Research on an improved yolov5s algorithm for detecting helmets on construction sites

Qi Liu*
School of Mechanical Engineering and Automotive Engineering, Xiamen University of Technology, Fujian Xiamen, China, Corresponding author's e-mail: liu08082023@126.com

Fenggang Han
School of Mechanical Engineering and Automotive Engineering, Xiamen University of Technology, Fujian Xiamen, China, email: stariyhan@126.com

ABSTRACT

Aiming at the situation of irregular wearing of safety helmets on construction sites, a work on the detection of the wearing of safety helmets on construction sites is proposed to improve the yolov5s (You only look once) algorithm. To improve the problem that a large number of small targets are difficult to accurately identify due to dense construction site targets, combined with the structured weighted bidirectional pyramidal feature network (BiFPN) to improve the process of feature merging; in the Yolov5s feature extraction network in the introduction of the CBAM attention mechanism of the convolution module to obtain more feature details; using SIoU instead of CIoU as prediction box regression loss makes the network training and inference process faster and more accurate. The experimental findings show that the model with improved YOLOV5S helmet detection has higher detection precision compared to the YOLOV5S model, with MAP increased by 6.6 percentage points, recall increased by 8 percentage points and accuracy by 0.2 percentage points, making the improved algorithm more satisfactory for the helmet-wearing detection challenge.

CCS CONCEPTS

• **Computing methodologies** → Artificial intelligence, Computer vision; Computer vision problems; Object detection.

KEYWORDS

yolov5s, BiFPN, CBAM, SIoU loss function

ACM Reference Format:
Qi Liu* and Fenggang Han. 2023. Research on an improved yolov5s algorithm for detecting helmets on construction sites. In *2023 7th International Conference on Electronic Information Technology and Computer Engineering (EITCE 2023), October 20–22, 2023, Xiamen, China*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3650400.3650418

1 INTRODUCTION

The majority of helmet use detection on construction sites relies on manual video surveillance or manual inspection, but this method is time consuming, labour intensive and prone to missed inspections [1]. If workers are not wearing helmets on site, the consequences can be severe if an accident occurs. It is therefore important to use computer vision technology to detect the wearing status of helmets.

The existing research on the helmet status recognition using traditional methods of machine learning [2], in most of the recognition work is based on manually set features, but in practical applications due to the light source or angle problems make the difference from the object and the surrounding is not very obvious, the function of feature extraction is extremely cumbersome. With the evolution of advanced deep learning technology, especially the emergence of convolutional learning nets [3], the image feature extraction ability of this model has been greatly improved. In addition, classifiers such as support vector machines [4] can classify and predict the extracted image features, effectively reducing the difficulty of feature selection.

Currently, the mainstream target identification algorithms are of the R-CNN series [5] and the YOLO series [6]. However, R-CNN series detection algorithms are two-stage detection algorithms, which will take a lot of time even for the faster R-CNN algorithms. YOLO series algorithms have been developed to YOLOv5s [7], which is already a first-stage detection algorithm, and can complete the detection process of R-CNN series algorithms in one step, and the pace of detection will be able to satisfy the high-speed needs.

Research on helmet recognition attracted the interest of numerous scholars: Zhang et al [8] proposed a colour- and shape-based method to detect helmets, and Li et al [9] trained Support Vector Machines (SVMs) to classify pedestrians based on the results of HOG feature extraction. However, traditional object recognition methods are only applicable to certain scenes with low precision. Along with the evolution of object recognition technique, Deng et al [10] advanced the YOLOv4 method, which successfully solved the recognition of helmets under different recognition scales. The deep learning-based helmet wearing state detection method has high requirements for the experimental environment, and it is difficult to achieve the detection and analysis of large datasets in general experimental environments, and the detection accuracy needs to be improved.

In order to solve the situation of small-sized target miss detection, this paper uses YOLOv5s network as the base model, combined with the structure of weighted bidirectional pyramidal feature network (BiFPN) [11] to improve the feature fusion process to improve the detection accuracy of the model. In yolov5s feature extraction network the introduction of CBAM [12] attention mechanism of the convolution module to obtain more feature details, the use of SIOU [13] to replace CIOU as the prediction frame regression loss, improve the speed and accuracy of network training and inference.

EITCE 2023, October 20–22, 2023, Xiamen, China

ACM ISBN 979-8-4007-0830-5/23/10
https://doi.org/10.1145/3650400.3650418

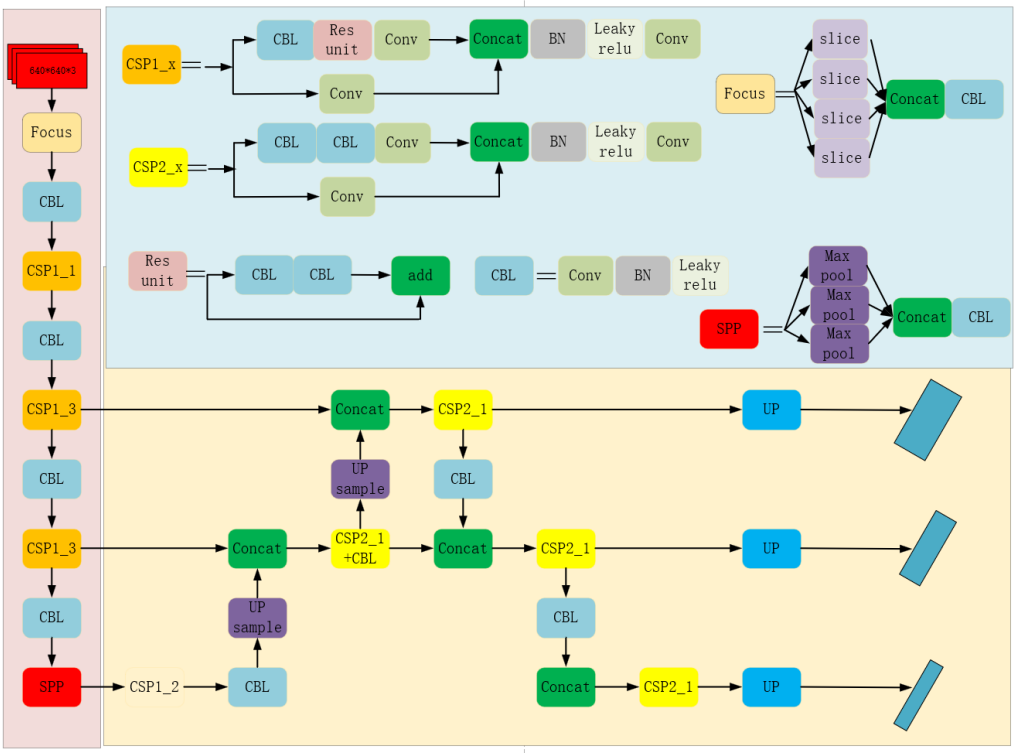

Figure 1: YOLOv5s structure

The target object can be accurately detected, and it shows good robustness and generalisation ability to detect the helmet wearing state in different sizes, angles and complex backgrounds.

2 YOLOV5S ALGORITHM

Yolov5s is fast and flexible and consists of the following four parts: Input, Backbone, Neck and Head.Input performs the pre-processing of the input data set, including operations like mosaic data enhancement, calculation of the anchor frame adaptively, etc.Backbone employs the CSPDarknet53 network to obtain the highly informative features from the input image. The core in Neck is the Feature Pyramid (FPN) and Path Aggregation Network (PAN) structure to achieve the fusion of feature information at different scales.Head is the recognition feature structure of yolov5s, outputting three different size feature maps, big, middle and tiny, corresponding to the detection of small, middle and big targets, respectively. Among them, Backbone is the backbone structure of yolov5s, which includes modules such as Foucs, Conv, C3, SPP, etc. The Foucs module slices the input at longitudinal and transverse intervals and then splices them together, Conv contains convolution, regularisation and activation layers, and C3 contains N residual networks Bottleneck, the input feature maps are first passed through the convolution layers of 1×1 and 3×3, and then the result is summed with the input features. SPP is the Spatial Pyramid Pooling Layer.

In this study, the yolov5 algorithm is improved in three aspects to solve the problems related to missed and wrong recognition in helmet-wearing scenarios at construction sites. First, a helmet detection object model using the enhanced yolov5s using a model structure network is proposed as shown in Figure 1. The CBAM attention mechanism is added in the network backbone module so as to strengthen the ability of character feature extraction; the structure of the PANet network is changed from PANet to BiFPN in Neck so as to strengthen the multi-scale feature fusion; in addition, in view of improving the accuracy of prediction and the speed of regression, the SIOU as the loss of prediction frame regression is used to replace the CIOU.

3 IMPROVEMENT OF YOLOV5S HELMET DETECTION ALGORITHM

3.1 BiFPN Improvement

YOLOv5s uses an FPN+PAN structure for combining feature maps derived from the backbone network for extracting features, using a bottom-up PAN network structure that simply sums feature maps from different paths and with the same width and height dimensions in the channel dimension. This aggregated network structure assumes that all input features contribute consistently to the output features, but the method has some shortcomings. Since the scale of the construction site safety helmet target varies greatly, the original feature fusion technique will destroy the feature consistency of smoke at different scales. To tackle this issue, this work adopts the BiFPN structure in the part of feature fusion to improve the neck part. BiFPN is a weighted bidirectional feature pyramid, and there

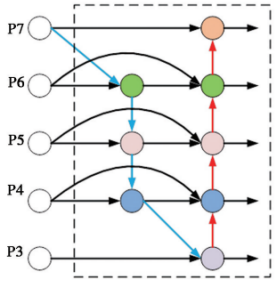

Figure 2: BiFPN structure

are two main ideas in BiFPN: one is the efficient bidirectional cross-scale connectivity, and the other is the fusion of weighted feature maps. Using the four terms of bidirectional fusion, constructing both top-down and bidirectional bottom-up channels, fusing pieces of feature data from various levels of the backbone network, fusing between different levels by up-sampling and down-sampling the same scales of feature resolution, and adding lateral connections among the initial input and output vertices of the same function, more features can be fused without increasing the cost, and in addition, the BiFPN is treated as a basic unit, i.e., a pair of paths in the feature pyramid. e.g., a pair of paths in the BiFPN is considered to be a weighted bipartite feature pyramid, and a pair of paths in the BiFPN is considered to be a weighted bipartite feature pyramid. In addition, the BiFPN is considered as a basic unit, i.e., a pair of paths in the BiFPN is considered as a feature layer, and then it is repeated many times to obtain more high-level feature fusion, and the network structure is shown in Figure 2.

3.2 Introduced CBAM

Introducing CBAM attention mechanism in yolov5s backbone network provides a Simple and Powerful Neural Network Attention Module for the feed-forward type of convolutional neural networks. CBAM introduces channel and spatial attention, can selectively focusing on key details and suppressing unimportant details, thereby enhancing the ability of the model to extract important information and improving the ability of convolutional neural networks to focus on images. CBAM introduces channel and spatial attention, can selectively focusing on key details and suppressing unimportant details, thereby improving the ability of convolutional neural networks to extract important information and focus on images. CBAM introduces channel and spatial attention, allowing selective focus on important features and suppression of unimportant features, thereby enhancing the model's ability to extract important information and improving the ability of convolutional neural networks to focus on images.CBAM computes the feature map's attention map using its channel and spatial extent, and then multiplies the intention map by the input feature map to adaptively learn features. For the feature maps in the middle layer, CBAM will sequentially derive the one-dimensional channel attention map and the bidimensional spatial attention map, the symbol ⊗ denotes the sincere element, and $F^{'}$ denotes the new features obtained by the channel attention, and in as the inlet of the spatial attention, and finally get the features $F^{''}$ of the whole CBAM output. Its formula is:

$$F^{'} = M\_C\ (F) \otimes F \tag{1}$$

$$F^{''} = M_S\left(F^{'}\right) \otimes F^{'} \tag{2}$$

The CBAM module is added to YOLOv5s to enhance the model's ability to detect facial features and improve detail capture. This enhancement is used to improve overall target recognition accuracy and performance. The structure of the CBAM module is shown in Figure 3.

3.3 SIoU loss function improvement

For the target recognition model, accurate target localisation is achieved using the regression module of the boundaries. For the YOLOv5s network model, the overall loss function has three components: localisation loss, classification loss and confidence loss. The localisation loss uses the CIoU (complete intersection and merger ratio) as a computational criterion to measure the agreement of the prediction bounding box with the true bounding box. The confidence loss and classification loss use a dichotomous cross-entropy loss function to measure the confidence in the existence of the target and the classification accuracy of the target category.

When the centroids of the real and predicted frames coincide, but the aspect ratios are different and the intersection and fusion ratios are the same, the CIoU adds a penalty term to this, and there will be relatively large fluctuations in the convergence process. To solve this problem, this study introduces the SIOU loss function, increases the angular cost, and uses the angular cost to re-describe the distance, which reduces the probability of the penalty term being 0 and makes the loss function converge more uniformly. The CIoU is calculated as shown in equation 5):

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{3}$$

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

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{5}$$

b——prediction box

b^{gt}——true box

c^2——The square of the difference between the real frame's aspect ratio and the predicted frame's normalised aspect ratio.

α——equilibrium parameter

v——indicates whether the measure of aspect ratio is consistent or not

From Equation 4), when the aspect ratio of the predicted frame and the real frame are the same, the aspect ratio penalty term in the localisation loss loses its effect, and the localisation loss function only considers the positional offset and scale difference between the two frames and does not additionally penalise the aspect ratio, i.e. the CIoU loss function cannot converge more uniformly.

In this study, using SIoU loss function instead of CIoU loss factor in YOLOv5s, introducing the angle information of the regression vector, and improving accuracy and robust target recognition by redefining the penalty term. The above improvements can better handle various target shapes and angles and improve the quality of detection results. The parameters used in the SIoU loss function are shown in Figure 4.

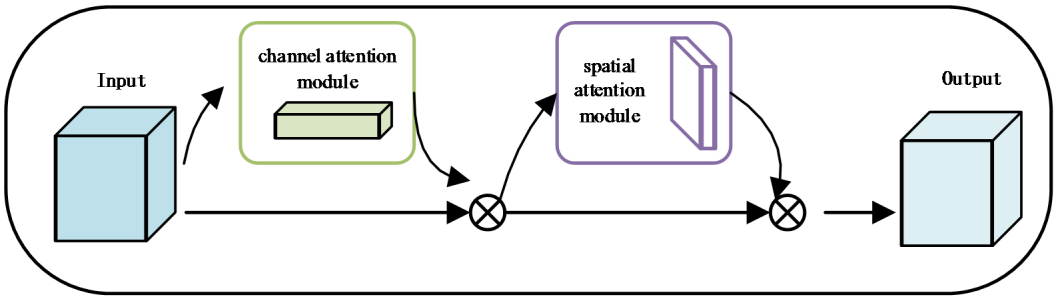

Figure 3: CBAM module structure diagram

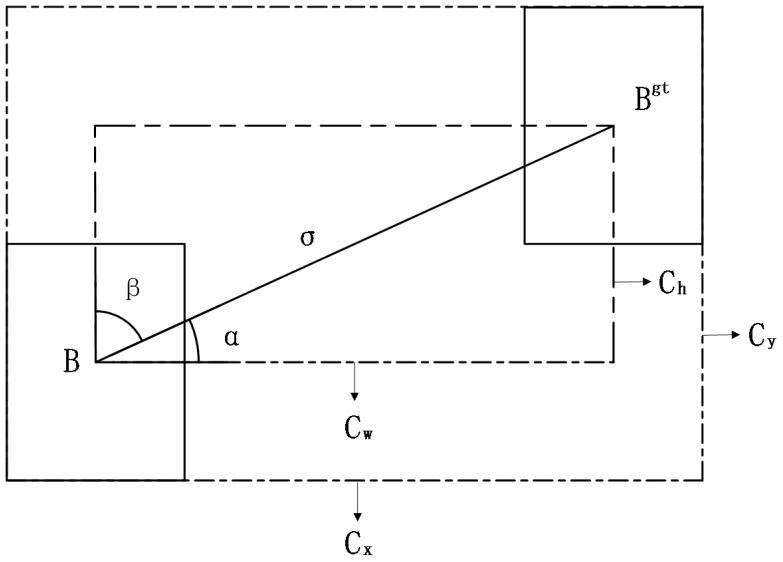

Figure 4: SIoU loss function calculation plot

4 EXPERIMENTS AND ANALYSIS OF RESULTS

4.1 Experimental Configuration

The experimental operating system is Windows 11 (64-bit), 16 GB RAM, RTX4050 graphics card driver with AMD Ryzen 7 7735H with Radeon Graphics 3.20 GHz processor, Anaconda 4.12.0 programming platform, CUDA 11.6, PyTorch development environment, Python 3.8 programming.

4.2 Experimental data set

The dataset used for the experiment is the images of hard hats on construction sites obtained from the Vision China website, the AI Studio website, as well as on-site photography and web crawling. The 2000 images were collected as a dataset and the dataset was labelled using the labelling library that comes with Python. It is then randomly assigned as Training and Verification sets in an 8:1:1 ratio. Some of the images are shown in Figure 5.

4.3 Assessment of indicators

Precision (P) is a measure of model performance as shown in equation 6); Recall (R) is a measure of classification model performance as shown in equation 7). A higher recall rate means that the model has a lower chance of missing detections, which may lead to an increase in false detections, thus reducing the precision rate. To evaluate the algorithm's overall performance, precision and recall can be combined using Average Precision (AP) by calculating precision and recall at different thresholds and calculating the combined metrics. For multi-classification problems, the AP can be averaged over the dimensions of the categories to obtain the Mean Average Precision (mAP):

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

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{7}$$

$$\mathrm{AP} = \int_0^1 \mathrm{P}\,(\mathrm{r})\,\mathrm{dr} \tag{8}$$

$$\mathrm{mAP} = \frac{AP}{Num_{class}} \tag{9}$$

TP——Sample size of correctly predicted helmets

Figure 5: Example chart of selected data sets

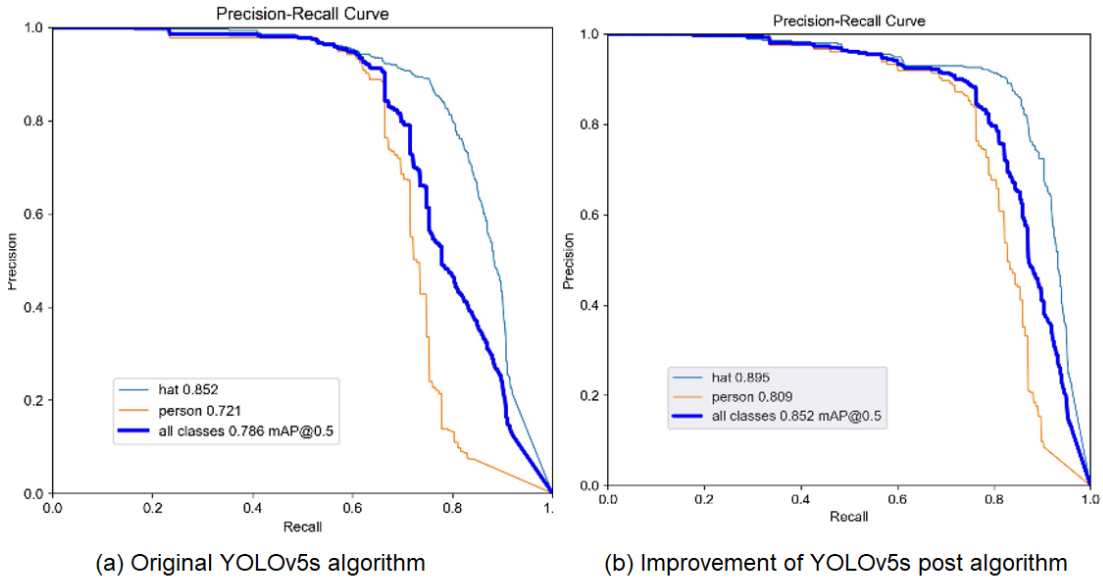

Figure 6: Comparison of algorithms before and after improvement

TP——Predicting background error as helmet sample size
FN——is the number of samples that misclassify the target frame as background

4.4 Experimental results of the improved algorithm

The YOLOv5s algorithm and the improved algorithm are trained separately using the training set on the dataset processed by the method of this paper and then validated on the test set, respectively. The P-R curves of the original YOLOv5s algorithm and the improved algorithm network on the test set are shown in Figure 6.

From Figure 6, it can be concluded that in the P-R curves derived from the tests of the improved YOLOv5s algorithm, the trend of decreasing precision with increasing recall is faster, and the improved YOLOv5 algorithm has a larger P-R area for each category as well as for the total category, and the curves converge better.

4.5 Ablation experiment

To verify the effectiveness of the improved YOLOv5s algorithm, ablation comparison experiments are set up to verify the effect of each improved strategy on the model performance, which is divided into five sets of experiments. Let the original model be Experiment 1, Experiment 2 adds the improved bidirectional feature pyramid network (BiFPN) structure based on Experiment 1, Experiment 3 adds the CBAM attention mechanism based on Experiment 1, Experiment 4 adds the improved SIoU loss function based on Experiment 1, and Experiment 5 performs the ablation comparison test for all added modules. The results of the experiments are shown in Table 1.

Table 1: Ablation experiment

algorithmic model	P%	R%	mAP%	mAP@0.5%
YOLOv5s	86.2	71.2	78.6	52.9
YOLOv5s+ BiFPN	88.5	78.3	83.7	56.8
YOLOv5s+CBAM	88.5	79.6	84.7	57.2
YOLOv5s+SIoU	89.3	78.2	84.9	57.1
The algorithms in this paper	86.8	79.2	85.2	57.6

Figure 7: Detection effect diagram

From Table 1, it can be seen that for the same number of iterations, the original model Precision, Recall and mAP values of YOLOv5s are relatively low, and after adding the BiFPN structure, CBAM attention mechanism and SIoU loss function, the average accuracy mAP is improved by 6.6 percentage points.

4.6 Results

The driver dataset is trained and validated using the improved model and tested on the test set and the detection results are shown in Figure 7. The improved algorithm is tested on construction workers wearing helmets and the improved model is very effective in detecting construction site helmets.

5 CONCLUSION

For the construction site construction site exists in the safety helmet wearing irregularities and and other helmets similar to the hat of the misdetection of the problem, visual China website, AI studio website, as well as on-site shooting and network crawler way to obtain on the construction site safety helmets of the image has a certain degree of authenticity, And proposes a construction site safety helmet based on the improvement of YOLOV5S algorithm for the detection of helmets model, which combines the weighted bidirectional feature pyramid network (BiFPN) structure to improve the feature fusion process and increase the model's detection accuracy of helmets; The convolution module in yolov5s feature extraction network introduces the CBAM attention mechanism to obtain more details of helmets, and predicts targets of different sizes; the SIOU is used to replace the CIOU as the prediction frame regression loss to makes the network training and inference process faster and more accurate. The network is applied to detect the helmet wearing state in real scenarios, and from the experimental results, in the helmet wearing detection task, the mAP of the improved YOLOv5s detection model reaches 85.2%, which is 6.6% higher than that of the traditional YOLOv5s, ensuring higher detection accuracy to meet the requirements of real-time detection.

REFERENCES

[1] Ma Junbo, YAO Wenli, SUN Shimei. Research on all-round monitoring and correction system of unsafe behavior in construction accidents. Science and Technology Innovation, 2021, (20): 161-162.
[2] Shenoy, M. Ashwin, Pranam R. Betrabet, and NS Krishnaraj Rao. Helmet Detection using Machine Learning Approach. 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2022.
[3] Li, Zewen, *et al.* A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 2021.
[4] Zou Z, Chen K, Shi Z, *et al.* Object detection in 20 years: A survey. Proceedings of the IEEE, 2023: 257-276.
[5] Li L, Hassan M and Yang S, *et al.* Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies. The Crop Journal 10.5, 2022: 1303-1311.
[6] Anushkannan, N. K., *et al.* YOLO Algorithm for Helmet Detection in Industries for Safety Purpose. 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2022.
[7] Wang, Lijun, *et al.* Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN. IEEE Access 10, 2022: 60622-60632.
[8] Zhang, Geng, *et al.* The method for recognizing recognition helmet based on color and shape. 2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017). Atlantis Press, 2017.
[9] Li, Jie, *et al.* Safety helmet wearing detection based on image processing and machine learning. 2017 Ninth international conference on advanced computational intelligence (ICACI). IEEE, 2017.
[10] Benyang, Deng, Lei Xiaochun, and Ye Miao. Safety helmet detection method based on YOLO v4. 2020 16th International conference on computational intelligence and security (CIS). IEEE, 2020.
[11] Zhang, Huanlong, *et al.* A recursive attention-enhanced bidirectional feature pyramid network for small object detection. Multimedia Tools and Applications 82.9, 2023: 13999-14018.
[12] Fu H, Song G and Wang Y. Improved YOLOv4 marine target detection combined with CBAM [J]. Symmetry, 2021, 13(4): 623.

[13] Gevorgyan Z. SIoU loss: More powerful learning for bounding box regression [J]. arXiv preprint arXiv:2205.12740, 2022,