International Journal of Science and Engineering Applications
Volume 13-Issue 10, 110 – 117, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1023

A Safety Helmet Target Detection Algorithm Based on Improved YOLOv7

`Lin Qihuang
College of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou City, Hubei Province
China

Luo Litao
College of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou City, Hubei Province
China

Abstract: Aiming at the problems of low detection accuracy and poor effect of safety helmet wearing in heavy industrial sites such as construction industry and construction site, a YOLO (you only look once) v7 detection algorithm for safety helmet is proposed. The Convolutional Block Attention Module (CBAM) is combined with the neck network of YOLOv7 to reduce the interference of complex background and improve the attention of operators. Then, the detection head of YOLOv7 is replaced by a dynamic detection head DyHead (Dynamic Head) to unify multiple attention operations and improve the efficiency of feature fusion, so as to effectively solve the feature fusion problem of small-size target detection. By optimizing the model bounding box regression, the W-IOU (weighted Intersection over Union) loss function replaces the original loss function to improve the model training speed and accuracy. The experimental results on the SHWD safety helmet dataset show that the improved algorithm improves the average detection accuracy by 2.2 %, the accuracy rate by 1.1 %, and the recall rate by 1.8 %.The improvement of this paper makes the model more accurate to identify the target, and the detection effect is greatly improved.
Keywords: Target detection ; attention mechanism ; dynamic detection head ; loss function ; Yolov7

1. INTRODUCTION

Risky companies such as construction and heavy engineering focus on safe production management, and in such a construction and production environment, the role of the helmet becomes crucial. [1] Although the traditional manual monitoring method has a series of problems such as low efficiency, narrow management coverage, insufficient timeliness, and inability to conduct comprehensive monitoring, the safety helmet wearing monitoring method based on target detection [2] is being popularized and used by many companies.

In recent years, the industry 's enthusiasm for target detection methods using deep learning [3] has continued to rise. In the field of target detection, the detection of small targets [4] has always been a difficult point in this field. Sample images of safety helmet wear collected at the industrial site. These sample images usually have defects such as low resolution [5], dense wearing of safety helmets, and easy occlusion and fewer pixels of related targets, resulting in higher detection difficulty [6]. Therefore, in recent years, many scholars have studied it and proposed a series of target detection algorithms based on deep learning.

In 2020, Zhang et al. [7] proposed a method to realize target detection by weighted fusion of shallow and deep feature maps before and after sampling by convolutional neural network. In 2021, Zhou et al. [8] proposed a method to enhance important features by optimizing the generation of default boxes on the network feature map based on the lightweight SSD model. The above two-stage model detection methods greatly improve the detection accuracy of the model. In the field of target detection, traditional algorithms usually have high detection accuracy, but they perform slowly in detection speed and are difficult to meet the needs of actual scenes. The Redmon team [9] first proposed the YOLO detection algorithm in 2015, and the algorithm can not only maintain the ideal detection accuracy, but also achieve the speed of detecting real-time video (45 frames / second). With the subsequent development of the YOLO series model, it not only has been greatly improved in accuracy, but also has reached the standard of real life needs in detection speed.In 2017, YOLO v2 detection algorithm and YOLO v3 detection algorithm were proposed by Redmon et al. [10]. Among them, YOLO v3 algorithm has better detection effect. The detection effect of YOLO v3 algorithm on COCO dataset is 3.8 times faster than that of RetinaNet algorithm in achieving similar mAP (mean Average Precision) accuracy. Therefore, the YOLO model with better performance and practicability has received more and more researchers ' favor. Based on the YOLOv4 network model, Ilhamu Yarmaiti et al. [11] designed a multi-scale context feature fusion mechanism, and used a new feature fusion method to enhance the multi-scale space and channel information representation ability of the network, so as to improve the accuracy of target detection. In 2021, Wang et al. [12] proposed to design a new feature extraction network based on the YOLOv5 network model by combining the SE attention mechanism, thus highlighting the feature information of the detection target and enhancing the detection performance of the target.In 2023, Qi et al. [13] proposed to improve the MPconv convolution module through the design idea of feature separation and merging for the YOLOv7 network model to reduce feature loss.

When dealing with the real-time and small detection target safety helmet wearing supervision task, because YOLOv7 has the characteristics of high detection rate and high detection accuracy [14], it is suitable for the above basic requirements of the supervision task. Therefore, this paper proposes an improved YOLOv7 algorithm for safety helmet wearing detection. The main algorithm innovations and contributions of this paper are as follows :

(1) Add attention mechanism. The CBAM [15] attention mechanism is integrated in the neck network to improve the feature representation ability, thereby improving the object

International Journal of Science and Engineering Applications
Volume 13-Issue 10, 110 – 117, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1023

detection performance and enabling the model to better locate the attention area in the dense object scene.

(2) DyHead [16] replaces the detection head of the original YOLOv7 model. The dynamic detection head is mainly calculated by combining three kinds of attention, which are scale perception, spatial location perception and task perception. It can better integrate the feature scale diversity brought by the difference of target scale, as well as the potential relationship between the shape and spatial position of each target.

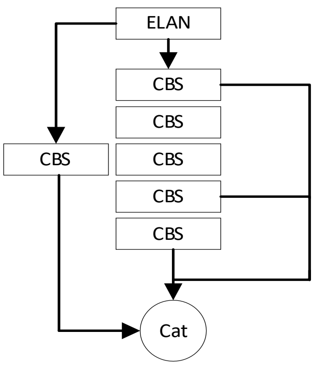
(3) Replacing the original loss function CIoU (Complete IoU) of YOLOv7 with the W-IOU loss function can better balance positive and negative samples, directly optimize the target detection index, adjust the significance weight of the target, and provide a training process that is easier to optimize and converge.

（4） On the SHWD public data set, the ablation comparison experiment is designed to analyze the influence of each module on the performance of YOLOv7, and the CBAM module improvement based on the attention mechanism and DyHead dynamic detection head are used. After repeatedly adjusting the parameters to obtain the final model, the experimental results show that compared with the original YOLOv7 algorithm, the performance of the new model in many aspects has been improved.

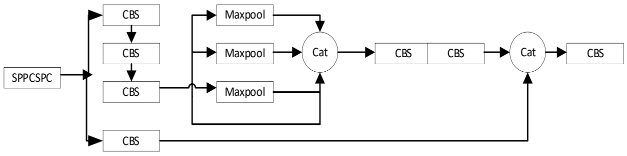
2. RELATED JOB

2.1 Network structure of YOLOv7

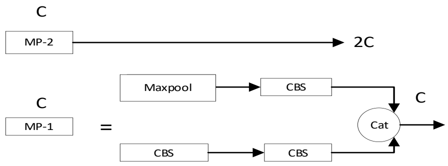
The YOLOv7 model consists of four key network layers : input layer, backbone network layer, feature fusion layer and output layer. The input layer adjusts the image size to 640 × 640. The backbone network uses deep convolution to extract different scale features. The feature fusion layer fuses different scale feature information. The output layer performs non-maximum suppression on the predicted anchor frame coordinates, categories and confidence, and finally outputs the result. The YOLOv7 model contains several highly integrated convolution operation modules, such as ELAN module, SPPCSPC module and MP module. These integrated structures are shown in Figure 1, which plays a key role in optimizing the model structure and improving the feature extraction ability.



(a) ELAN Model



（b）SPPCPC Model



（c）MP Model

Figure 1. Integrated convolutional structure

The ELAN module is shown in Fig.1 (a). By changing the feature extraction path, the feature information extracted by the backbone layer is increased to improve the robustness of the model. The SPPCSPC module is shown in Fig.1 (b), which mainly increases the receptive field, reduces the amount of calculation, and improves the efficiency of the model. The multi-channel feature map is divided into two parts. One part uses multi-size maximum pooling operation to obtain different receptive fields, and the other part uses conventional convolution operation to extract features. The MP module is shown in Fig.1 (c). The combination of maximum pooling and downsampling convolution is used to realize the deep downsampling of the feature map.

However, the YOLOv7 model is often disturbed by factors such as image noise and complex background when detecting small targets, and the network model itself fails to fully extract medium and shallow texture and contour information in small target detection, which affects the detection accuracy. These are the problems that many scholars are committed to solving.

2.2 Attention Mechanism

Attention mechanism is a data processing method in machine learning. In the task of target detection, the addition of attention module can improve the performance of target detection model in information screening, target location, model robustness and target feature enhancement. The attention mechanism enables the model to pay more attention to the important information in the image, thereby improving the accuracy and robustness of target detection.

In the feature extraction and fusion stage of YOLOv7, there is a problem of feature redundancy, and the middle and shallow texture and contour information of small targets cannot be fully extracted. These defects are easy to cause problems such as target error detection and missed detection in the process of small target detection. Liu Mingrui [17] et al.proposed in 2023 that the YOLOv7-based network enhances the feature extraction ability of the model for small targets by combining the SE attention mechanism module. In 2023, Zhang Yanjun [18] and others proposed a method based on the YOLOv7 network model. By fusing the CBAM of multiple convolutional attention modules into the output channel of the main network feature of the model, the key features of the output are weighted to improve the model 's anti-interference ability and target detection ability for non-essential and secondary features. Therefore, it can be seen that the fusion attention mechanism can better improve the network 's perception ability to the target, weighted feature selection, scale and location sensitivity, so as to further improve the detection accuracy of small targets and avoid errors and omissions.

2.3 Loss function

The bounding box loss function [19] plays a key role in object detection, and the rationality of its definition can significantly improve the performance of the object detection model. In recent years, a large number of studies have generally assumed that the training data contains high-quality examples to improve the fitting ability of the bounding box loss function. However, the observation that the small target detection training set often contains a large number of low-quality examples shows that too much emphasis on the regression of bounding boxes to low-quality examples may damage the performance improvement of the model in small target detection tasks. Therefore, it is necessary to optimize the loss function of the original model to avoid the blind regression of the model to the low-quality examples that account for a large proportion of the small target data set, resulting in the phenomenon of missed detection and wrong detection of small targets in the model detection.

3. ANALYSIS OF ALGORITHM IN THIS PAPER

3.1 Improvement of the Algorithm

The YOLO v7 algorithm has achieved good detection results in the application of target detection in common data sets. However, for a specific safety helmet wearing data set [20], due to the low resolution of the data set, dense targets, easy occlusion and sparse pixels, the target detection results will be missed and misdetected. Therefore, it is necessary to adjust and improve the YOLO v7 model for the safety helmet wearing detection task. Aiming at the problem that the YOLOv7 neck network cannot fully fuse important features and too many redundant features interfere with subsequent detection, the CBAM attention mechanism is integrated into the neck network to improve the efficiency of feature fusion. Aiming at the problem of how to focus more on the anchor frame of ordinary quality, the loss function is replaced to improve the overall recognition performance of the detector. The detection head of the YOLOv7 model is replaced by the dynamic detection head DyHead, which integrates scale, spatial location and task perception to adapt to the diversity features and relative positional relationships in the safety helmet data set.

3.2 CBAM Attention Module

The CBAM attention mechanism aims to enhance the modeling ability of the convolutional neural network for image space and channel dimensions, so as to better capture the global and local features of the image. The specific composition of the module is shown in Figure 2.

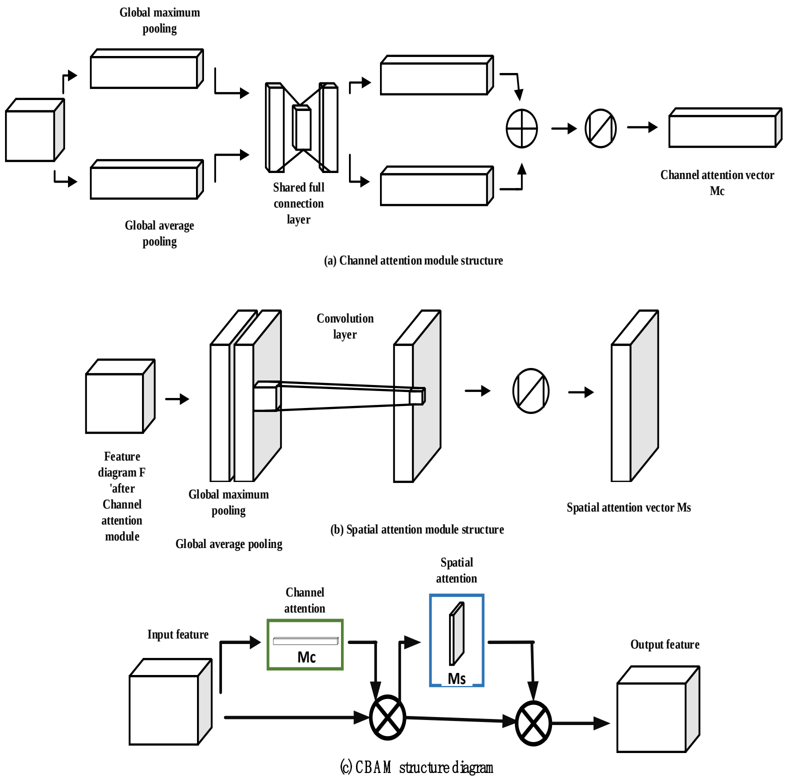


Figure 2. Schematic diagram of the structure of the attention mechanism of CBAM

The CBAM attention mechanism consists of two modules [21] : channel attention module and spatial attention module. The channel attention module is shown in Figure 2 (a), which helps to determine the importance of feature channels, eliminate the interference of general features, and focus on key channels. Through the combination of maximum pooling and average pooling, more comprehensive feature information is provided to ensure the learning of key features in the helmet image, thereby improving model performance. The spatial attention module is shown in Fig.2 (b), which can identify

International Journal of Science and Engineering Applications
Volume 13-Issue 10, 110 – 117, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1023

the location of small targets, and can effectively capture small targets that are easily overlooked and missed in the helmet. CBAM is a lightweight attention mechanism that can be easily integrated into any CNN convolutional neural network to achieve end-to-end training, which can effectively improve the performance of the model. The CBAM attention mechanism is integrated with the neck network of the YOLOv7 network model. The network structure is shown in Figure 3, which can effectively enhance the detection performance of the helmet without significantly increasing the memory overhead and network depth.

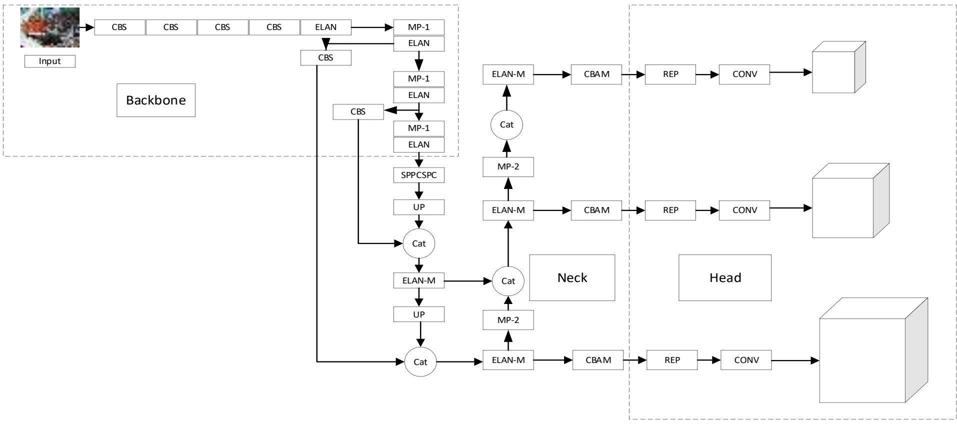


Figure 3. Schematic diagram of CBAM-YOLOv7 structure

3.3 DyHead module

In the process of target detection, the algorithm of YOLOv7 model performs down-sampling operation in the backbone network, in which the size of the feature map is reduced and the receptive field is increased by down-sampling operation. After several down-sampling operations, YOLOv7 can generate feature maps of different scales in order to detect targets of various sizes, but at the same time, its multiple down-sampling operations will lose a lot of location information and feature information, which makes the model When detecting small targets, there may be missed detection of smaller targets. Therefore, in order to better deal with the difference of target scale and the relationship between target shape and position, the YOLOv7 model replaces the detection head with a dynamic detection head called DyHead, which can realize the unification of multiple perceptual attention at the same time. The structure is shown in Figure 4.

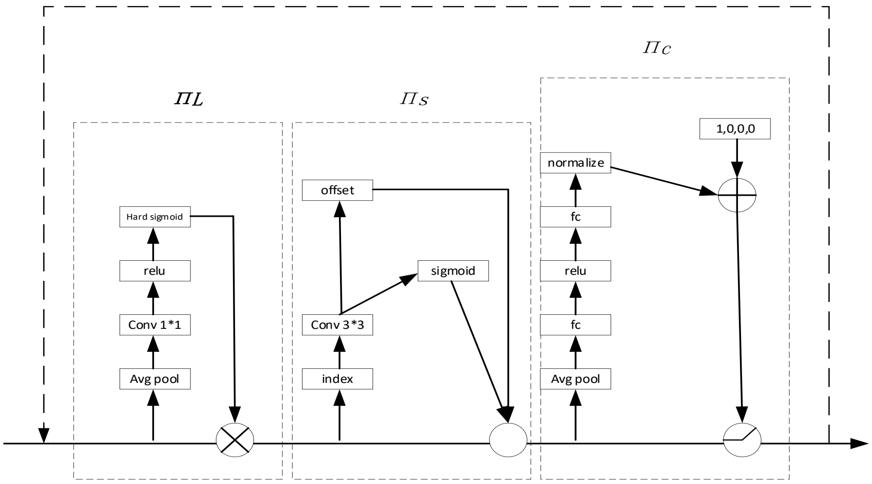


Figure 4. Schematic diagram of the DyHead block structure

That is, an attention mechanism is added to each specific dimension of the feature tensor. The attention function calculation formula is described as formula (1) :

$$W(F)=\pi_c\left(\pi_s\left(\pi_L(F)\cdot F\right)\cdot F\right)\cdot F \quad (1)$$

Where W (F) denotes the attention function, F is a three-dimensional characteristic tensor. The shape of the tensor is $L\times S\times C$, where L represents the different levels of the feature map generated by the middle layer of the neural network, S represents the product of the width and height of the feature map, which can be used to calculate the total number of pixels of the feature map, and finally C represents the number of dimensions of the feature map in the depth direction, referred to as the number of channels of the feature map.

Among them, and $\pi_{L}(\cdot)$ 、 $\pi_{S}(\cdot)$ 、 $p_j+\Delta p_j$ represent the scale, space, and task-aware attention modules, which are performed on L, S, and C, respectively.

The calculation process of the scale-aware attention module is expressed as Equation (2) :

International Journal of Science and Engineering Applications
Volume 13-Issue 10, 110 – 117, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1023

$$\pi_L(F)\cdot F = \sigma\left(f\left(\frac{1}{SC}\sum_{S,C}F\right)\right)\cdot F \qquad (2)$$

The F function is a linear function, which realizes feature dimension reduction by approximate 1 × 1 convolution operation. $\sigma(x)$ is a Hard-Sigmoid activation function. Compared with the traditional Sigmoid function, it is simpler and more efficient in calculation. The calculation process of the spatial perception attention module is expressed as Equation (3) :

$$\pi_S(F)\cdot F = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K} w_{l,k}\cdot F(l;p_k+\Delta p_k;c)\cdot\Delta m_k \qquad (3)$$

K is the number of sparse sampling positions, $p_j+\Delta p_j$ representing the movable position, which is used to focus the discriminant position. Δm_k is a self-learning importance scalar, which is determined by the input features of the F intermediate level.

The calculation process of the task-aware attention module is expressed as Equation (4) :

$$\pi_c(F)\cdot F = MAX\left(\alpha^1(F)\cdot F_C+\beta^1(F),\alpha^2(F)\cdot F_C+\beta^2(F)\right) \qquad (4)$$

Using three kinds of attention mechanisms, the dynamic detection head DyHead is formed by stacking multiple times in order to achieve different activation of different channels. Among them, F_C The feature slice of the C channel is shown, $\theta(\bullet)$ is a hyperfunction used to learn the control activation threshold, similar to Dynamic ReLU, α and β is a learnable parameter.

In this experiment, the original detection head is replaced by the dynamic detection head DyHead to provide the detection model with the advantages of dynamic receptive field, efficient information interaction and adaptive network architecture, so as to improve the performance and adaptability of the target detection algorithm.

3.4 Loss function

The YOLOv7 model uses CIoU Loss to calculate the rectangular box loss, which consists of three parts :Predict the loss of rectangular frame position (L_{bbox})、Loss of confidence (L_{obj}) and the classification loss (L_{cls}). The specific calculation formula of CIoU Loss is shown in Formula (5) :

$$L = L_{bbox} + L_{obj} + L_{cls} \qquad (5)$$

Due to the influence of the environment and the monitoring angle of the safety helmet in real life, the target size of the safety helmet that can be detected is small, which may lead to the increase of the regression error, and then cause the imbalance of the training samples, that is, the low quality samples with large regression error are far more than the high quality samples with small error. In this study, the W-IOU loss function with dynamic non-monotonic focusing mechanism is used to improve the learning effect of the target detection model on difficult samples in the training process. By dynamically adjusting the weight of each sample in the loss function, the dynamic non-monotonic focusing mechanism makes the difficult samples get more attention in the optimization process, so as to improve the learning ability of the model to these difficult samples. In this way, W-IoU can focus more on the anchor frame of ordinary quality, thus improving the overall performance of the detector. Its calculation formula is shown in formula (6) :

$$L_{Wiou} = \frac{\beta}{\delta\alpha^{\beta-\delta}}\exp\frac{(x-x_{gt})^2+(y-y_{gt})^2}{(W_g^2+H_g^2)^*}L_{IOU} \qquad (6)$$

In the formula, x, x_{gt}, y, y_{gt} the center point coordinates of the prediction frame and the real frame are W_g, H_g the width and height of is the minimum circumscribed frame of the real frame and the prediction frame, and the learning parameters α,δ are used to adjust the calculation process. The quality of the bounding box indicates that as the quality of the bounding box β increases, the smaller the value, the higher the quality of the bounding box. The calculation formula $\beta = \frac{L_{IOU}^*}{L_{IOU}}$ is that * in the formula means that a variable is converted into a constant.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental platform

The Window11 operating system is used in the experiment, and the running environment based on the PyTorch framework is built. The environment is configured with the CUDA 11.7 acceleration toolbox. The hardware configuration includes Intel i5-12400F processor, 16GB memory and NVIDIA RTX3060Ti graphics card.

4.2 Experimental evaluation index

In this experiment, the evaluation of the performance of the model will use the mAP value [22] as the evaluation index of the experimental results, which is used to detect whether the prediction box category and position of the model to be detected are correct. It is calculated by the accuracy rate and the recall rate, where the accuracy rate represents the proportion of data predicted as positive samples and predicted correctly. Recall rate refers to the proportion of all real positive examples that are correctly predicted as positive by the model. Its calculation formula is as shown in formula (7) :

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN} \qquad (7)$$

In the formula, the TP (true positive) classification model correctly predicts the number of positive samples ; fP (false positive) refers to the situation that the classification model incorrectly predicts negative samples as positive samples ; fN (false negative) refers to the case where the classification model incorrectly predicts positive samples as negative samples ; tN (true negative) refers to the case where the classification model correctly predicts negative samples as negative samples. In the data set, the mean of the average detection accuracy (AP) of each category is called the mean of average accuracy. The calculation formulas are as follows : (8), (9) :

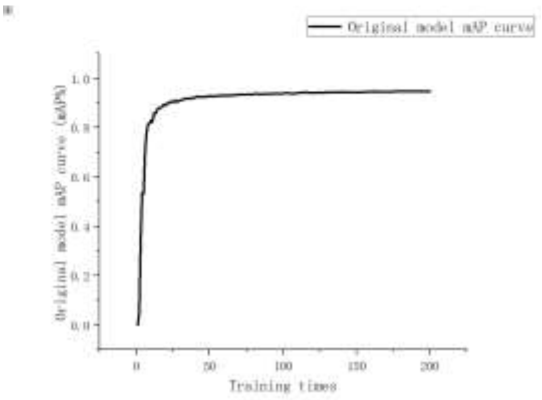
$$AP = \int_0^1 P(r)dr \qquad (8)$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \qquad (9)$$

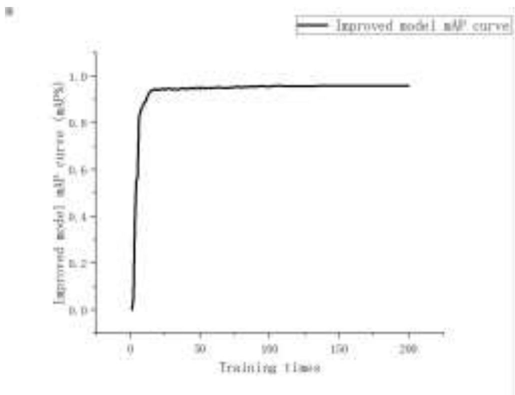
In the formula, n represents the total number of categories in the sample, and AP is the total area under the Precision-Recall curve.

4.3 Experimental training results

After 200 rounds of model training, the loss function of the original model was replaced by the W-IOU loss function, and the change of the average accuracy rate curve was observed. Figure 5 (a) shows the average accuracy curve of the original model when the IoU threshold is 0.5, while Figure 5 (b) shows the average accuracy curve of the improved model under different IoU thresholds. It can be seen from the curve trend that in the first 30 rounds of training after the model is improved, the model is observed to converge faster. After 50 rounds of training, the mAP value was stable, and no under-fitting or over-fitting was observed. On the whole, this shows that the model training effect is good.



（a） map_0.5:0.95 curve



(b) map_0.5:0.95 curve

4.4 Ablation experiment

In this paper, a series of ablation experiments are carried out on the SHWD safety helmet data set to verify the actual effect of the adopted method in the safety helmet target detection. Table 1 lists the application of each module, √ represents the use of the module.

In the table, YOLOv7 _ 4, YOLOv7 _ 5, YOLOV7 _ 11 corresponds to the addition of W-IOU, CBAM, Dyhead modules ; the W-IOU used in this study performs better than SIOU and Focal IOU in optimizing the network loss function. Compared with the above two loss functions, the mAP in this paper is 1.3 % and 1 % higher, respectively. Compared with the CIOU used in YOLOv7, the mAP of the two targets to be inspected is also improved, and the overall mAP is increased by 0.9 %.

In order to improve the attention of the network to the target, the attention module CBAM used in this paper is improved by 2.4 %, 3.1 %, 2.5 %, 0.6 % and 1 % on the average detection accuracy mAP0.5 compared with the commonly used attention modules GAM, SE, ECA [23], MHSA [24] and EA, while for the original YOLOv7, mAP0.5 is improved by 0.9 %.

In order to better integrate the difference of target scale [25] and the potential positional relationship characteristics of target shape and spatial position, a dynamic detection head DyHead [26] is introduced. It can be seen from the table that the effect of using this module is improved by 0.8 % compared with the original YOLOv7 network.

After adding W-IOU, CBAM and DyHead modules at the same time, the effect of the three targets to be detected is significantly improved. On the whole, compared with the original YOLOv7 network, it is improved by 2.2 %. From the comprehensive effect point of view, the method in this paper performs best compared with the previous methods, and the detection effect reaches the optimal level.

Table 1. Comparison of experimental results under different conditions

Network Name	Siou	Facal IOU	W-iou	CBAM	GAM	SE	MHSA	EA	ECA	DyHead	mAP@0.5%		all mAP_0.5%	all mAP_0.5%-0.95%
											person	hat		
YOLOv7_1											93.92	95.2	94.6	61
YOLOv7_2	√										93.2	94.6	93.9	60.4
YOLOv7_3		√									92.1	93.2	92.6	60.7
YOLOv7_4			√								94.9	96.2	95.5	61.9
YOLOv7_5				√							94.9	96.1	95.5	61.6
YOLOv7_6					√						94.5	91.7	93.1	58.8
YOLOv7_7						√					90.9	93.8	92.4	58.5
YOLOv7_8							√				94.1	91.9	93	59.1
YOLOv7_9								√			93.7	95.3	94.5	60
YOLOv7_10									√		94.2	95.5	94.9	60.7
YOLOv7_11										√	95	95.8	95.4	61.6

YOLOv7_12		√	√	95.1	96	95.5	61.9
YOLOv7_13	√		√	94.9	95.8	95.3	61.4
YOLOv7_14	√	√		95.2	96.1	95.6	61.6
Proposed Algorithm	√	√	√	**96.5**	**97.1**	**96.8**	**62.7**

4.5 Experimental comparison

For the verification of helmet target detection performance, we need to compare the improved network with the original network and other representative networks (such as Faster RCNN, SSD, EfficientDetV2, RT-DETR, CentNet, YOLOv5, YOLOX, etc.) for mAP indicators. In the person category, the proposed algorithm is improved by 7.27 %, 13.04 %, 9.25 %, 20.82 %, 40.15 %, 2.96 %, 4.27 % and 2.58 % respectively compared with the above mainstream algorithms. Compared with the above mainstream algorithms, the proposed algorithm is improved by 11.37 %, 11.48 %, 7.76 %, 18.87 %, 38.67 %, 2.48 %, 3.85 % and 1.87 % respectively. As shown in Table 2, the improved YOLOv7 network shows a significantly improved average accuracy compared with the classical YOLOX network, YOLOv5 network and the original YOLOv7 network when detecting two targets. Different from the yolo series algorithm, it also has a large lag in the average detection progress compared with the improved yolov7.Therefore, it can be concluded that the detection accuracy of the proposed algorithm in the target detection for helmets has a greater advantage than the current mainstream algorithms.

Table 2. Comparison of experimental results of different network models

Network Name	mAP@0.5%	
	person	hat
Faster Rcnn	89.23	85.73
EfficientDetV2	83.46	85.62
RT-DETR	87.25	89.34
CentNet	75.68	78.23
SSD	56.35	58.43
YOLOv5	93.54	94.62
YOLOX	92.23	93.25
YOLOv7	93.92	95.23
Proposed Algorithm	96.5	97.1

5. CONCLUSION

This paper improves the YOLOv7 algorithm, including the fusion of the CBAM attention module, the improved loss function, and the replacement of the detection head as the dynamic detection head Dyhead. After training on the SHWD safety helmet dataset, compared with the original YOLOv7 algorithm, our model has improved accuracy, reduced false detection and missed detection, and effectively improved the efficiency of safety helmet wearing supervision.

The improved algorithm can be integrated into the complete safety inspection framework, and can be widely deployed in various production sites, which has strong application and promotion potential. Next, the algorithm structure will be further improved to ensure its safe and reliable application in the safety helmet detection work of the actual factory scene.

6. REFERENCES

[1] Xu Shoukun, Wang Yaru, Gu Yuwan, et al., research on helmet wearing detection based on improved Faster Rcnn [J].Computer application research, 2020.37 (3) : 901-905.

[2] Li Mingshan, Han Qingpeng, Zhang Tianyu, etc. An improved SSD safety helmet detection method [J].Computer Engineering and Applications, 2021,57 (8) : 192-197.

[3] Wu Xue, Song Xiaoru, Gao Song et al. Overview of target detection algorithms based on deep learning [J].Sensors and microsystems, 2021,40 (02) : 4-7 + 18.DOI : 10.13873 / J.1000-9787 (2021) 02-0004-04.

[4] Liang Hong, Wang Qingwei, Zhang Qian et al. Review of research on small target detection technology [J].Computer engineering and application, 2021,57 (01) : 17-28

[5] Lv Zongzhe, Xu Hui, Yang Xiao et al. YOLOv5 helmet detection algorithm for small targets [J]. Computer Applications, 2023, 43 (06) : 1943-1949.

[6] Vang M, Wa Z,Zhou Z.Fie-yraied ietitatonreseach ofcopests and diseases based on improved CBAM va attention]J Trans Cchin Soc Ag rc Mach, 2021,52.29-247.

[7] Zhang Shihui, Wang Honglei, Chen Yuxiang and so on. Target detection method based on deep learning using feature map weighted fusion [J].Statistics, 2020,41 (11) : 1344-1351.

[8] Zhou Yan, Liu Wenping, Luo Youqing and so on. A method for detecting small target affected trees based on deep learning [J]. Forestry Science, 2021,57 (03) : 98-107.

[9] REDMON J, DIVVALAS,GIRSHICK R,et al.You Only Look Once:Unified,Real-Time Object Detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR) . Las Vegas, NV, USA, 2016.

[10] REDMON J， FARHADI A. YOLO9000: Better, Faster ,Stronger[C]//2017 IEEE Conference on Computer Visionand Pattern Recognition(CVPR). Honolulu, HI.USA,2017.

[11] REDMONJFARHADIA.YOLOW3:An Incremental Improvement[J]. arXiv e-prints ,2018 :1804 -2767.

[12] Ilhamu Yarmaiti, Bai Pengfei. Research and application of small target detection method based on YOLOv4 [J].Computer Engineering and Design, 2023,44 (01) : 277-283.DOI : 10.16208 / j.issn1000-7024.2023.01.037.

[13] Wang Yang, Cao Tieyong, Yang Jibin, etc. Research on camouflage target detection technology based on YOLO v5 algorithm [J].Computer Science, 2021,48 (10) : 226-232.

[14] Qi Linglong, Gao Jianling. Small target detection based on improved YOLOv7 [J]. Computer Engineering, 2023,49 (01) : 41-48.DOI : 10.19678 / j.issn.1000-3428.0065942.

[15] WANG C, BOCHKOVSKIY A, LIAO H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv preprint arXiv: 2207.02696 (2022)

[16] DAI X, CHEN Y, XIAO B, et al. Dynamic Head: Unifying object detection heads with attentions[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, June 20-15, 2021. Piscataway: IEEE, 2021: 7373-7382.

[17] Chen L,Liu C,Zhishuang X, et al. A Novel Safety Helmet Wear Detection Method based on Improved RFBnet[J]. Journal of Physics: Conference Series,2020,1617(1).

[18] Liu Mingrui, Che Ben, Dong Hongbo and so on. Small

target detection of UAV remote sensing image in open-pit coal mine stope [J / OL]. Coalfield geology and exploration : 1-9 [2023-10-08]. http://kns.cnki.net/kcms/detail/ 61.1155.P.20230831.1012.002.html

[19] Zhang Yanjun, Shen Ping, Guo Anhui, etc. Research on intelligent detection method of pavement defects based on CBAM-YOLOv7 model [J / OL].Journal of Chongqing University of Technology (Natural Science) : 1-7 [2023-10-09].http://kns.cnki.net/kcms/detail/50.1205 T.2023062 9.1123.002.html.

[20] Lv Zongzhe, Xu Hui, Yang Xiao and so on. YOLOv5 helmet detection algorithm for small targets [J]. Computer Applications, 2023,43 (06) : 1943-1949.

[21] Jin Yufang, Wu Xiang, Dong Hui, etc. Helmet wearing detection algorithm based on improved YOLO v4 [J].Com puter Science, 2021,48 (11) : 268-275.

[22] Fu Guodong, Huang Jin, Yang Tao, et al. Improved lightweight attention model of CBAM [J].Computer Eng ineering and Application, 2021,57 (20) : 150-156.

[23] Ge Qingqing, Zhang Zhijie, Yuan Long et al. Fusion of environmental features and improved safety helmet wearing detection of YOLOv4 [J].Chinese Journal of Image Graphics, 2021, 26 (12) : 2904-2917.

[24] FENG C, YAN Y, et al. Exploring Separable Attention for Multi-Contrast MR Image Super-Resolution [J].arXiv preprint arXiv, 2109.01664.2022

[25] Zhang Zhenghua, Wu Yu, Jin Zhiqi. Detection of wheat scab infection rate based on MHSA-YOLOv7 [J / OL].Radio engineering : 1-8 [2023-10-16].http://kns.cnki.net/kcms/de tail/13.1097.TN.20230608.1716.004.html

[26] Wang Jianlin, Fu Xuesong, Huang Zhanchao and so on. Improved multi-type cooperative target detection of YOLOv2 convolutional neural network [J].Optical precision engineering, 2020,28 (01) : 251-260.