IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 January 2024, accepted 5 February 2024, date of publication 13 February 2024, date of current version 20 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365530

RESEARCH ARTICLE

YOLO-ESCA: A High-Performance Safety Helmet Standard Wearing Behavior Detection Model Based on Improved YOLOv5

PEIJIAN JIN, HANG LI, WEILONG YAN, AND JINRONG XU
School of Emergency Science and Engineering, Jilin Jianzhu University, Changchun, Jilin 130118, China

Corresponding author: Hang Li (li1520133943@gmail.com)

This work was supported in part by the Ministry of Emergency Management of the People's Republic of China under Grant Jilin-0050-2018AQ.

ABSTRACT To solve the problem of workers incorrectly wearing helmets, this study proposes a standard helmet wear detection model, YOLO-ESCA based on improved YOLOv5n. This model can monitor workers' helmet wear in real time via UAVs and other means and automatically reduce video streaming detection results. The model is trained using a self-built dataset that containing 4400 images. To address the shortcomings of the original YOLOv5, an improved version of the proposed approach, in which the efficient intersection over union loss function (EIOU-loss), Soft-NMS nonmaximal suppression, and the convolutional block attention module (CBAM) are employed, is proposed, and a small target detection layer (ADL) is added to improve model performance. The experimental results show that the mAP@0.5 of the improved model is up to 94.7%, the FPS is up to 65.3, the model size is only 4.47MB, and that the number of detections on the self-constructed dataset and SHWD dataset is 41.7% and 73% greater, respectively, than that of the original model, respectively.

INDEX TERMS Convolutional neural networks, deep learning, object detection, safety management.

I. INTRODUCTION

Safety accidents frequently occur in the construction industry due to factors such as labor intensiveness, the intersection of multiple processes, and complex operating environments. Safety helmets, one of the "three treasures" of construction, can prevent most of the injuries that occur during the construction process. Moreover, wearing a helmet can prevent fatal injuries. Nevertheless, in some accidents, such as those involving people falling from high heights, the improper wearing of helmets can cause secondary injuries, which eventually lead to tragedy. According to the national standard "head protective safety helmet" (GB2811-2019), a safety helmet should be adjusted according to the size of the head circumference cap or chin belt to ensure that it is firmly worn, not accidentally offset or slipped. Even if an accident occurs, even the most straightforward brain injury may require physical and psychological treatment to treat memory problems, behavioral changes, depression, and personality changes. Therefore, workers must wear correctly safety helmets to address potential dangers. However, due to low safety awareness, construction workers do not comply with national standards. Traditional manual management is inefficient, consumes resources, and hinders effective accident prevention accidents. Therefore, automatic detection of helmet-wearing situations is critical.

Target recognition based on deep learning has recently been a research hotspot in computer vision. Unlike traditional methods that require manual design and feature extraction, deep learning can improve model accuracy by automatically learning features [1], [2]. Deep learning detection algorithms are divided into two-stage detection algorithms based on candidate regions and end-to-end single-stage detection algorithms [3]. The two-stage detection algorithm is represented by R-CNN [4], a deep learning algorithm for target detection proposed by R.Girshick et al., Fast R-CNN [5], and Faster R-CNN [6] algorithms with higher performance. Two-stage detection algorithms have the characteristics of high detection

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

accuracy and long detection time, so it is unsuitable for real-time detection. The single-stage object detection algorithm represented by SSD [7] and YOLO [8], [9], [10], [11] have fast detection speeds and high accuracy. Compared with other algorithms, the YOLO series has the advantages of a more straightforward network structure, more vital generalization ability, and better performance. Therefore, many scholars propose applying the YOLO algorithm to real-time construction scene detection.

Fang et al. proposed a helmet detection algorithm based on YOLOv2; they added a dense network to the feature extraction network and utilized a lightweight MobileNet network structure to reduce the model complexity and improve the detection speed [12]. Wu et al. employed a DenseNet network instead of the Darknet53 feature extraction network to YOLOv3 to improve helmet detection accuracy [13]. Shi et al. added a feature pyramid in YOLOv3 to improve the recognition accuracy of people and helmets [14]. Yang et al. investigated helmet-wearing detection based on YOLOv3 and used a support vector machine (SVM) to classify the detection results [15]. Wang et al. based their helmet detection algorithm on YOLOv5s, introduced the CA (coordinate attention) attention mechanism in the backbone network structure and utilized a weighted bidirectional feature pyramid (BiFPN) network structure to improve the model detection accuracy [16]. Alateeq et al. proposed a personal protective equipment (PPE) and heavy equipment detection model based on the YOLOv5s algorithm and incorporated weather conditions into the model. It is possible to analyze whether the area around the equipment is dangerous based on the prevailing weather [17]. Lo et al. constructed a new PPE dataset, trained three PPE detection models using the YOLOv3, YOLOv4 and YOLOv7 algorithms and summarized the advantages and disadvantages of each algorithm [18]. Zhu et al. proposed a detection model for electric power based on the YOLOv5s algorithm by using a self-constructed dataset to detect power staff protection equipment [19]. Fu et al. used K-means to recluster based on YOLOv5s and added a detection layer to improve the detection accuracy [20]. Zhao et al. used the DenseBlock module instead of the Focus structure in the YOLOv5 main network and added the SE-Ne attention module to improve the detection performance [21]. Du et al. employed the Swin Transformer as a feature extractor for the YOLOv5s network and introduced a dense spatial pyramid pooling module to improve model detection [22]. Chen et al. propose a YOLOv5n-based for helmet and reflective undershirt detection algorithm; in their method, they used the efficient intersection over union loss function (EIOU-loss), a mixed convolutional block attention module (CBAM) and a CA attention mechanism in the network structure, and subsequently added a detection layer to improve the model performance [23].

The above research is very important, but there are still the following problems: (1) some algorithms have high detection accuracy, but the number of parameters and calculation amount still greatly burden the computing equipment. (2) Some detection models have low computational effort but also low detection accuracy. (3) Helmet detection has been widely investigated attention to, and some studies have been conducted on reflective undershirts. However, no scholars have explored whether helmets are worn correctly. (4) all of the above studies optimize and improve the detection performance of algorithms without considering the needs of practical applications. Notably, the YOLO series algorithms have been updated to the eighth version (YOLOv8). Nevertheless, in recent years, the vast majority of scholars have based their research on the YOLOv5 algorithm, because YOLOv5 has a simpler network structure and has premodels with different network depths so that the scholars can choose a premodel that is more suitable for their research. More importantly, YOLOv5 has lower model size and higher FPS.

Therefore, this paper proposes a standard helmet wearing detection model based on improved YOLOv5. To address the problem of the high computational effort of the above algorithms, YOLOv5n, which is the least computationally intensive, is used as the pre-model for training. In response to the low accuracy of model detection, considering the characteristics of small detection targets, high overlap rate, and easy occlusion at construction sites, the algorithm uses the EIOU-loss [24] loss function to replace the CIOU-loss loss function to improve the model performance. Aimed at the original YOLOv5n detection of dense targets with high leakage rates, Soft-NMS [25] is employed instead of NMS to improve the recognition of occluded targets. By adding the CBAM [26] attention module to improve the attention given to target features, the problems of small size and easy confusion with the lower chin strap in the helmet can be solved. A small target detection layer (ADL) is added to improve the detection performance for small targets over a long range. The contributions of this study are summarized as follows:

1. The first standard wearing helmet image dataset was established and included 4400 images in different environments such as dense targets, long-distance targets, dense long-distance targets, and insufficient illumination.

2. For the first time, we propose the theory of whether helmet wearing is standard for target detection research, and apply it to standard helmet wearing detection at construction sites based on the YOLOv5 algorithm to fill research gap on standard helmet wearing detection.

3. The experiment showed that ADL reduces the model's accuracy. Nevertheless, by cooperating with the CBAM, the model can meet the real-time detection accuracy requirements of construction sites and significantly reduce the missed detection rate.

4. From the perspective of improving the practicability of detection results, this study developed an automatic preservation function for video stream detection results, which can be utilized as an important basis and support for

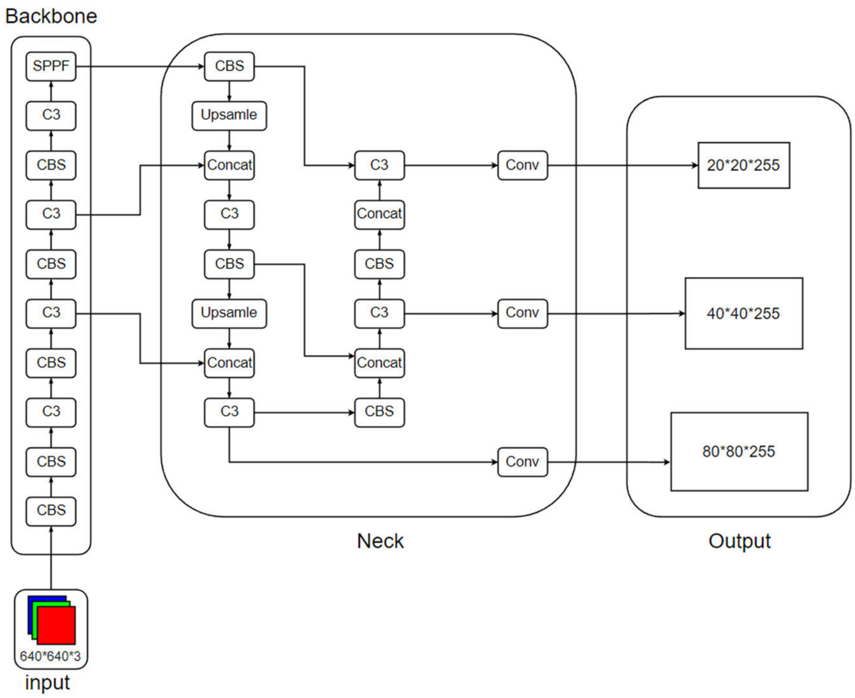


FIGURE 1. YOLOv5n network structure.

decision-making and the implementation of construction site safety management.

II. METHOD

1) YOLOV5

YOLOv5 is one of the most advanced single-stage target detection algorithms; it was released on June 10, 2020, and is still being updated. There are currently eight versions. This paper selects the latest version, 6.1. YOLOv5 officially provides five versions of the network model, according to the network depth from low order to high order for YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. This paper uses the YOLOv5n model. The YOLOv5n network model has the smallest volume and the fastest detection speed, and the detection accuracy can also meet the actual needs. The can be deployed on low-performance UAVs and has extremely high versatility.

The YOLOv5n network model is divided into four parts: the input, backbone, neck, and output. The network structure of YOLOv5n is shown in Fig. 1

2) EIOU-LOSS

The IOU (intersection-over-union) loss represents the difference between the predicted values and the true values of the target position and can be used to correct the position coordinates of the prediction box. However, when the initial IOU prediction box and the real box do not intersect, the difference does not reflect the distance between the two boxes or the size of the overlap. Therefore, we use the EIOU-loss to improve the accuracy of the prediction box.

YOLOv5 uses CIOU-loss (Complete-IOU) as the loss function of the bounding box. The specific formulas are (1), (2), (3), and (4).

$$L_{CIOU} = 1 - IOU + \frac{p^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{1}$$

$$\alpha = \frac{v}{(1 - IOU) + v} \tag{2}$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \tag{3}$$

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

where A represents the prediction box, B represents the target box, and $\rho(\cdot)$ represents the Euclidean distance between the two centroids of the predictor frame and the target frame, c is the length of the diagonal that minimally encloses the two bounding boxes, and b and b^{gt} denote the centroids of A and B, respectively. α is the weight function; v is used to measure the similarity of the aspect ratio between the anchor frame and the target frame; w and w^{gt} are the widths of A and B, respectively, and h and h^{gt} are the heights of A and B, respectively.

The CIOU-loss does not calculate the true difference between width or height and their confidence, which sometimes hinders the convergence of the model. In response to this problem, the EIOU-loss is used instead of the CIOU-loss. The specific formula presented is as follows.

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} \tag{5}$$

$$L_{IOU} = 1 - IOU + \frac{\rho^2\left(b, b^{gt}\right)}{(w^c)^2 + (h^c)^2} \tag{6}$$

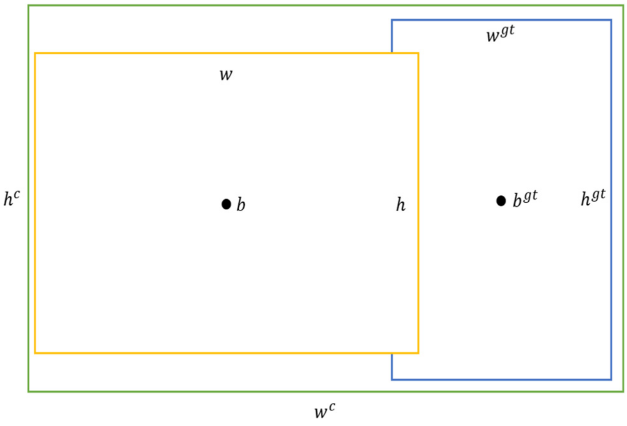


FIGURE 2. Schematic of the mechanism of EIOU-loss.

$$L_{dis} = \frac{\rho^2\left(w, w^{gt}\right)}{(w^c)^2} \tag{7}$$

$$L_{asp} = \frac{\rho^2\left(h, h^{gt}\right)}{(h^c)^2} \tag{8}$$

In the illustration in Fig. 2, yellow is the prediction box, blue is the target box, green is the minimum closed box, and w^c and h^c are the width and height, respectively, of the smallest enclosing box that covers both boxes. The EIOU-loss is divided into three parts: the IOU loss L_{IOU}, the distance loss L_{dis}, and the aspect loss L_{asp}. In this way, the difference between the width and height of the target frame and the anchor frame can be reduced while retaining the advantages of CIOU, thereby obtaining faster convergence speed and better positioning results.

3) SOFT-NMS

Soft-NMS was used to replace NMS to increase the accuracy and recall of obscured target detection. The original NMS determines whether to remove the detection frame when removing the redundant detection frame based on the IOU's value. The detection frame is removed when the IOU exceeds the set threshold value. When the targets are dense, mutual occlusion leads to an enormous IOU value, and NMS incorrectly removes the detection frame, causing the target to be missed. In a helmet-wearing environment, there is often overlap of occlusions, so Soft-NMS is used to improve missed detection.

The standard suppression of NMS and the IOU exceeds the threshold of the detection frame score, which is directly set to 0, as shown in (9). Moreover, Soft-NMS advocates the penalty decay of its score. There are two types of penalties. The first penalty function is as shown in (10), but the above equation is not continuous; this leads to an abrupt change in the detection sequence. The continuous penalty function has no penalty when there is no overlap and a very high penalty when there is a high overlap. Moreover, the number of sentences should gradually increase when the overlap is low. Thus, the second Gaussian penalty function is proposed as shown in (11) so that Soft-NMS can avoid setting the threshold size.

$$s_i = \begin{cases} s_i, & IOU\left(M, b_i\right) < N_t \\ 0, & IOU\left(M, b_i\right) \geq N_t \end{cases} \tag{9}$$

$$s_i = \begin{cases} s_i, & IOU\left(M, b_i\right) < N_t \\ s_i\left(1 - IOU\left(M, x_i\right)\right), & IOU\left(M, b_i\right) \geq N_t \end{cases} \tag{10}$$

$$s_i = s_i e^{-\frac{IOU(M,x_i)^2}{\sigma}}, \quad \forall b_i \notin D \tag{11}$$

where s_i denotes the classification score, M indicates the prediction box with the highest prediction score, x_i is used to determine whether the prediction box needs to be removed, and N_t denotes the threshold value of NMS.

Soft-NMS is a greedy algorithm that does not find a globally optimal rescoring detection frame. Soft-NMS is a generalized nonmaximal suppression, and conventional NMS is a particular case of Soft-NMS.

4) CBAM

Because the chin strap target of a helmet is small and the number of pixels is low, it is easy to confuse or miss. This paper adds the CBAM before the SPPF module to the Backbone section.

The CBAM consists of two submodules: the CAM and SAM. As shown in Fig. 3(a), a feature map is input, and the attention feature map is reasoned along two dimensions: channel and space. Then, the two feature maps are multiplied for adaptive operation, and the refined feature map is outputted. The structure of the CAM is shown in Fig. 3(b). The input feature map F is subjected to global maximum pooling and global average pooling in spatial dimensions. The two feature maps that are obtained are subsequently fed into a two-layered shared selective linking layer. The two features are summed to obtain the channel attention feature M_c after the sigmoid activation function. The structure of the SAM is shown in Fig. 3(c). M_c and the input feature map F are elementwise multiplied to obtain the input feature F' of the SAM and F' is pooled with the maximum and average in the channel dimension and convolved with a convolution operation to reduce its dimensionality. The spatial attention feature M_s is generated by the sigmoid activation function. The CBAM is a lightweight module that needs to be added only to the needed parts when used, without additional training, and the impact on the detection time is negligible. The structure of the CBAM is shown in Fig. 3. The improved network model is shown in Fig. 4.

5) SMALL TARGET DETECTION LAYER

The original YOLOv5n model has only three detection layers, as shown in Fig. 1; these layers are used to detect large, medium, and small targets, and the sizes of the corresponding detection layer feature maps are 20*20, 40*40, and 80*80, respectively. [27]. Due to the small size of the chin strap of the helmet, it is easy to occlude the strap, and the construction site staff are all over the site. Moreover, the helmets in the

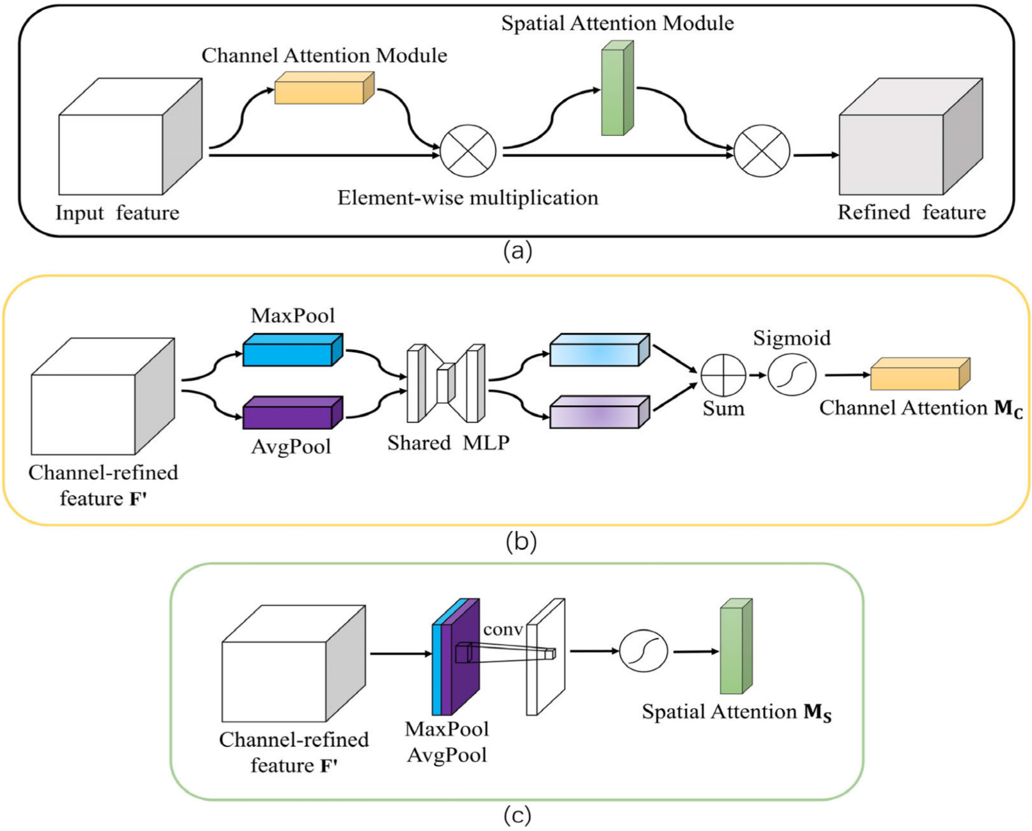

Spatial Attention Module
Channel Attention Module
Input feature
Element-wise multiplication
Refined feature
(a)
MaxPool
Sigmoid
AvgPool
Shared MLP
Sum
Channel Attention $\mathbf{M_C}$
Channel-refined feature F'
(b)
conv
Spatial Attention $\mathbf{M_S}$
Channel-refined feature F'
MaxPool
AvgPool
(c)

FIGURE 3. CBAM: (a) Structure of the CBAM, (b) structure of the channel attention module, and (c) structure of the spatial attention module.

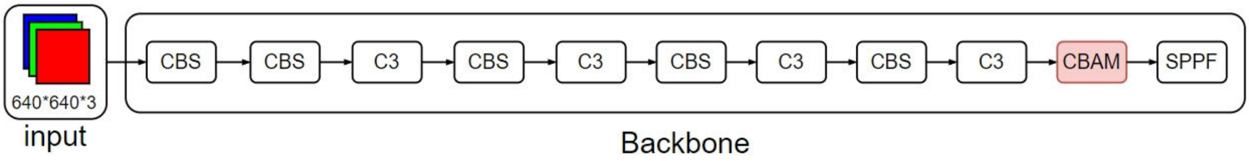


FIGURE 4. Backbone network structure with CBAM.

detection image also differ in size, especially the images taken by the drone. Therefore, based on the YOLOv5n network, another target detection layer with a feature image size of 160*160 is added to improve the accuracy under the above complex conditions This layer has a smaller receptive domain and richer position information. The featured image can better utilize the multilevel feature information of dense objects, thus improving the detection performance of the model in long-range scenes. The improved YOLOv5n network structure is shown in Fig. 5.

6) AUTOMATIC STORAGE OF VIDEO STREAM DETECTION RESULTS

Construction sites are generally equipped with video monitoring systems (e.g., CCTV). Nevertheless, images in the monitoring room are broadcast on the same screen in multiple venues, and the use of manual labor is not only time-consuming and laborious but also inefficient. Therefore, with the use of an existing video monitoring device, the model proposed in this paper is used to perform real-time detection of safety helmets worn at construction sites, and the detection results are automatically extracted and saved to terminals, which plays an important role in improving the pertinence of safety measures at construction sites. Safety managers punish construction personnel for on-site violations, and test results containing environmental information about the work site are needed as the basis. The original YOLOv5 can only save the clipping map of the target frame and lose important information, such as the working site environment, which cannot serve as a basis. For example, if a worker puts his or her helmet in his or her hand for a brief adjustment due to a problem such as a loose hatband, this behavior is fine, but the

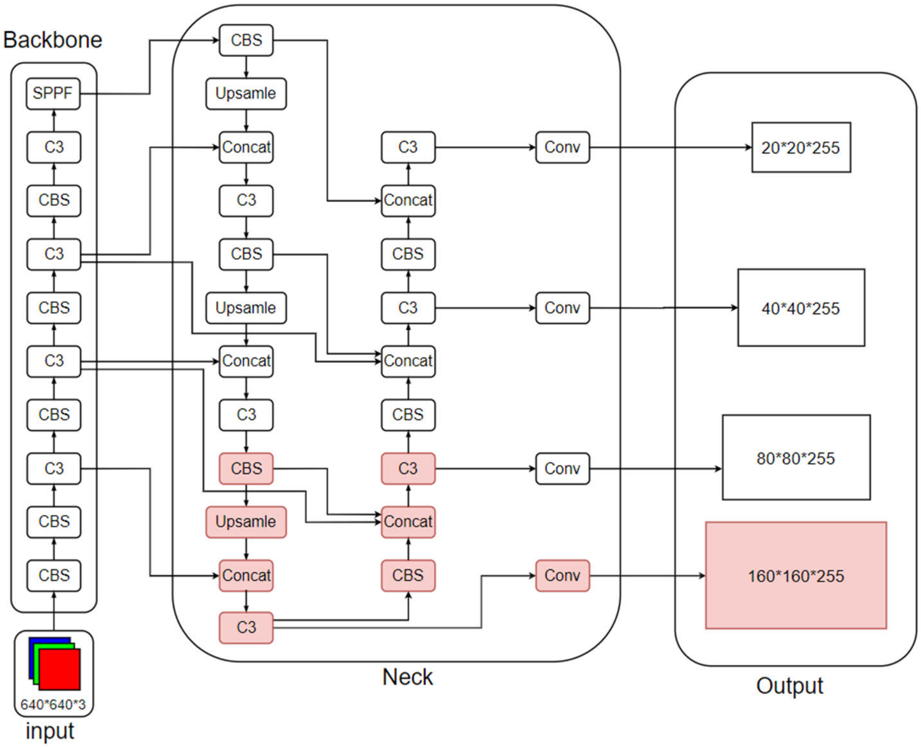


FIGURE 5. YOLOv5n network structure with ADL.

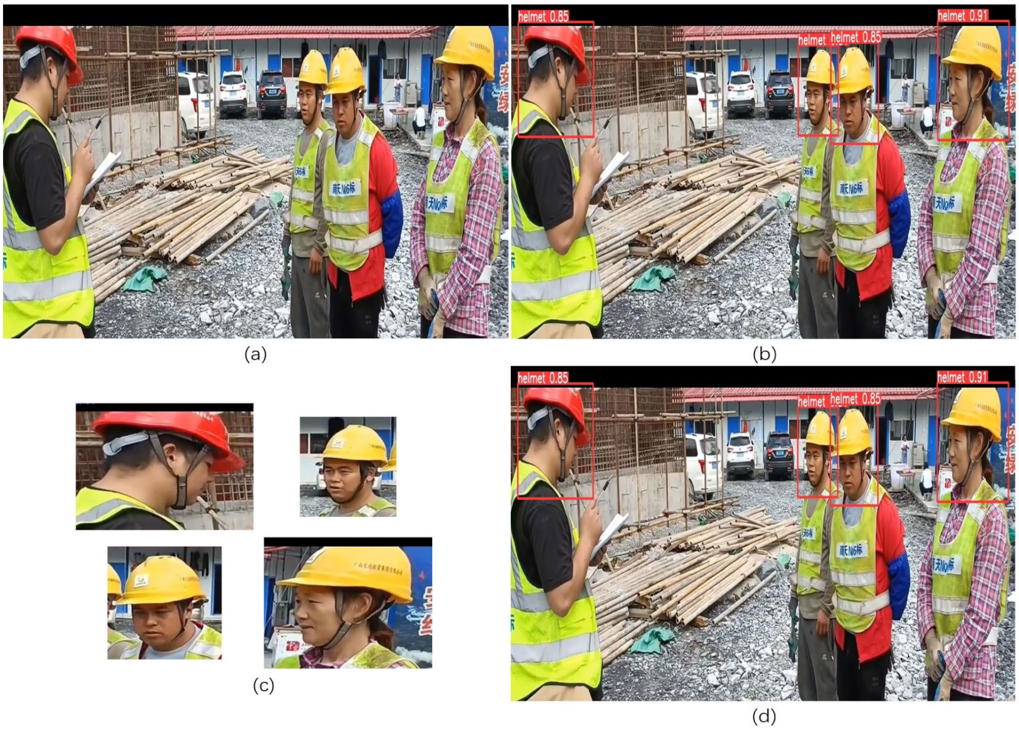


FIGURE 6. Automatic storage function renderings: (a) Video-captured image, (b) detection result, (c) original YOLOv5n capture function, and (d) YOLO-ESCA automatic savings function.

model will only save a screenshot of the worker's head and will not be able to provide a reason for why the worker did not wear the helmet. It would be unreasonable if the worker were penalized for this behavior. YOLOv5 has the function

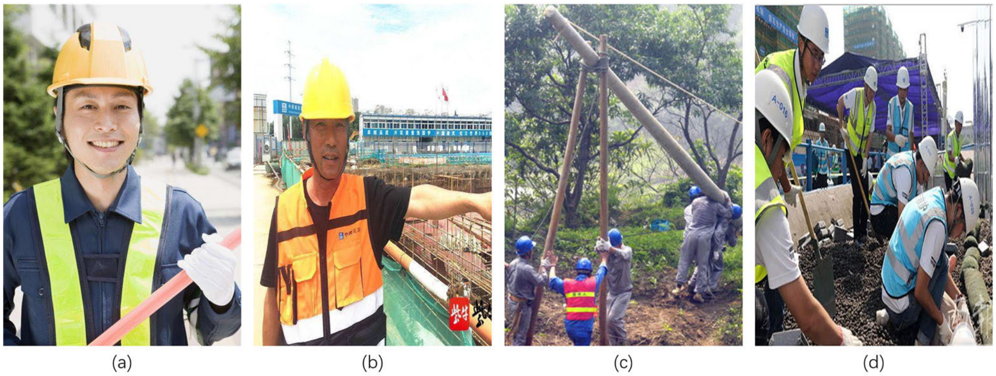


FIGURE 7. Distribution of some dataset categories: (a) Large target sample, (b) medium target sample, (c) small target, and (d) intensive target sample.

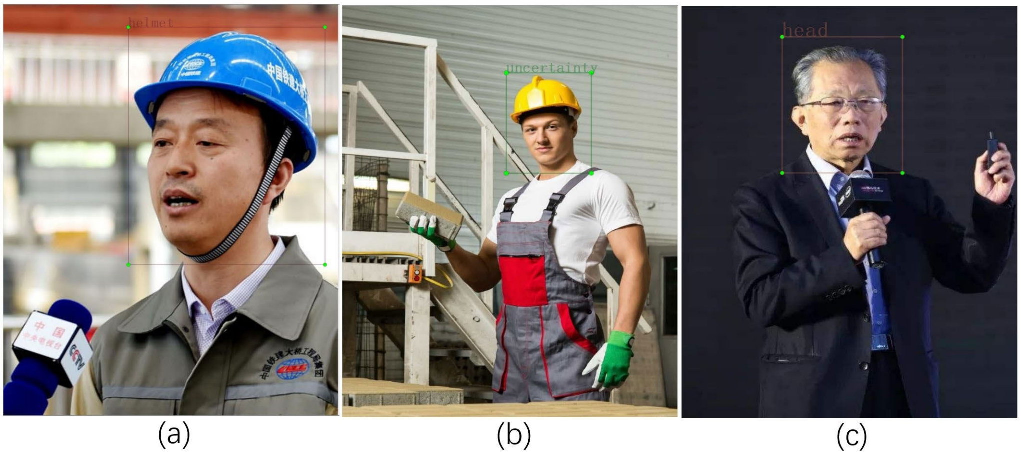


FIGURE 8. Image annotation: (a) Standard wearing of helmets, (b) nonstandard wearing of helmets, and (c) failure to wear a helmet.

to save the video, but manually checking the surveillance video is not only time-consuming and laborious but also has the possibility of errors, even if the surveillance video contains the detection results. Therefore, this study adds a video stream detection results preservation function, when it detects a target, it will only save the image of the current frame to a terminal such as a monitoring device. These images will only be made available to security managers, so there will be no legal implications. Fig. 6 shows the 1080P definition of the construction site safety education video inspection effect.

III. EXPERIMENT AND ANALYSIS

1) DATASETS

Since there is no research on the standard wearing behavior of helmets at home or abroad, there is a lack of open-source datasets. Therefore, in this paper, according to the national standard "Head Protection: Helmets" (GB2811-2019), the wearing style of wearing with the hat band fastened is recorded as the standard wearing of helmets, the wearing style of wearing without the hat band fastened is recorded as the nonstandard wearing of helmets, and then images are collected. The dataset consists of 4400 pictures, which will not involve privacy and interest issues, and the details are shown in Table 1. The dataset contains images of large, small, and small and dense targets; some data visualization is shown in Fig. 7. The labelimg tool was subsequently used to label the pictures. The label format was YOLO, which was divided into three categories—category 0, 1, and 2—corresponding to standard wearing a helmet (helmet), not wearing a helmet (head), not standard wearing a helmet (uncertainty). The process of labeling is shown in Fig. 8.

TABLE 1. Dataset details.

Datasets	Quantity
Train	3000
Val	1000
Test	400

2) EXPERIMENTAL ENVIRONMENT AND MODEL TRAINING

The experimental equipment used was a Shinelong M7-E6S3 notebook computer. Parameter selection was based on

TABLE 2. Experimental conditions.

Experimental Environment	Details
GPU	NVIDIA GeForce RTX2070 8G
CPU	AMD Ryzen 5 3600
Memory	Samsung 32G 3200 MHz
Operating system	Windows 10
Programming language	Python3.8.0
Deep learning framework	Pytorch1.10.1
Acceleration environment	CUDA11.3 + cudnn8.2.1

TABLE 3. Training parameters.

Parameters	Details
Image-size	640*640
Epochs	200
Batch-size	24
Warmup	10
Initial learning rate	0.01
Optimization algorithm	SGD
Premodel	YOLOv5n

previous studies, and the parameters for data preprocessing were selected from the default data in the hyp.scratch-low.yaml file. The specific configuration is shown in Table 2, and the training parameters are shown in Table 3.

3) EVALUATION CRITERIA

Precision is the assessment of the accuracy of the forecast.

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

Recall is an evaluation of the completeness of the search.

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

The single-category accuracy (AP) refers to the average of all accuracies obtained under the possible values of all the recall rates. mAP@0.5 is the average accuracy of all categories when the IOU threshold is 0.5, where m is the total number of categories.

$$AP = \int_0^1 P(r)dr \tag{14}$$

$$mAP = \frac{1}{m}\sum\nolimits_{i=1}^{m} AP_i \tag{15}$$

4) IMPROVED YOLOV5

To verify whether the above four performance improvements can enhance the model performance, we conduct a separate improvement comparison experiment before the ablation experiment. The data before and after the model is improved in a separate place are shown in Table 4, and the default parameters are utilized for training. The improved network structure is shown in Fig. 9, and the improved detection effect is shown in Fig. 10-13.

In Table 4, we show the single class performance and average performance of each model. In terms of average performance, after adopting the EIOU loss function, the models' precision, recall, mAP@0.5 and mAP@0.5:0.95 of the model were increased by 1.4%, 0.2%, 0.9% and 1.5%, respectively; the FPS improved by 0.7; the model size decreased by 0.02 MB; and the size of the target box was closer to the real situation than before the improvement. When the CBAM attention module is added in front of the SPPF module, image feature extraction is enhanced, the target false detection rate is reduced, and the model precision, recall, mAP@0.5, and mAP@0.5:0.95 are increased by 1.2%, 0.3%, 0.7%, and 0.5%, respectively. When the FPS decreased by 0.7, the model size increased by 0.11 MB. After ADL, the precision, recall, mAP@0.5 and mAP@0.5:0.95 of the model decreased by 10.5%, 4.7%, 3.8% and 7.2%, respectively. When the FPS decreases by 8.2, the model size increases by 0.7 MB. Since only part of the images in the dataset contains small targets, the addition of a small target detection layer will reduce the performance of the model. However, as shown in Figure 9, ADL has an acceptable impact on detection accuracy and can effectively detect small targets, significantly improving missed detections. After using Soft-NMS, the model precision, recall, mAP@0.5 and mAP@0.5:0.95 of the model increased by 1.8%, 0.3%, 1.2% and 4%, respectively. When the FPS increases by 1 and the model size decreases by 0.02 MB, the leakage rate can be effectively reduced when the target is dense. However, as shown in Fig. 10, neither algorithm can detect the leftmost helmet, and we believe that the photographer did not focus on the leftmost helmet at the time of the shot and that this portion of the image was somewhat blurred and that some features were still obscured, causing the algorithm to miss the detection.

In terms of single-class performance, the precision performance of the uncertainty category is the lowest for all models because the detection targets of uncertainty and helmets are too similar, and the model considers some of the targets uncertain at the pretraining stage. However, at the late stage of the training stage, as the performance improves, these targets are again considered by the model as helmets, which reduces the uncertainty detection accuracy, and the detection performance of other categories improves after improvement. Adding a small target detection layer and CBAM decreases the FPS of the model and increases the model size, where the effect of CBAM is negligible. However, these two improvements effectively improve the detection performance, so they are necessary.

5) ABLATION EXPERIMENTS

In this study, an ablation experiment was performed to verify the effect of mixing improvements on the model performance. To ensure the effectiveness of the experiment, 11 ablation experiments were created by arranging and combining four improvements, of which seven groups included ADL. The experimental data are shown in Table 5

In Table 5, we divided all the models into two categories. The first category is the model without adding the small target detection layer, and the second category is the model with adding the small target detection layer. In these two categories, the two models with the best performance, named

TABLE 4. Comparison of the improvement in performance for each part of the model.

Model	classes	Precision	Recall	mAP@0.5	mAP@0.5:0.95	FPS	Model size
YOLOv5n	all	96	98	98.1	79.9	80.4	3.67MB
	helmet	96.8	98.2	98.2	81.2		
	head	97.1	98.1	98.3	80		
	uncertainty	94.1	97.8	97.9	78.2		
EIOU	all	97.4	98.2	99	81.4	81.1	3.65MB
	helmet	98.2	98.8	99.4	83.3		
	head	97.9	98.8	99	81.6		
	uncertainty	96.1	97	98.7	79.1		
CBMA	all	97.2	98.3	98.9	80.4	79.1	3.78MB
	helmet	98.3	98.9	99.3	83.2		
	head	97.5	98.9	99.1	80.9		
	uncertainty	95.8	97.3	98.3	76.9		
ADL	all	85.5	93.3	94.3	72.7	72.2	4.37MB
	helmet	90.7	98.9	96.9	76.6		
	head	84.7	93.4	95.3	73.5		
	uncertainty	81.3	87.5	91.6	68.1		
Soft-NMS	all	97.8	98.3	99.3	83.1	81.4	3.65MB
	helmet	98.4	98.4	99.4	84.7		
	head	98.4	98.8	99.3	83.5		
	uncertainty	96.7	97.7	99.1	81		

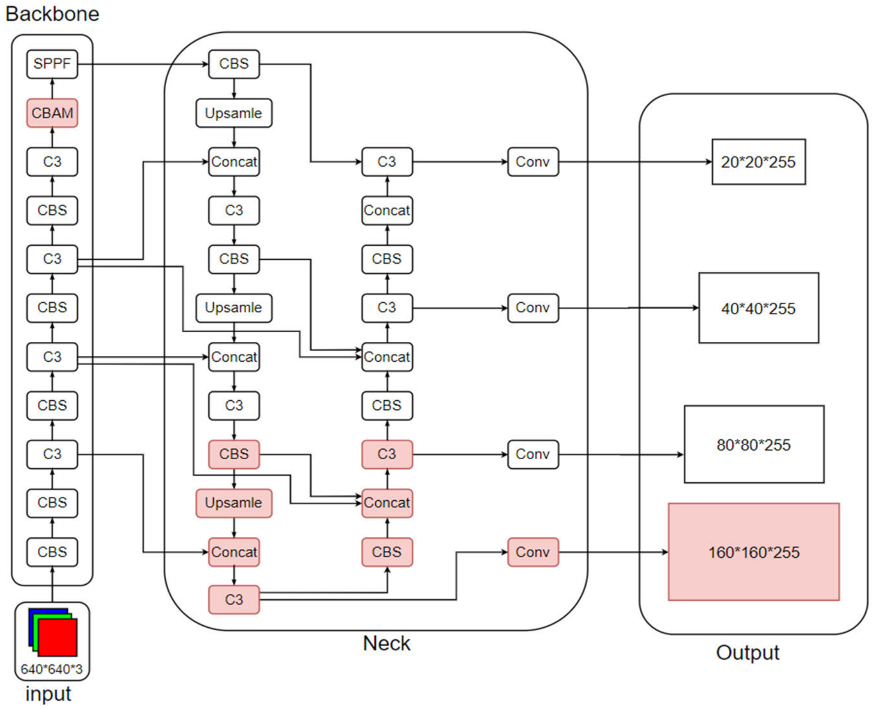


FIGURE 9. Improved YOLOv5n network structure.

YOLO-ESC and YOLO-ESCA after the acronym of the improved method, are selected as representatives. Based on the original YOLOv5n model, method 1-YOLO-ESC shows that at any time, the improvements combined in various ways, with the exception of the addition of a small target detection layer, will improve the performance of the model to varying degrees. The precision and recall increase by 0.13% and 0.1%, respectively, on average. mAP@0.5 has an average increase of 0.45%, mAP@0.5:0.95 has an average increase of 0.2%, FPS has an average increase of 1.9, and the model size has an average increase of 0.04 MB. Moreover, the YOLO-ESC model, which simultaneously achieves three improvements, achieves the best performance, which indicates that these three improvements not only have no negative effects but also may complement each other. Thus, better results are obtained. As demonstrated by method 4-YOLO-ESCA, when a small target detection layer is added without a CBAM attention module, the model precision decreases by 11.3% on average, the recall decreases by 6.5% on average, the mAP@0.5 decreases by 3.9% on average,

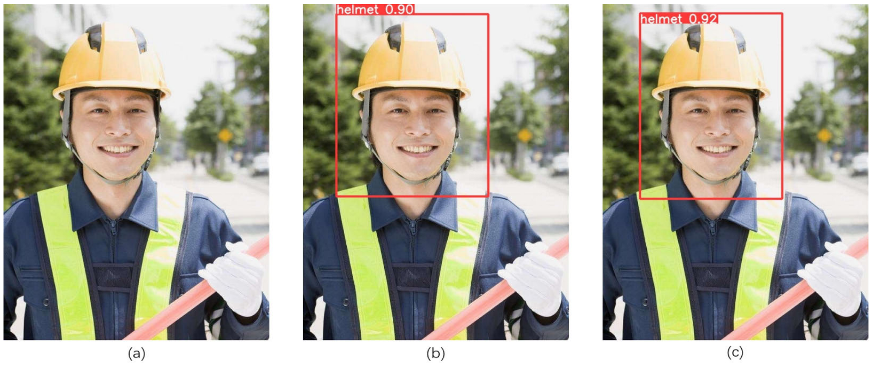


FIGURE 10. Comparison of detection before and after EIOU-loss improvement: (a) Image, (b) YOLOv5n detection results, and (c) EIOU-loss detection results.

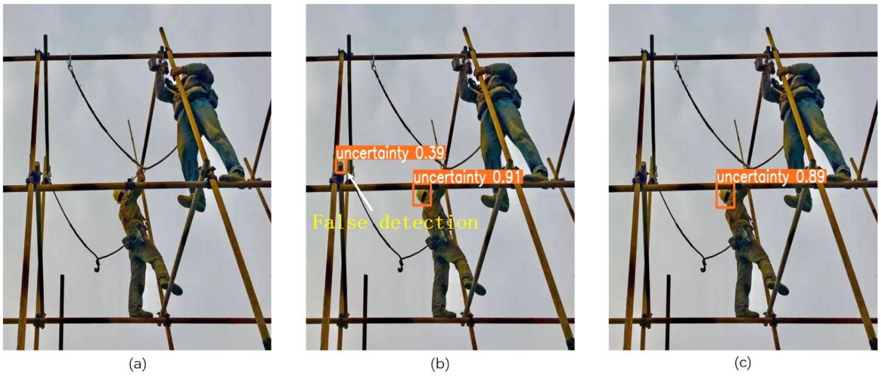


FIGURE 11. Comparison of detection before and after improvement of the CBAM: (a) Image, (b) YOLOv5n detection results, and (c) CBAM detection results.

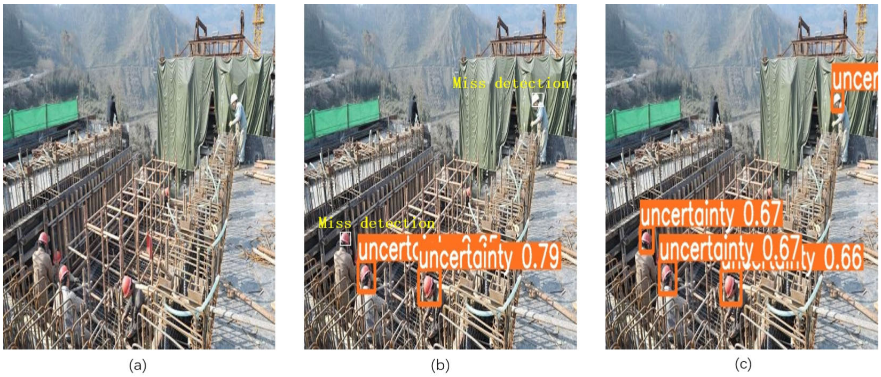
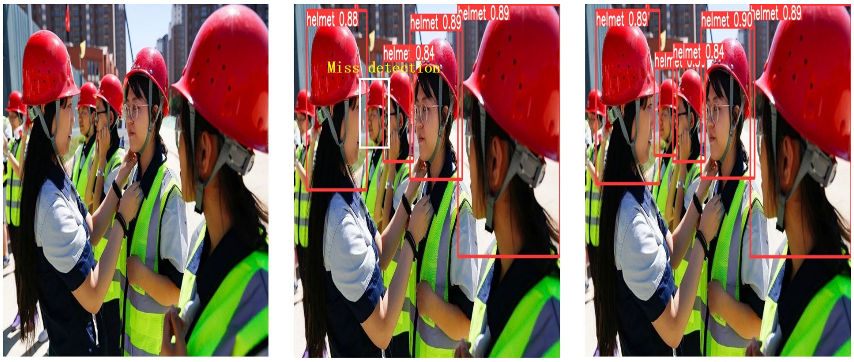


FIGURE 12. Comparison of detection before and after ADL improvement: (a) Image, (b) YOLOv5n detection results, and (c) ADL detection results.

and the mAP@0.5:0.95 decreases by 6.7% on average. FPS decreased by 12.8 on average, and the model size increased by 0.7 MB on average. After the two models are combined, the precision decreases by 10.53% on average, the recall decreases by 4.78% on average, the mAP@0.5 decreases by 3.83% on average, the mAP@0.5:0.95 decreases by 6.58% on average, the FPS decreases by 14.2 on average, and the model size increases by 0.74 MB on average. Compared with other improvements, the CBAM attention module significantly reduces the impact of adding a small target detection layer but also increases the complexity of the model. This result is consistent with the conclusion drawn in the previous section. In summary, both the EIOU-loss and Soft-NMS improve the prediction stage of the model; therefore, improving neither the loss nor the Soft-NMS will reduce the impact of ADL. Although adding a small target detection layer will reduce the model performance, especially the impact on precision, the experiments in the previous chapter have shown its necessity.



(a) (b) (c)

FIGURE 13. Comparison of detection before and after Soft-NMS improvement: (a) Image, (b) YOLOv5n detection results, (c) Soft-NMS detection results.

TABLE 5. Results of ablation experiments.

Model	EIOU	Soft-NMS	CBAM	ADL	classes	Precision	Recall	mAP@0.5	mAP@0.5;0.95	FPS	Model size
YOLOv5n	×	×	×	×	all	96%	98%	98.1%	79.9%	80.4	3.67MB
					helmet	96.8%	98.2%	98.2%	81.2%		
					head	97.1%	98.1%	98.3%	80%		
					uncertainty	94.1%	97.8%	97.9%	78.2%		
1	×	√	√	×	all	95.9%	97.8%	98.8%	79.9%	78.7	3.76MB
					helmet	97.9%	99.3%	99.4%	82.8%		
					head	96.8%	98.1%	99.2%	80.6%		
					uncertainty	93.1%	95.9%	97.9%	76.2%		
2	√	√	×	×	all	96.2%	98.3%	98.1%	80.3%	80.5	3.65MB
					helmet	97.1%	99.1%	98.4%	82%		
					head	96.4%	98.8%	98.2%	80.9%		
					uncertainty	95.1%	97.2%	97.8%	78%		
3	√	×	√	×	all	95.9%	98.1%	98.6%	80.1%	77.8	3.76MB
					helmet	98.3%	98.9%	99.3%	83.2%		
					head	96.8%	98.9%	99%	81.9%		
					uncertainty	92.5%	96.6%	97.5%	76.3%		
YOLO-ESC	√	√	√	×	all	96.5%	98.2%	98.7%	80.1%	76.7	3.67MB
					helmet	98.1%	99.6%	99.1%	82.7%		
					head	97.5%	99.1%	99.1%	80.4%		
					uncertainty	93.9%	95.8%	97.8%	77.4%		
4	×	√	×	√	all	84.8%	92.5%	93.6%	73.3%	67.8	4.37MB
					helmet	89.7%	98.3%	96.2%	77.2%		
					head	85%	90.8%	93.3%	73.3%		
					uncertainty	79.7%	88.5%	91.2%	69.3%		
5	×	×	√	√	all	85.2%	92.8%	93.6%	72.8%	66.7	4.39MB
					helmet	88.6%	97.2%	96.9%	77.8%		
					head	85%	92%	94.4%	73.4%		
					uncertainty	81.9%	89.1%	89.5%	67.2%		
6	√	×	×	√	all	84.3%	92.5%	93.7%	72.5%	67.4	4.37MB
					helmet	89.1%	97.2%	96.7%	75.8%		
					head	84.3%	91.8%	95%	73.6%		
					uncertainty	79.6%	88.6%	89.5%	68%		
7	√	√	×	√	all	85.1%	92.5%	94.2%	73.2%	67.6	4.37MB
					helmet	89.9%	98.3%	97.1%	77.3%		
					head	85.5%	91.6%	94.3%	73.4%		
					uncertainty	79.8%	87.5%	91%	68.9%		
8	×	√	√	√	all	85.7%	94%	94.1%	73.7%	65.9	4.39MB
					helmet	89.8%	97.6%	96.5%	78.1%		
					head	85.6%	94.1%	84.3%	73.8%		
					uncertainty	81.9%	90.1%	91.3%	69.3%		
9	√	×	√	√	all	85.4%	92.3%	94.4%	72.6%	65.6	4.39MB
					helmet	89.8%	97.5%	97.2%	78.4%		
					head	86.3%	92.3%	94.8%	74.8%		
					uncertainty	80.1%	87.2%	91.1%	68%		
YOLO-ESCA	√	√	√	√	all	85.6%	93.8%	94.7%	74.2%	66.3	4.47MB
					helmet	89.1%	98.1%	97.8%	78.8%		
					head	84.8%	93.9%	95.8%	74.7%		
					uncertainty	79.8%	89.5%	90.6%	68.9%		

TABLE 6. Method comparison results.

Model	classes	Precision	Recall	mAP@0.5	mAP@0.5:0.95	FPS	Model size
YOLOv3	all	98.9%	98.7%	99.4%	89.3%	25.5	117MB
	helmet	99.1%	98.9%	99.4%	87.4%		
	head	99.6%	98.9%	99.5%	91.6%		
	uncertainty	98%	98%	99.3%	88.9%		
YOLOv3tiny	all	97.8%	98.3%	99.3%	80.3%	70.3	16.6MB
	helmet	98.9%	98.6%	99.3%	81.6%		
	head	98.9%	98.8%	99.5%	80.7%		
	uncertainty	95.5%	97.9%	99.1%	78.6%		
YOLOv5s	all	97.3%	98.2%	98.9%	84.1%	73.9	13.7MB
	helmet	97.6%	98.2%	98.9%	84.9%		
	head	98.2%	98.4%	99.1%	84.7%		
	uncertainty	96.1%	98%	98.7%	82.8%		
YOLOv5m	all	98.6%	99.1%	99.5%	88.4%	57.4	40.2MB
	helmet	99.5%	99.1%	99.5%	87.7%		
	head	99.1%	99.6%	99.5%	89.6%		
	uncertainty	97.2%	98.7%	99.4%	88%		
YOLOv5l	all	98.9%	99.2%	99.5%	90.6%	32	88.5MB
	helmet	99.3%	98.9%	99.5%	88.9%		
	head	99.2%	99.5%	99.5%	92.2%		
	uncertainty	98.2%	99.3%	99.4%	90.7%		
YOLOv5x	all	99.2%	99%	99.4%	90.1%	17.4	165MB
	helmet	99.5%	98.4%	99.5%	88.9%		
	head	1%	99.9%	99.5%	92.1%		
	uncertainty	98%	98.6%	99.3%	89.4%		
YOLO-ESC	all	96.5%	98.2%	98.7%	80.1%	75.7	3.67MB
	helmet	98.1%	99.6%	99.1%	82.7%		
	head	97.5%	99.1%	99.1%	80.4%		
	uncertainty	93.9%	95.8%	97.8%	77.4%		
YOLO-ESCA	all	85.6%	93.8%	94.7%	74.2%	65.3	4.47MB
	helmet	89.1%	98.1%	97.8%	78.8%		
	head	84.8%	93.9%	95.8%	74.7%		
	uncertainty	79.8%	89.5%	90.6%	68.9%		

6) COMPARISON OF METHODS

To evaluate, the proposed method is compared with existing mainstream lightweight target detection algorithms, and the results are shown in Table 6.

As shown in Table 6, although the precision, recall, mAP@0.5 and mAP@0.5:0.95 performances of the YOLOv3, YOLOv5m, YOLOv5l and YOLOv5x models are better than those of the YOLO-ESC and YOLO-ESCA models, their FPS and model size are unacceptable, and these models hinder deployment. The precision, recall, mAP@0.5 and mAP@0.5:0.95 performances of YOLOv3tiny and YOLOv5s are not much different from those of YOLO-ESC; however, YOLO-ESC has a higher FPS and a smaller model size, so YOLO-ESC is easier to deploy. Therefore, the final question is which of the two models, YOLO-ESC or YOLO-ESCA, is more suitable for practical applications?

7) STATISTICAL SIGNIFICANCE TEST

In this section, a statistical significance test (t test) is conducted to assess and ensure the generalizability of the proposed model. A t test is a statistical hypothesis test that determines whether there is a significant difference between the means of two groups or samples by calculating the t test and the p value of the two groups or samples. The t test is a measure of the difference in the mean values of two groups with respect to the within-group variance of each group and indicates how much the means of two groups differ from each other in terms of standard error. The p value is the probability that the t-statistic will reach an extreme value if the null hypothesis holds. The null hypothesis indicates that the means of the two groups are equal. A significance level of 0.05 was used as the cutoff for determining statistical significance. A p value less than 0.05 indicated that the original hypothesis was rejected, and a statistically significant difference existed between the means of the two groups.

In this work, we choose mAP@0.5 for comparison and apply the paired t test to evaluate the performance of YOLO-ESCA against other models. The smaller the t statistic is, the better the performance of the model. The P value was obtained by comparing YOLO-ESCA with different models. The p value of our model relative to the other models is less than 0.05, indicating a significant difference from the benchmark model, as shown in Table 7. Therefore, YOLO-ESCA is a considerable improvement over YOLO-ESC, YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x.

8) DETECTION EXPERIMENT

The comprehensive performance of the model cannot be singularly based on the level of the performance indicators.

TABLE 7. Statistical significance test results.

Model	T statistic	P value
YOLO-ESCA vs. YOLO-ESC	3.3915	0.0146
YOLO-ESCA vs. YOLOv5s	3.6030	0.0113
YOLO-ESCA vs. YOLOv5n	3.0926	0.0213
YOLO-ESCA vs. YOLOv5m	3.9867	0.0072
YOLO-ESCA vs. YOLOv5l	3.9867	0.0072
YOLO-ESCA vs. YOLOv5x	3.9523	0.0075

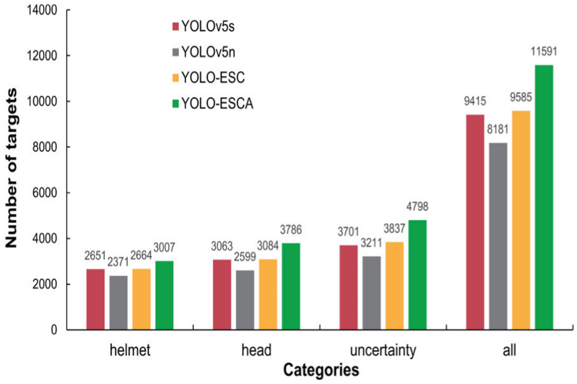


FIGURE 14. Self-made dataset detection results.

As a result, the effect of the actual application of the model is also highly important. To further validate the reasonableness of the improvement, 4400 images within the homemade dataset were selected as a detection dataset, and simultaneously, to validate the generalizability of the model, we also selected all the images from the open-source SHWD helmet-wearing detection dataset, which is employed by most scholars, as another detection dataset. YOLOv5s, YOLOv5n, YOLO-ESC and YOLO-ESCA, which have similar performances, were selected according to the above experiments for the detection comparison experiments, and the specific data are shown in Figs. 14 and 15.

From Fig. 14, we can see that on the homebrew dataset, YOLOv5s detects a total of 9415 targets, YOLOv5n detects a total of 8181 targets, and YOLO-ESC detects 9585, which is not much different from the number of detections of YOLOv5s and improves by 17.2% compared to YOLOv5n, whereas the total number of detected targets in YOLO-ESCA is as high as 11591, a 41.7% boost compared to YOLOv5n and a 20.9% boost compared to YOLO-ESC. In terms of categorization, YOLOv5s detected 3701 uncertainty targets, 3063 head targets, and 2651 helmet targets; YOLOv5n detected 3211 uncertainty targets, 2599 head targets, and 2371 helmet targets; YOLO-ESC detected 3837 uncertainty targets, 3084 head targets, and 2664 helmet targets, which is not much different from the number of detections of YOLOv5s and improved by 19.5%, 18.7%, and 12.4%, respectively, compared to YOLOv5n; and YOLO-ESCA detected 4798 uncertainty targets, 3786 head targets, and 3007 helmet targets, which are 49.4%, 45.7%, and 26.8% higher, respectively, than YOLOv5n and 25%, 22.8%, and 12.9% greater than YOLO-ESC.

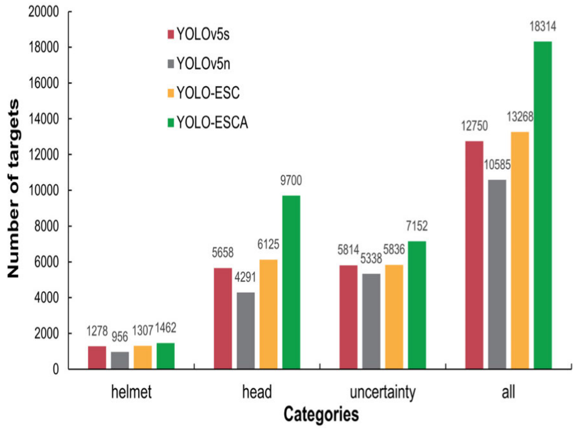


FIGURE 15. SHWD dataset detection results.

Fig. 15 shows that on the SHWD dataset, YOLOv5s detects a total of 12750 targets, YOLOv5n detects a total of 10585 targets, and YOLO-ESC detects 13268, which is not much different from the number of detections of YOLOv5s and improves by 25.3% compared with that of YOLOv5n; moreover, the detection of YOLO-ESCA's total number of targets reaches 18314, which is a 73% improvement compared to YOLOv5n and a 38% improvement compared to YOLO-ESC. In terms of categorization, YOLOv5s detected 5814 uncertainty targets, 3063 head targets, and 1278 helmet targets; YOLOv5n detected 5338 uncertainty targets, 4291 head targets, and 956 helmet targets; and YOLO-ESC detected 5836 uncertainty targets, 6125 head targets, and 1307 helmet targets, which is not much different from the number of detections of YOLOv5s and improves by 9.3%, 42.7%, and 36.7%, respectively, compared to YOLOv5n. The number of targets detected by YOLO-ESCA includes 7152 uncertainty targets, 9700 head targets, and 1462 helmet targets, which are 34%, 126.1%, and 52.9% improved, respectively, compared to YOLOv5n, and 22.5%, 58.4%, and 11.9% improved.

9) DISCUSSIONS

First, although YOLO-ESCA has good performance, it cannot be denied that our dataset is not large or representative enough, which will lead to the model failing to detect the target if it encounters a situation in which the training set does not have or contains fewer detection scenarios. Therefore, it is necessary to expand the dataset by further collecting images from various environments at the construction site. Second, to reduce the number of parameters of the model and facilitate its deployment, we selected YOLOv5n, which has the smallest volume, as the premodel for training. Although YOLO-ESCA based on YOLOv5n has excellent performance, as shown in Fig. 16, there are still cases of missed detections, which is an unavoidable side effect caused

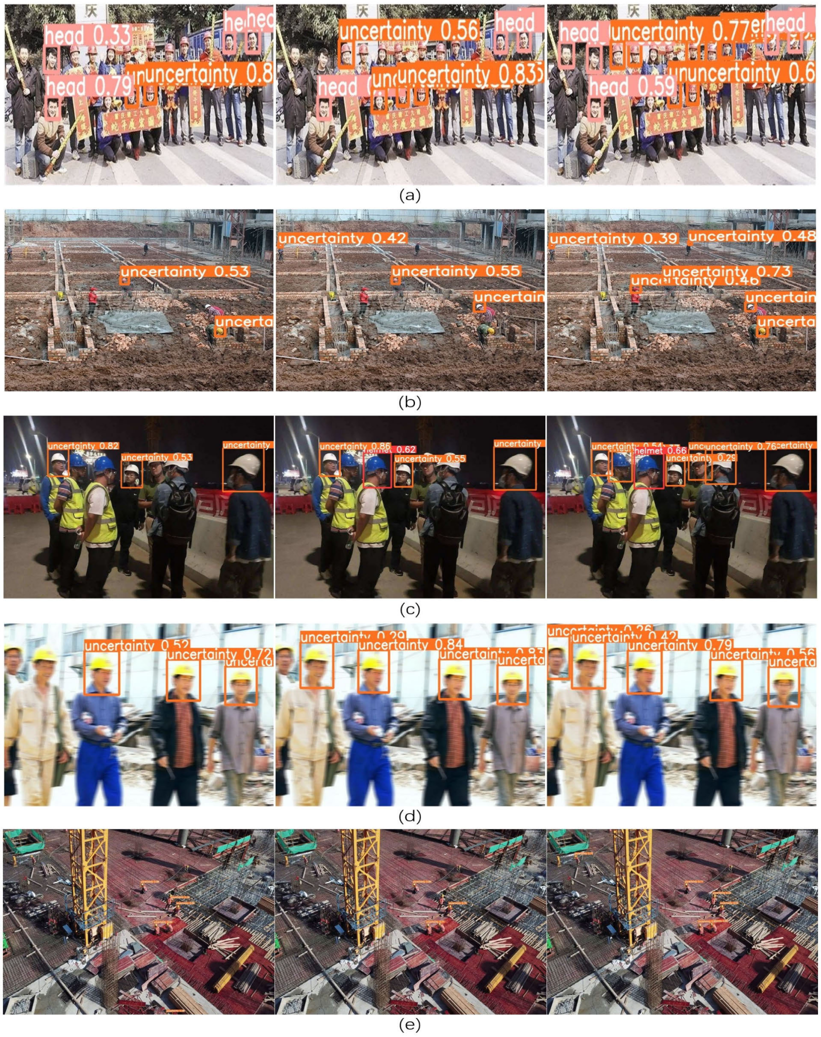


FIGURE 16. Visual comparison of several results. Columns from left to right are YOLOv5n, YOLO-ESC, and YOLO-ESCA: (a) Intensive target detection results, (b) long-range, small-target detection results, (c) dark environment intensive target detection results, (d) fuzzy target detection results, and (e) target detection results from a UAV perspective.

by a drastic reduction in volume, and the addition of a small target detection layer increases the volume of the model. Therefore, first, we will improve the model by decreasing the weight in the future and then improve the detection performance of the model without increasing the volume. Third, we have not applied the model to real construction work, and we do not know the actual performance of the model; however, our model has a very low model size and high FPS, and the hardware requirement is not high, which is highly suitable for deploying on UAVs with low computational power.

IV. CONCLUSION

In this paper, we propose a standard helmet wear detection model, YOLO-ESCA; the model can detect not only whether the worker is wearing a helmet but also whether the way he wears the helmet is standard. To improve the performance of the model, we first develop the automatic savings function of video streaming detection to improve the model utility. Second, we improve YOLOv5n by using EIOU-loss, a Soft-NMS nonlinear suppression module, a CBAM attention module and a small target detection layer. Although all the performance indices of the model decrease, the detection experiments prove that YOLO-ESCA is better than the original YOLOv5n and YOLO-ESC models without a small target detection layer in this application. Notably, our model also misses detection when detecting small targets at long distances. More importantly, our model size is only 4.47 MB with FPSs up to 65.3, which is conducive to deploying the model. Our ongoing work is focused on developing reliable target detection models. The goal of future work will be to continue to improve the model and other models for use in terminals, such as Raspberry Pi or NVIDIA Jetson Nano devices.

REFERENCES

[1] N. Sharma, R. Sharma, and N. Jindal, "Machine learning and deep learning applications—A vision," *Global Transitions Proc.*, vol. 2, no. 1, pp. 24–28, Jun. 2021, doi: 10.1016/j.gltp.2021.01.004.

[2] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in *Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697857.

[3] X. Bao and S. Wang, "Survey of object detection algorithm based on deep learning," *Transducer Microsyst. Technol.*, vol. 41, no. 4, pp. 5–9, 2022, doi: 10.13873/J.1000-9787(2022)04-0005-05.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[5] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C Berg, "SSD: Single shot multibox detector," in *Computer Vision—ECCV*, Amsterdam, The Netherlands. Berlin, Germany: Springer, 2016, pp. 21–37.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[9] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 7263–7271.

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

[11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOV4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[12] M. Fang, T. Sun, and Z. Shao, "Fast helmet-wearing-condition detection based on improved YOLOv2," *Opt. Precis. Eng.*, vol. 27, no. 5, pp. 1196–1205, 2019.

[13] F. Wu, G. Jin, M. Gao, Z. He, and Y. Yang, "Helmet detection based on improved YOLO v3 deep model," in *Proc. IEEE 16th Int. Conf. Netw., Sens. Control (ICNSC)*, May 2019, pp. 363–368.

[14] H. Shi, X. Chen, and Y. Yang, "Safety helmet wearing detection method of improved YOLOv3," *Comput. Eng. Appl.*, vol. 55, no. 11, pp. 213–220, 2019.

[15] L. Yang, L. Cai, and S. Gu, "Detection on wearing behavior of safety helmet based on machine learning method," *J. Saf. Sci. Technol.*, vol. 15, no. 10, pp. 152–157, 2019.

[16] L. Wang, J. Duan, and L. Xin, "YOLOv5 helmet wear detection method with introduction of attention mechanism," *Comput. Eng. Appl.*, vol. 58, no. 9, pp. 303–312, 2022.

[17] M. M. Alateeq, P. P. R. Fathimathul, and M. A. S. Ali, "Construction site hazards identification using deep learning and computer vision," *Sustainability*, vol. 15, no. 3, p. 2358, Jan. 2023.

[18] J.-H. Lo, L.-K. Lin, and C.-C. Hung, "Real-time personal protective equipment compliance detection based on deep learning algorithm," *Sustainability*, vol. 15, no. 1, p. 391, Dec. 2022.

[19] W. Zhu, Y. Shu, and S. Liu, "Power grid field violation recognition algorithm based on enhanced YOLOv5," *J. Phys., Conf.*, vol. 2209, no. 1, Feb. 2022, Art. no. 012033.

[20] D. Fu, L. Gao, T. Hu, S. Wang, and W. Liu, "Research on safety helmet detection algorithm of power workers based on improved YOLOv5," *J. Phys., Conf.*, vol. 2171, no. 1, Jan. 2022, Art. no. 012006.

[21] Z. Rui, L. Hui, L. Peilin, L. Yin, and L. Da, "Safety helmet detection algorithm based on improved YOLOv5s," *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 49, no. 8, pp. 2050–2061, 2021.

[22] X. Du, Y. Wang, R. Yan, D. Gu, X, Zhang, and T. Lei, "Accurate helmet wearing detection algorithm based on YOLO-ST," *J. Shaanxi Univ. Sci. Technol.*, vol. 40, no. 6, pp. 177–183 and 191, 2022.

[23] Z. Chen, F. Zhang, H. Liu, L. Wang, Q. Zhang, and L. Guo, "Real-time detection algorithm of helmet and reflective vest based on improved YOLOv5," *J. Real-Time Image Process.*, vol. 20, no. 1, p. 4, Feb. 2023.

[24] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022.

[25] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS—Improving object detection with one line of code," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5562–5570.

[26] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[27] K. Zhang, Y. Wu, J. Wang, Y. Wang, and Q. Wang, "Semantic context-aware network for multiscale object detection in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

PEIJIAN JIN is currently an Associate Professor; a Master Tutor and a Visiting Scholar of the University of Wollongong, Australia; an Expert in the expert pool of the Jilin Emergency Management Department; an Expert in the expert pool of the Changchun Emergency Response Bureau; the Head of the Emergency Response Technology and Management Program with Jilin Jianzhu University; and the Standing Director of the First Council of the Society of Safety Science and Engineering of Jilin Province. He is a national-level safety evaluator. He has been engaged in the professional construction and scientific research of emergency technology and management for a long time. He studies mainly power disaster monitoring and early warning systems, regional emergency capacity assessment, emergency technology, and management.

HANG LI received the B.S. degree in safety engineering from Jilin Jianzhu University, Jilin, China, in 2021, where he is currently pursuing the M.S. degree in safety engineering. His research interests include building construction safety, target detection, and convolutional neural networks.

WEILONG YAN is currently pursuing the master's degree in safety engineering with Jilin Jianzhu University, Changchun, Jilin, China. He joined the Safety Engineering Program, Jilin Jianzhu University, in 2022. His research interests include coal rock mechanics, graph neural networks, and complex networks.

JINRONG XU received the B.S. degree in mechatronics engineering from Shanghai Dianji University, Shanghai, China, in 2020. He is currently pursuing the M.S. degree in safety engineering with Jilin Jianzhu University. His research interests include lithium battery safety, target detection, and graph neural networks.