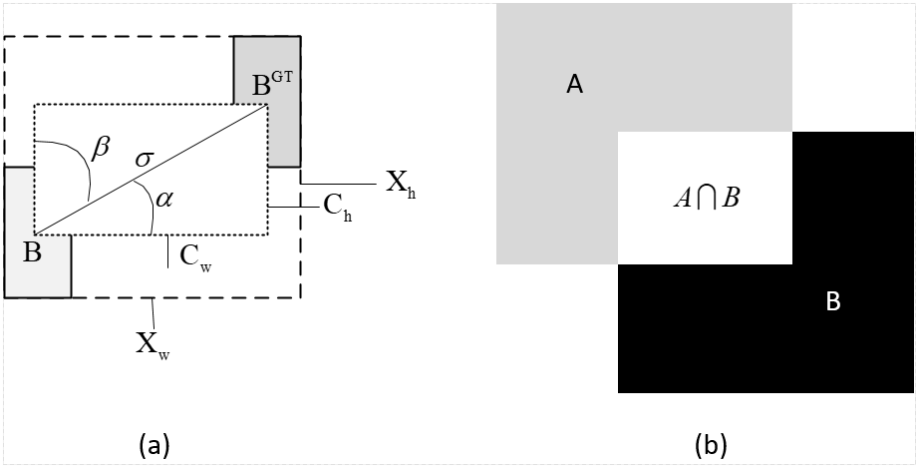


Figure 8. Schematic diagram of the calculation: (a) schematic diagram of the SIoU; (b) schematic diagram of IoU calculation.

3.3.1 Angle cost

$$\Lambda = 1 - 2 * \sin^2\left(\arcsin\left(\frac{C_h}{\sigma}\right) - \frac{\pi}{4}\right) \tag{5}$$

$$\sigma = \sqrt{\left(b_{c_x}^{gt} - b_{c_x}\right)^2 + \left(b_{c_y}^{gt} - b_{c_y}\right)^2} \tag{6}$$

$$\mathrm{C}_h = \max\left(b_{c_y}^{gt}, b_{c_y}\right) - \min\left(b_{c_y}^{gt}, b_{c_y}\right) \tag{7}$$

In Eq (6), where σ represents the distance between the center points of prediction bounding box A and real bounding box B. In Eq (7), C_h is the difference of height between the center point of the real bounding box and the predicted bounding box. $b_{c_x}^{gt}$, $b_{c_y}^{gt}$ are the real bounding box center coordinate. b_{c_x}, b_{c_y} are the predicted bounding box center coordinate.

3.3.2 Distance cost

$$\Delta = \sum_{t=x,y}\left(1 - e^{-\gamma\rho t}\right) \tag{8}$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{\mathrm{X}_w}\right)^2 \tag{9}$$

$$\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{\mathrm{X}_h}\right)^2 \tag{10}$$

$$\gamma = 2 - \Lambda \tag{11}$$

In Eqs (9) and (10), X_w and X_h are the width and height of the minimum bounding rectangle of the real bounding box and the prediction bounding box. As the angle increases, γ is assigned a value of time-preferred distance.

3.3.3 Shape cost

The definition formula of shape loss is shown in Eq (12):

$$\Omega = \sum_{t=w,h}\left(1 - e^{-w_t}\right)^{\theta} \tag{12}$$

$$W_w = \frac{\left|w - w^{gt}\right|}{\max\left(w, w^{gt}\right)} \tag{13}$$

$$W_h = \frac{\left|h - h^{gt}\right|}{\max\left(h, h^{gt}\right)} \tag{14}$$

w, h, w^{gt}, h^{gt} are the width and height of the prediction and real bounding box. In order to avoid paying too much attention to shape cost and reduce the movement of the prediction bounding box, this paper sets θ to 2.

In summary, the final definition of the SIoU loss function is shown in Eq (15):

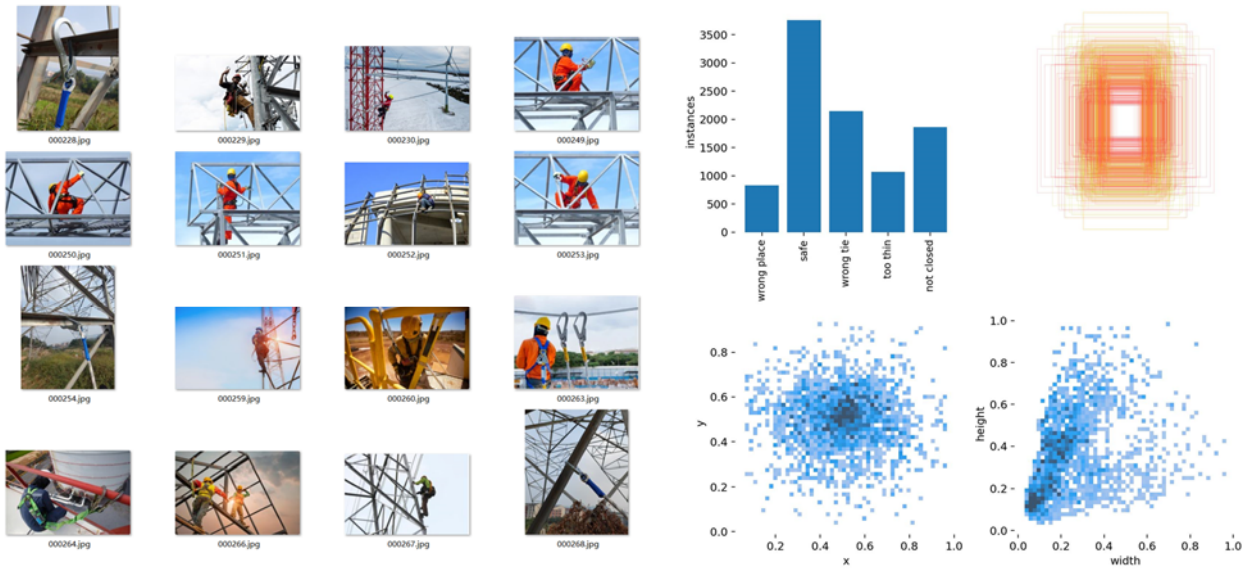
$$Loss_{SIOU} = 1 - IOU + \frac{\Delta + \Omega}{2} \tag{15}$$

Due to the increased angle cost, the loss function is more fully expressed, reducing the likelihood of obtaining a zero-penalty term and facilitating the smoother convergence of the final loss function. In turn, this enhances regression accuracy and minimizes prediction errors.

4. Experiments and results

4.1. Dataset

To assess the detection performance of the improved YOLOv5s algorithm in this paper, the hook dataset was created using selfies and images obtained from the internet. The objective was to enable the algorithm to achieve better hook detection results under various complex scenes and extreme weather conditions. For this purpose, hooks were initially hung in different ways and at various locations on a domestic electric power tower, simulating the common arrangement of safety harness hooks during the operations of electric workers. Subsequently, a Xiaomi 11 cell phone was used to capture photos of the hooks, ensuring variations in lighting conditions, time periods (noon, evening, etc.), distances and focal lengths.



(a)Partial Picture of the Hook Dataset (b)Number of samples in each category

Figure 9. Hook dataset.

After undergoing collation, a total of 3378 photos of safety harness hooks encompassing four types of violations, one type of safety and five types of hook hanging were obtained. Considering the small size of the dataset, precautions were taken to prevent the overfitting phenomenon resulting from an insufficient number of samples, which could potentially affect the detection effectiveness of the seat belt hook. Consequently, a data enhancement tool was employed to expand the original dataset. This

involved augmenting the images through random rectangle masking and horizontal flipping, boosting the total count to 9738. By doing so, the scale of the training set was effectively increased, enhancing the model's ability to generalize. Furthermore, the Labelimg annotation software was utilized to annotate each image according to the required txt format for YOLOv5. Finally, the dataset was split into three parts: a training set, a validation set and a test set distributed in an 8:1:1 ratio. Figure 9 displays examples of images from the dataset.

4.2. Experimental environment and parameter setting

The hook detection method proposed in this research was implemented in a Windows 10 Professional environment. PyTorch, a deep learning framework, was utilized for model construction, training and testing. The programming software of choice was PyCharm community edition. To expedite the model training process, CUDA and CUDNN were employed for acceleration. A comprehensive list of the training platform parameters is provided in Table 2.

Table 2. Experimental operating environment.

Name	Configuration
CPU	Intel(R)CPU E5-2695 v4
GPU	NVIDIA TITAN XP
Memory	256 G
CUDA	11.1.96
CUDNN	8.4
Pytorch	1.10.2

The training configuration for the improved YOLOv5s model is outlined as follows: the size of the input image is set to 640 × 640, the number of epochs is specified as 300, the batch size is set at 32, the initial learning rate is defined as 0.01 and the weight decay is established as 0.0005.

4.3. Evaluation metrics

To provide a comprehensive and objective evaluation of the improved YOLOv5s model proposed in this paper, precision and recall are employed as commonly used evaluation metrics for neural network models. Precision represents the ratio of correctly predicted targets to all predicted targets, while recall indicates the ratio of correctly predicted targets to all actual (correct) targets. The calculation of precision and recall is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

where true positive (TP) refers to the correct target in the predicted target, false positive (FP) refers to the wrong target in the predicted target and false negative (FN) refers to the right target that is not predicted. Where the target prediction is considered correct when IoU ≥ the threshold and incorrect

when IoU < threshold. In this paper, the detection threshold is set to 0.5, when the IoU value between the detection box and the real box exceeds 0.5 the detection box is considered accurate. The two most common evaluation metrics for target detection tasks are the average precision (AP) and the mAP. AP is the area enclosed by the curve of different accuracy and recall rates. Generally, the classifier exhibits superior performance as the AP value increases. A larger value indicates better detection accuracy by the network model, while MAP represents the average AP value calculated across all categories. Its value ranges from 0 to 1. The calculation formula is given as follows:

$$AP = \frac{1}{m}\sum_{i}^{m} P_i \tag{18}$$

$$mAP = \frac{1}{c}\sum_{j}^{c} AP_j \tag{19}$$

In this paper, N = 5 represents the number of target detection categories. The measure used to evaluate the model's detection effectiveness in this paper is mAP@0.5. This allows us to measure the comprehensive performance of the model under different IoU thresholds. Higher numbers suggest a better model effect and a more accurate fit between the predicted and real bounding boxes.

4.4. Analysis of experimental results

4.4.1. Ablation experiment

Table 3. Results of the ablation experiments.

Model	mAP @0.5	AP				
		Wrong place	Safe	Wrong tie	Too thin	Not closed
Original YOLOv5s (SPPF+CIoU+Coupled Head)	0.882	0.836	0.919	0.913	0.908	0.837
YOLOv5s-1 (SPPF+SIoU+Coupled Head)	0.890	0.843	0.931	0.92	0.917	0.839
YOLOv5s-2 (SPPF+SIoU+Decoupled Head)	0.900	0.902	0.885	0.961	0.904	0.849
YOLOv5s-3 (HOOK-SPPF+SIoU+Coupled Head）	0.908	0.875	0.945	0.933	0.932	0.857
HDS-YOLOv5 (HOOK-SPPF+SIoU+Decoupled Head)	0.912	0.919	0.900	0.947	0.957	0.870

In this study, we conducted ablation experiments to comprehensively validate the optimization impact of each enhancement module. Specifically, we set up multiple ablation experiments between YOLOv5s (SPPF + CIoU + Coupled Head), YOLOv5s-1 (SPPF + SIoU + Coupled Head), YOLOv5s-2 (SPPF + SIoU + Decoupled Head), YOLOv5s-3 (HOOK-SPPF + SIoU + Coupled Head), and HDS-YOLOv5 (HOOK-SPPF + SIoU + De-coupled Head). Table 3 shows the results of the experiment.

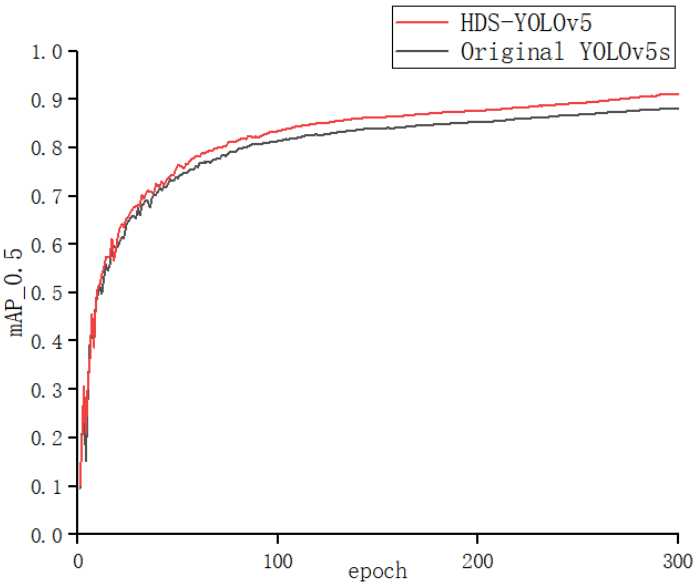


Figure 10. Comparisons of mAP@0.5 curve.

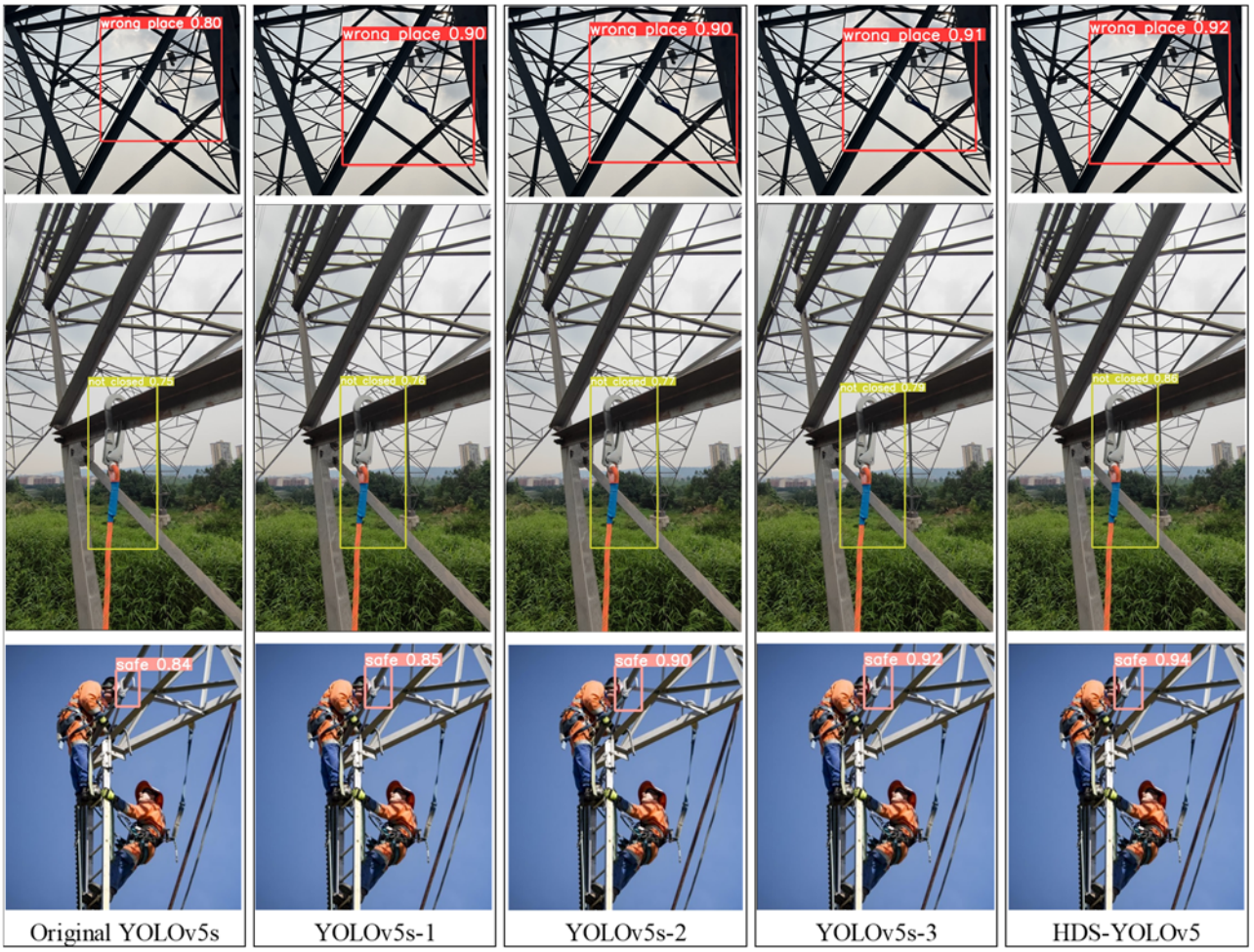


Figure 11. Detection results of ablation experiments.

The mAP@0.5 of the original YOLOv5s is 88.2%. YOLOv5s-1 replaced the original CIoU with SIoU, resulting in an increase in mAP to 89.0%. This indicates that SIoU improves the accuracy of the network. YOLOv5s-2 integrated SIoU and decoupled head, further improving the mAP to 90.0%. YOLOv5s-3 utilized SIoU and HOOK-SPPF, leading to an improvement in mAP to 90.8%. HDS-YOLOv5 incorporates SIoU, decoupled head, and HOOK-SPPF, significantly enhancing the model's accuracy and achieving a 91.2% mAP, which is 3.0% higher than the original YOLOv5s. The effectiveness and superiority of the proposed network as described in this paper are clearly demonstrated.

To demonstrate the increased efficiency of the model more clearly, we tested it on several images from the test set and the results are shown in Figure 11. It is evident from Figure 11 that the improved algorithm HDS-YOLOv5 successfully extracts the desired features and the performance of the safety harness hook suspension method is better than YOLOv5s in both backlit conditions and complex background environments. Figure 10 shows the change curve of the mAP@0.5 during training.

4.4.2. Comparison experiment

To assess the effectiveness of the object detection methods described in this paper, comparative tests were performed between the improved model proposed in this study and various existing algorithms including fast R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5 and YOLOv7. The comparative experiment was conducted using identical experimental settings and dataset with four metrics：mAP@0.5, mAP@0.5:0.95, FPS and model size utilized as measurement criteria. The findings of these experiments are presented in Table 4.

Table 4. Experimental comparisons with other models.

Model	mAP@0.5	mAP@0.5:0.95	FPS	Model Size/MB
SSD	0.894	0.592	10.1	92.6
Faster R-CNN	0.908	0.612	2.1	521
YOLOv3	0.71	0.430	18	118
YOLOv4	0.783	0.508	3.8	244
YOLOv5	0.882	0.620	27	14
YOLOv7	0.90	0.623	33	72
HDS-YOLOv5	0.912	0.638	24.0	33.3

As shown in Table 4, when compared to the SSD, faster R-CNN and YOLOv4 models, the HDS-YOLOv5 model significantly reduces the model size while greatly improving detection speed. In comparison to the original YOLOv5s and the YOLOv7 model, our improved model achieves the highest mAP value. Although the detection speed (FPS) of our improved model is lower than the original YOLOv5s and YOLOv7 models, a comprehensive analysis of the experimental results shows that the improved YOLOv5 model strikes a balance between detection speed and performance, resulting in superior overall performance.

To intuitively verify the effectiveness of the improved algorithm in object detection and its robustness in different complex settings, we selected the same test datasets for experimental comparison of SSD, Faster R-CNN, YOLOv4, YOLOv5 and YOLOv7. The experimental results are depicted in Figure 12.

(a) FASTER RCNN

(b) SSD

(c) Original YOLOV5s

(d) YOLOv7

Continued on next page

(e) HDS-YOLOV5

Figure 12. Comparison of detection results of our model with other models.

From Figure 12, it is evident that all five detection algorithms effectively identify the suspended state of the hooks. However, the superiority of the HDS-YOLOv5 model can be intuitively observed from the specific results. The HDS-YOLOv5 model exhibits a greater level of confidence in detecting the hooks. Moreover, the HDS-YOLOv5 showcases improved accuracy and stability when detecting targets, particularly in complex backgrounds, thereby substantially enhancing the detection capabilities of hook-shaped objects.

5. Conclusions

This article presents HDS-YOLOv5, an algorithm designed to identify the suspended state of safety harness hooks during power tower inspections. The HOOK-SPPF module is designed to enhance the feature extraction capability of the backbone network and improve the model's ability to extract deep and crucial features of the hooks. The decoupled head replaces the coupled head in the original network, reducing the negative conflict between the classification and regression tasks thereby improving accuracy and reducing missed detections of hooks in complex environments. The CIoU loss function is replaced by the SIoU loss function, SIoU further considers the vector angle between the real box and the predicted box, redefining four loss functions: angle cost, distance cost, shape cost and IoU cost thereby enhancing the regression accuracy of the model. Comparative experiments were conducted on a homemade hook dataset. The improved model achieves a 3% increase in mAP@0.5, reaching 91.2%, compared to the original network. However, the detection speed of the improved model is 24FPS, which is 3FPS slower than the original YOLOv5s. To address this issue, the next optimization plan involves implementing a more lightweight MobileNetV3 network as the backbone network for YOLOv5, further reducing model parameters and computational load to improve detection speed.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part by the Project of Sichuan Provincial Science and Technology Department under grant 2022ZHCG0035 and 2023NSFSC1987, the Artificial Intelligence Key Laboratory Project of Sichuan Province under grant 2021RYY04, the Opening Project of Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things under grant 2021WYY01 and Sichuan University of Science and Engineering Postgraduate Innovation Fund in 2022 under grant Y2022113.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. Li, H. Liu, T. Wang, M. Jiang, S. Wang, K. Li, et al., Safety helmet wearing detection based on image processing and machine learning, in *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)*, (2017), 201–205. https://doi.org/10.1109/icaci.2017.7974509
2. X. Xu, X. Wang, Z. Q. Sun, S. X. Wang, Face recognition technology based on CNN, XGBoost, model fusion and its application for safety management in power system, in *IOP Conference Series: Earth and Environmental Science*, **645** (2021), 012054. https://dx.doi.org/10.1088/1755-1315/645/1/012054
3. Z. Sun, Y. Xuan, L. Fan, R. Han, Y. Tu, J. Huang, et al., Security monitoring strategy of distribution community operation site based on intelligent image processing method, *Front. Energy Res.*, **10** (2022), 931515. https://doi.org/10.3389/fenrg.2022.931515
4. B. Weng, W. Gao, W. Zheng, G. Yang, Newly designed identifying method for ice thickness on high-voltage transmission lines via machine vision, *High Voltage*, **6** (2021), 904–922. https://doi.org/10.1049/hve2.12086
5. J. X. Li, Y. Y. Liu, H. Wang, Safety supervision method of power work site based on computer machine learning and image recognition, *J. Phys. Conf. Ser.*, **2074** (2021), 012021. https://dx.doi.org/10.1088/1742-6596/2074/1/012021
6. R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 580–587. https://doi.org/10.1109/cvpr.2014.81
7. R. Girshick, Fast R-CNN, in *2015 IEEE International Conference on Computer Vision (ICCV)*, (2015), 1440–1448. https://doi.org/10.1109/iccv.2015.169
8. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031
9. K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 2980–2988. http://doi.org/10.1109/ICCV.2017.322
10. T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 936–944. http://doi.org/10.1109/CVPR.2017.106

11. K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, **37** (2015), 1904–1916. http://doi.org/10.1109/TPAMI.2015.2389824
12. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Y. Fu, et al., SSD: Single shot multibox detector, in *European Conference on Computer Vision*, preprint, arXiv:1512.02325. https://doi.org/10.48550/arXiv.1512.02325
13. C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, A. C. Berg, DSSD: Deconvolutional single shot detector, preprint, arXiv:1701.06659. http://doi.org/10.48550/arXiv.1701.06659
14. M. Tan, R. Pang, Q. V. Le, EfficientDet: Scalable and efficient object detection, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 10778–10787. http://doi.org/10.1109/CVPR42600.2020.01079
15. T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 318–327. http://doi.org/10.1109/TPAMI.2018.2858826
16. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788. http://doi.org/10.1109/CVPR.2016.91
17. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 6517–6525. http://doi.org/10.1109/CVPR.2017.690
18. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, preprint, arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767
19. A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, preprint, arXiv:2004.10934. https://doi.org/10.48550/arXiv.2004.10934
20. C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, preprint, arXiv:2207.02696. https://doi.org/10.48550/arXiv.2207.02696
21. Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, YOLOX: Exceeding YOLO series in 2021, preprint, arXiv:2107.08430. https://doi.org/10.48550/arXiv.2107.08430
22. K. Yan, Q. Li, H. Li, H. Wang, Y. Fang, L. Xing, et al., Deep learning-based substation remote construction management and AI automatic violation detection system, *IET Gener. Transm. Distrib.*, **16** (2022), 1714–1726. https://doi.org/10.1049/gtd2.12387
23. W. Fang, L. Ding, H. Luo, P. E. D. Love, Falls from heights: A computer vision-based approach for safety harness detection, *Autom. Constr.*, **91** (2018), 53–61. https://doi.org/10.1016/j.autcon.2018.02.018
24. C. Fang, H. Xiang, C. Leng, J. Chen, Q. Yu, Research on real-time detection of safety harness wearing of workshop personnel based on YOLOv5 and OpenPose, *Sustainability*, **14** (2022), 5872. https://doi.org/10.3390/su14105872
25. J. Li, C. Liu, X. Lu, B. Wu, CME-YOLOv5: An efficient object detection network for densely spaced fish and small targets, *Water*, **14** (2022), 2412. https://doi.org/10.3390/w14152412
26. R. Chang, S. Zhou, Y. Zhang, N. Zhang, C. Zhou, M. Li, Research on insulator defect detection based on improved YOLOv7 and multi-UAV cooperative system, *Coatings*, **13** (2023), 880. https://doi.org/10.3390/coatings13050880
27. M. Chen, Z. Duan, Z. Lan, S. Yi, Scene reconstruction algorithm for unstructured weak-texture regions based on stereo vision, *Appl. Sci.*, **13** (2023), 6407. https://doi.org/10.3390/app13116407

28. M. J. Chen, T. T Liu, X. Z. Xiong, Z. X. Duan, A. L. Cui, A transformer-based cross-window aggregated attentional image inpainting model, *Electronics*, **12** (2023), 2726. https://doi.org/10.3390/electronics12122726
29. M. O. Lawal, Tomato detection based on modified YOLOv3 framework, *Sci. Rep.*, **11** (2021), 1447. https://doi.org/10.1038/s41598-021-81216-5
30. A. M. Roy, J. Bhaduri, DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism, *Adv. Eng. Inf.*, **56** (2023), 102007. https://doi.org/10.1016/j.aei.2023.102007
31. Z. Xu, J. Huang, K. Huang, A novel computer vision-based approach for monitoring safety harness use in construction, *IET Image Process.*, **17** (2023), 1071–1085. https://doi.org/10.1049/ipr2.12696
32. Z. Gevorgyan, SIoU Loss: More powerful learning for bounding box regression, preprint, arXiv:2205.12740. https://doi.org/10.48550/arXiv.2205.12740
33. K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, **37** (2014), 1904–1916. https://doi.org/10.1007/978-3-319-10578-9_23
34. C. Y. Wang, H. Y. M. Liao, I. H. Yeh, Y. H. Wu, P. Y. Chen, J. W. Hsieh, CSPNet: A new backbone that can enhance learning capability of CNN, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2019), 1571–1580. https://doi.org/10.48550/arXiv.1911.11929
35. Y. Wu, Y. Chen, L. Yuan, Z. Liu, L. Wang, H. Li, et al., Rethinking classification and localization for object detection, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 10183–10192. https://doi.org/10.1109/cvpr42600.2020.01020
36. G. Song, Y. Liu, X. Wang, Revisiting the sibling head in object detector, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 11560–11569. https://doi.org/10.1109/cvpr42600.2020.01158
37. Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, et al., Enhancing geometric factors in model learning and inference for object detection and instance segmentation, *IEEE Trans. Cybern.*, **52** (2022), 8574–8586. https://doi.org/10.1109/tcyb.2021.3095305

AIMS Press