Universe
Scientific Publishing

Research on real-time detection algorithm of safety helmets in complex operating environment

Tong Xiao[1], Guodong He[1]*, Mingxing Fang [1], Shaoguo Xie[2]

1. School of Physics and Electronic Information, Anhui Normal University, Wuhu 241002, China.

2. School of Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, Anqing 246133, China.

Abstract: In order to solve the problems of low detection accuracy when the background of safety helmets is complex at construction sites, and the safety helmet target is too small to be easily detected, this paper proposes a real-time detection algorithm for safety helmets in complex working environments based on the YOLOv5 framework. An improved YOLOv5 detection algorithm is proposed to address the issues of missing safety helmets and low detection accuracy in the construction environment. Adding an attention mechanism to the YOLOv5 backbone network, adding a detection layer at the neck of the network, and integrating an ASFF module at the neck of the network have better detection performance when facing complex backgrounds and dense helmet detection; The experimental results show that compared to the original YOLOv5 model, the improved average accuracy has increased by 2.4%, reaching 91.3%, effectively improving the detection ability of safety helmets in complex environments.

Keywords:YOLOv5s; CoordAtt; ASFF: henlmet

1. Introduction

Safety helmet as a most common and practical personal protective equipment, can effectively prevent and reduce the external danger to the head injury. But in many construction sites, helmet wearing is easy to be ignored by the construction personnel, thus causing personal injury accidents. However, in most construction sites in China using manual helmet wearing detection, this primitive detection method is not only time-consuming and labor-intensive, and easy to produce errors. The real-time detection of construction workers wearing helmets has important research significance for the safety of construction sites.

Target detection algorithms currently fall into one of two categories [1]: one is based on region proposals and includes algorithms like R-FCN [2], R-CNN [3], Fast-RCNN [4], and Mask RCNN[5]; the other is based on border regression and includes algorithms like SSD[6], YOLOX[7], CenterNet[8], YOLO[9]etc. Numerous academics domestically and overseas have recently enhanced the identification algorithm based on the attributes of the helmet and the working environment. Numerous academics domestically and overseas have recently enhanced the identification algorithm based on the attributes of the helmet and the working environment.

In Sun et al., the anchor frame is improved to strengthen the expression information of the network for small targets, and the improved Faster R-CNN algorithm has a good detection effect on various scene types. This is because it adds a self-attention mechanism to the Faster R-CNN to obtain multi-scale global information, which gives it richer information of high-level semantic features and introduces a wider range of perceptual fields into the model. Shallow output was introduced by Fu et al. to the YOLOv5 model to increase the precision of small target helmet recognition. In , Yingli Wang introduced the lightweight attention force module CBAM in the YOLOv5 structure, and also used the SIoU_Loss loss function to replace the original loss function as a way to improve the convergence of the model. Finally, the precision of helmet wearing detection is improved. In CoordAtt coordinate attention mechanism module is added to the YOLOv5s network's backbone in order to take into account global information and enhance the network's capacity to detect small targets. Res2NetBlock structure substitutes the residual block in the backbone network to enhance feature fusion.

This article proposes an improved YOLOv5 detection algorithm to address the issues of missing safety helmets and low detection accuracy in construction environments. Adding an attention mechanism to the YOLOv5 backbone network, adding a detection layer at the neck of the network, and integrating an ASFF module at the neck of the network have better detection performance when facing complex backgrounds and dense helmet detection; Meanwhile, the improved YOLOv5 algorithm has a faster detection speed compared to other network structures, with FPS only lower than the original algorithm, but still able to meet the real-time detection of safety helmets on construction sites.

2. YOLOv5s

As shown in Figure 1,There are four YOLOv5 network models, namely YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5. Their networks are the same, but the difference is the depth and width of the network.

The YOLOv5s network is mainly composed of three parts, Backbone (backbone network),It consists of Neck (neck) and Head (detection head).

Among them, Backbone (backbone network) is composed of CBS (Conv+BN+SiLU), C3 module and fast spatial pyramid pooling SPPF (Spatial Pyramid Pooling - Fast), which is mainly used to extract image features; Neck is composed of Feature Pyramid FPN (Feature Pyramid Network) + path aggregation network PAN (Perceptual Adversarial Network), which mainly performs feature fusion; Head makes the final prediction of the image.

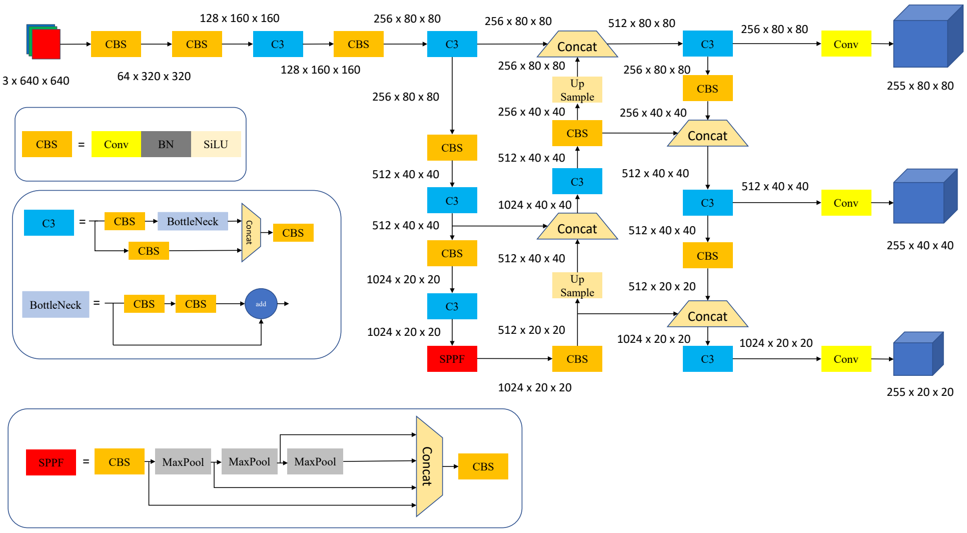


Fig.1 YOLOv5 structure diagram

3. Improvements based on YOLOv5s

3.1 Network structure

The model in this article uses the YOLOv5s network as the basic model, and adds the Coordinate Attention (CA) mechanism to the model Backbone to improve the model's ability to capture the features of the helmet; a detection layer is added to the three-layer detection layer of the model; Improve the accuracy of small target recognition; in the feature fusion module of the network, the Adaptively Spatial Feature Fusion (ASFF) module is introduced to perform spatial adaptation of semantic information and multi-scale feature fusion to improve the characteristics of the network Integration capabilities. The improved network structure is shown in Figure 2.

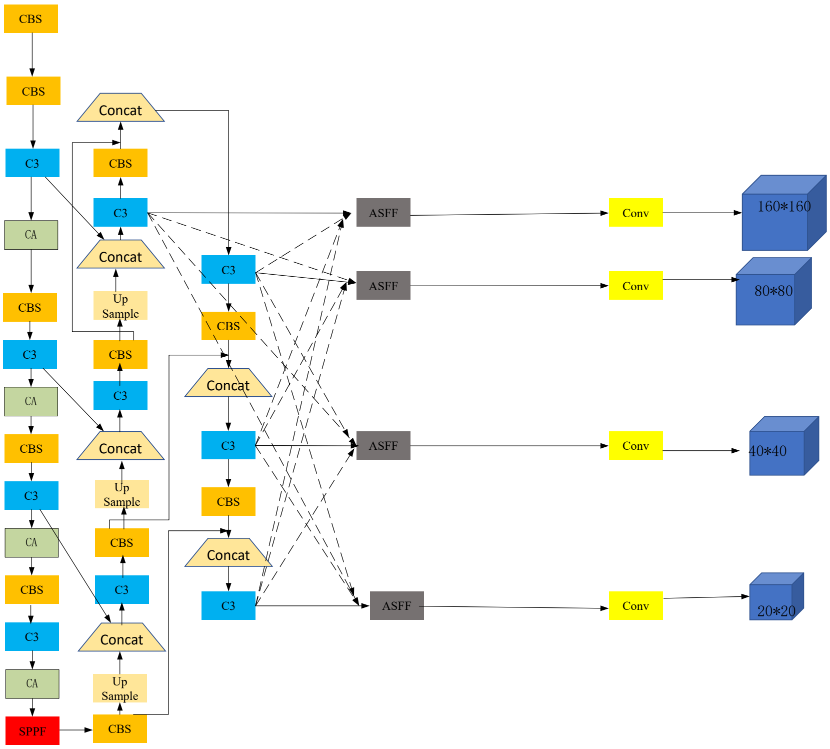


Fig.2 Improved YOLOv5 network structure

3.2 Coordinate Attention

In order for the network to quickly locate areas of interest when paying attention to things, like humans, people usually add an attention mechanism to the network structure to increase attention to the target area. This article introduces the coordinate attention mechanism (CA) into the network structure of YOLOv5 to improve the network's ability to detect targets, enhance the expressiveness of the backbone network, and thereby improve the accuracy of target detection. The structure of CA is shown in Figure 3.

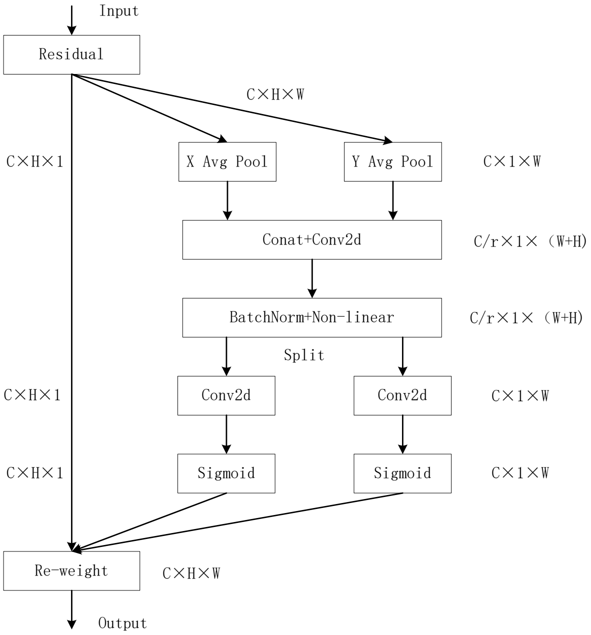


Fig.3 CoordAtt coordinate attention mechanism structure digiagram

The CoordAtt coordinate attention mechanism embeds location information into channel attention, allowing information to be captured on a larger scale. CoordAtt decomposes channel attention into two 1D feature encoding processes that aggregate features along different directions. This has the advantage of capturing long-range dependencies along one spatial direction and retaining precise location information along the other spatial direction. The generated feature maps are then encoded separately to form a pair of orientation-aware and position-sensitive feature maps, which can be complementarily applied to the input feature maps to enhance the representation of the target of interest. Briefly, Coordinate Attention is performed by averaging pooling in horizontal and vertical directions, then transform to encode the spatial information, and finally fusing the spatial information by weighting it over the channels.

The output of the cth channel at height h can be formulated as：

$$Z_c^h(h)=\frac{1}{W}\sum_{0\le i\le W} x_c(h,i) \tag{1}$$

Similarly, the output of the cth channel of width w can be written as：

$$Z_c^h(h)=\frac{1}{H}\sum_{0\le j\le W} x_c(j,w) \tag{2}$$

The above two transformations aggregate features along two spatial directions, respectively, to generate a pair of direction-aware feature mappings. Given the aggregated feature mappings generated by Eqs. (3) and (4), first stitch them together and then pass them to the shared 1 × 1 convolutional transform function F_1 , we obtain:

where [z^h , z^w] denotes the stitching operation along the spatial dimension, σ is a nonlinear activation function, and $f\in R^{C/r\times W}$ is an intermediate feature mapping that encodes the spatial information in the horizontal and vertical directions. Then f is split into two independent tensors along the spatial dimension $f\in R^{C/r\times H}$ and $f\in R^{C/r\times W}$. Then using two 1 × 1 convolutional transforms F_h and F_w are used to convert to f^h and f^w as tensor with the same channels as the input X , respectively, we obtain:

$$g^h=\sigma(F_h(f^h)) \tag{3}$$

$$g^w=\sigma(F_w(f^w)) \tag{4}$$

The outputs g^h and g^w are expanded and used as attention weights, respectively. Finally, the output Y of the coordinate attention mechanism can be expressed as：

$$y_c(i,j)=x_c(i,j)\times g_c^h)(i \times g_c^w j \tag{5}$$

CA not only considers channel information, but also adds coordinate position information, which can locate and identify target areas more quickly. By adding coordinate attention mechanisms in different places in YOLOv5, it is finally found that coordinate attention is added after each C3 module of the backbone network. The mechanism is the best. The changed backbone network is shown in Figure 4

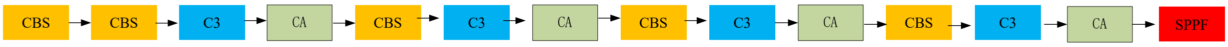


Fig.4 The changed backbone network

3.3 Add detection layer

YOLOv5 target detection algorithm, its backbone network is CSPDarknet53 structure. In the CSPDarknet53 backbone network, a total of 3 effective feature extraction layers are included, which are used to detect larger objects, medium objects and smaller objects. However, these three effective feature extraction layers also have some shortcomings in the problem of helmet wearing detection. When the pixel value of the safety helmet worn by a construction worker on the image is very small or less than 8×8 pixels, the safety helmet will be difficult to detect. The above situation corresponds to the smaller target in the helmet wearing detection. The resolution itself presented on the image is relatively low, and it looks blurry. When its pixel value is small, it is difficult to detect. arrived. This situation does not happen

It only reduces the accuracy of helmet wearing detection and has a greater impact on the detection performance of the algorithm. In order to effectively improve the problem that the hard hats worn by construction workers present small target information on the image, making it difficult to detect the hard hats. Therefore, an effective feature extraction layer is added to the backbone network of YOLOv5, CSPDarknet53, to improve the problem that smaller hard hat targets are difficult to detect. In YOLOv5's backbone network CSPDarknet53, as for its ability to extract feature information, as the network depth in the backbone network increases, the feature extraction ability for large targets will become stronger and stronger, which will be more conducive to the extraction of large targets. detection. Therefore, adding an effective feature extraction layer to the shallow network of the CSPDarknet53 backbone network is more conducive to the detection of smaller target information. The network structure is shown in Figure 5

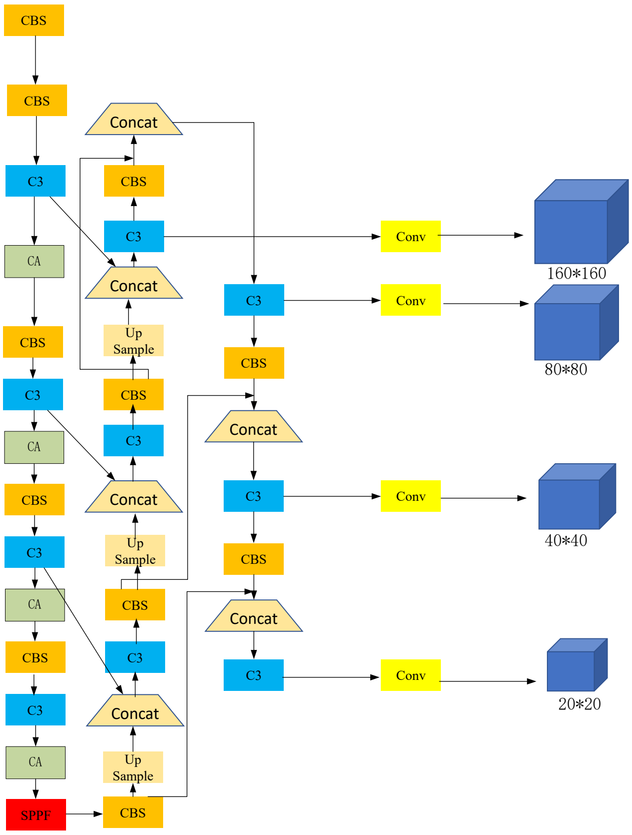


Fig.5 Add feature layer and detection layer

3.4 ASFF

The original YOLOv5 algorithm uses FPN+PAN as the neck feature fusion, integrating high-level semantic information and low-level features at each detection layer, enhancing the global feature extraction capabilities and effectively improving the target detection capabilities. However, FPN+PAN only accumulates different feature layers through upsampling to feature maps of the same size, without taking into account the importance of different feature layers, because the feature information of small target helmets is not evenly distributed in different feature layers, so we A more effective feature fusion algorithm is needed to obtain the weights of different feature layers, thereby improving the detection capabilities of safety helmets.

The ASFF module can adaptively determine the weight of each feature layer, allocate greater weight to the feature layer containing the target, reduce the noise when merging different feature layers, further extract more effective features, and improve the accuracy of target detection. This article introduces the ASFF module based on the YOLOv5 network to increase the network's ability to integrate helmet feature information and reduce the interference of complex backgrounds on targets in real environments. The structural design diagram of ASFF is shown in Figure 6:

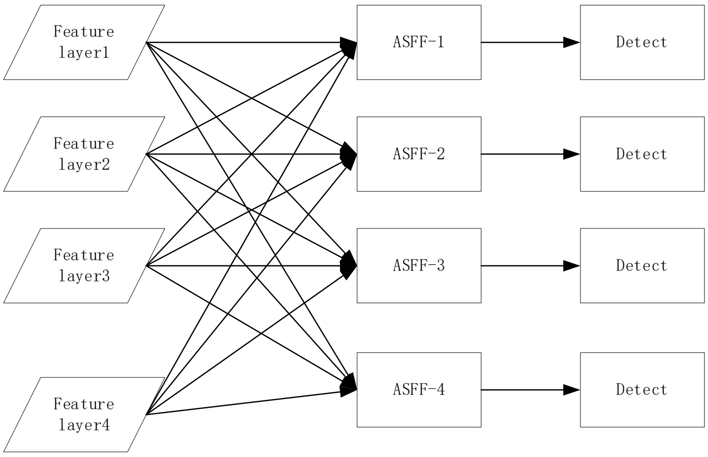


Fig.6 Diagram of ASFF

Among them, PANet in the original model has three output feature layers. Since the improved model PANet adds a detection layer, the four feature layers are recorded as level1, level2, level3, and level4. Take the calculation process of ASFF-4 as an example. First, level1, level2 and level3 are convolved with 1×1 to make the number of channels the same as level4. Then level1, level2 and level3 are adjusted through upsampling to form features of the same size as level4. Figure, recorded as new_level1, new_level2, new_level3; then new_level1, new_level2, new_level3 and resize3 are convolved to obtain the weights a, b, c and d, and then new_level1, new_level2, new_level3 and resize3 are multiplied by a, b, c, and d normalized a′, b′, c′ and d′ are then added to obtain the output feature map of ASFF-4. The formula of ASFF adaptive spatial feature sum is as follows:

$$y_{ij}^{1} = a_{ij}^{1} * X_{\mathrm{ij}}^{1\to1} + b_{ij}^{1} * X_{\mathrm{ij}}^{2\to1} + c_{ij}^{1} * X_{\mathrm{ij}}^{3\to1} + \mathrm{d}_{ij}^{1} * X_{\mathrm{ij}}^{4\to1} \tag{6}$$

Among them, y_{ij}^{1} represents the output feature layer of ASFF; $X_{\mathrm{ij}}^{1\to1}$ 、$X_{\mathrm{ij}}^{2\to1}$ 、$X_{\mathrm{ij}}^{3\to1}$ 、$X_{\mathrm{ij}}^{4\to1}$ is the feature vector from the three different output layers of PAN to this layer; a_{ij}^{1} 、b_{ij}^{1} 、c_{ij}^{1} 、d_{ij}^{1} represents the weight of each layer, $a_{ij}^{1} + b_{ij}^{1} + c_{ij}^{1} + \mathrm{d}_{ij}^{1} = 1$, a_{ij}^{1} 、b_{ij}^{1} 、c_{ij}^{1} and $\mathrm{d}_{ij}^{1} \subset [0,1]$

4. Result and Discussion

4.1 Dataset

For deep learning, the quality of the dataset determines the learning results to a certain extent. The dataset used in this article is based on SHWD. Some of these images are from classroom surveillance and are not suitable as data for hard hat detection. Therefore, this data is deleted and some data from complex construction sites is supplemented. The dataset includes a total of 6,000 images, 5,000 images in the training set, and 1,000 images in the validation set.

4.2 Experimental environment

The model experimental environment parameters are shown in Table 1.

Table1 Experimental environment

Experimental environment	parameter
operating system	Windows 10 Professional
PyTorch version	PyTorch 1.10.2
Python version	Python 3.7
CUDA version	CUDA11.8

GPU	NVIDIA GeForce GTX 3060
CPU	Intel® CoreTM i5-12600KF

4.3 Parameter settings

Table 2 shows some of the parameter settings in this experiment.

Table2 Parameter settings

parameter name	Parameter value
weight decay	0.0005
Bath Size	8
Optimization function	Adam
learning rate	0.01
epoch	100

4.3 Evaluation indicators

In this paper, average precision (AP) and mean average mAP are used as the evaluation indexes of model detection precision based on the open set, and average precision takes into account the precision (P) and recall (R) of target detection, and each evaluation index is calculated as follows:

$$precision = \frac{TP}{FP+TP} \tag{7}$$

$$recall = \frac{TP}{FN+TP} \tag{8}$$

$$AP = \int_0^1 P(R)dR \tag{9}$$

$$mAP = \frac{1}{n}\sum AP \tag{10}$$

4.4 Experimental results and analysis

The comparison test chart of the average accuracy mean curve of the improved YOLOv5 model and the original model with an IoU threshold of 0.5 is shown in Figure 7, where the abscissa is Epoch and the ordinate is map.

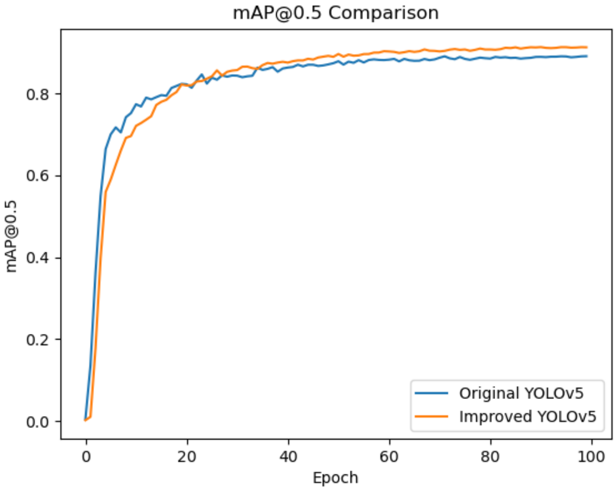


Fig.7 Average precision mean comparison curve

As can be seen from Figure 7, the average accuracy of the improved YOLOv5 model reaches 91.3%, which is better than the 88.9% of the YOLOv5 model, an increase of 2.4%. There was no overfitting phenomenon during the 100 rounds of model training, and the model training was ideal.

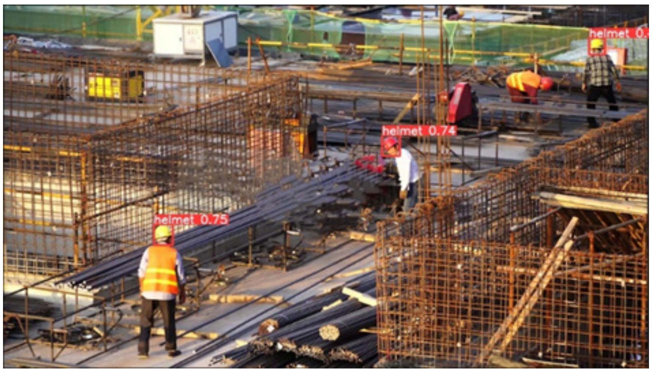
In order to verify the improvement effect of the improved YOLOv5 model in the helmet data set, an ablation test was set up, as shown in Table 3. Compare the improvement effects of the coordinate attention mechanism, adding a detection layer and ASFF on the original model respectively. The added detection layer is represented by ad. For the accuracy of the experiment, all models were trained under the same experimental conditions (training environment, training parameters, data set partitioning), and then the accuracy, average accuracy mean, parameter amount, model size and FPS of the models were compared.

Table3 Ablation test

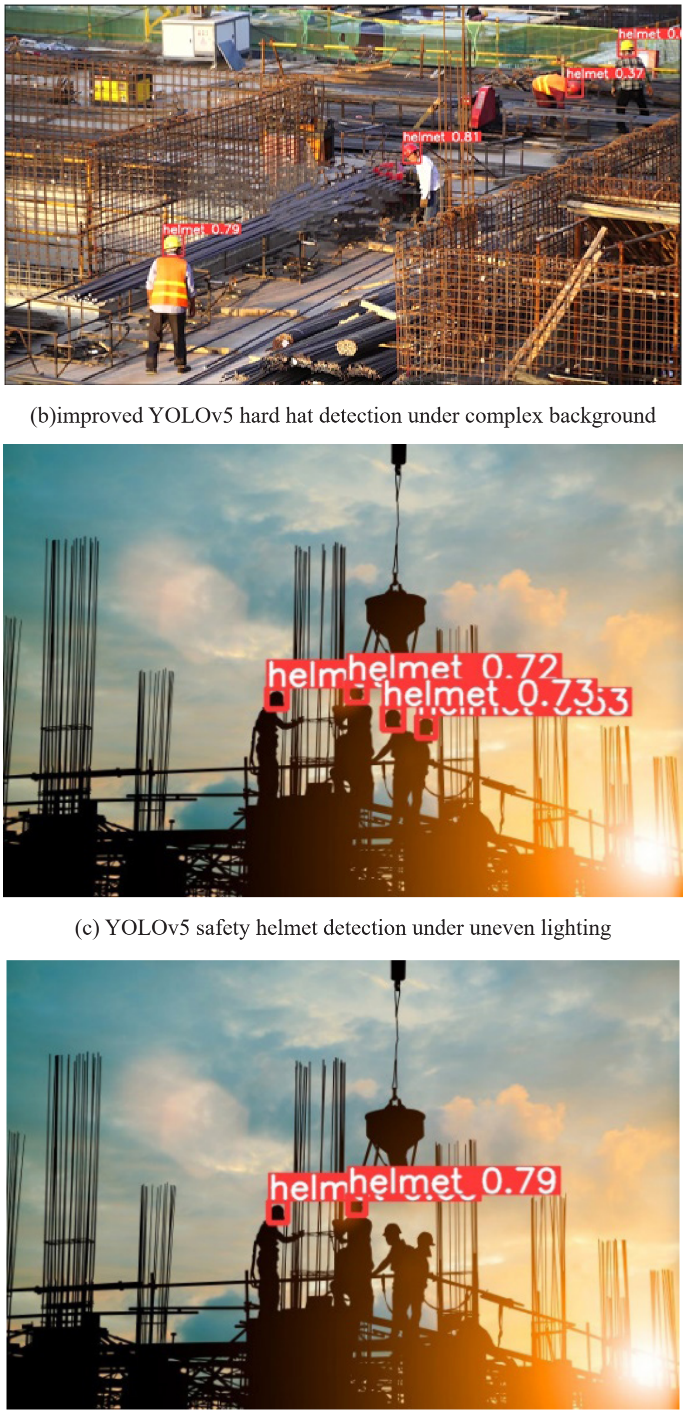
Methods	Precision	mAP	Parameters/106	Modelsize/M	FPS
YOLOv5s	0.898	0.889	7.015	13.8MB	172.5
YOLOv5s+CA	0.895	0.895	7.132	13.9MB	143.8
YOLOv5s+ad	0.901	0.898	7.158	14.7MB	135.6
YOLOv5s+ASFF	0.911	0.903	13.267	24.2MB	123.4
YOLOv5s+CA+ad+ASFF	0.915	0.913	13.267	26.5MB	79.2

From Table 3, it can be seen that all individual improvements have improved the accuracy of the YOLOv5 model on the helmet dataset, but have decreased the detection speed. After each C3 module in the backbone network, a coordinate attention mechanism was added to enhance the network's ability to capture helmet feature information. The average accuracy increased by 0.6%, the model size slightly increased, and FPS decreased by 28.7 compared to the original YOLOv5 algorithm; By adding a small target detection layer, the feature extraction of small targets is more effective. The average accuracy has been improved by 0.9%, and the model size has slightly increased. FPS has decreased by 36.9% compared to the original YOLOv5 algorithm; By adding ASFF to the feature fusion module, different weights were fused for different feature layers to extract more effective features. The average accuracy increased by 1.4%, while FPS decreased by 49.1; Finally, each improvement was integrated into a network structure, resulting in a model with an average accuracy of 91.3. Compared to the original YOLOv5, the average accuracy has been improved by 2.4%.

In order to fully demonstrate the superiority of this algorithm, this article compares the models of YOLOv5 and improved YOLOv5 algorithms. Select some pictures from the data set to display the results. Among them, (a), (c), and (e) in the first column are the detection results of the original YOLOv5 algorithm. (b), (d), (f) in the second column is the detection result of the improved algorithm, as shown in Figure 8.



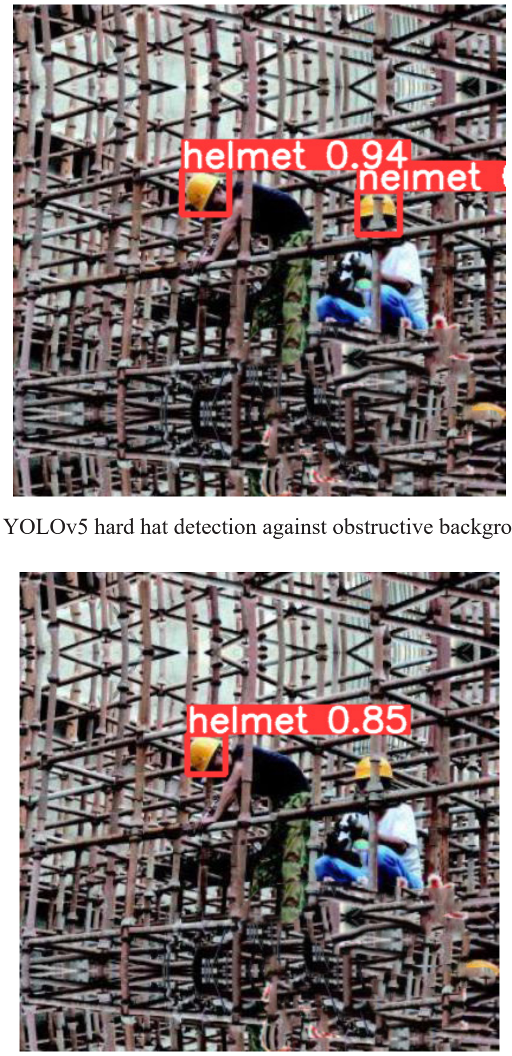
(a) YOLOv5 hard hat detection under complex background



(b)improved YOLOv5 hard hat detection under complex background

(c) YOLOv5 safety helmet detection under uneven lighting

(d) improvedYOLOv5 safety helmet detection under uneven lighting



(e) YOLOv5 hard hat detection against obstructive background

(f)improved YOLOv5 hard hat detection against obstructive background

Fig.8 Comparison of detection performance between YOLOv5 and improved YOLOv5 network models

As shown in Figure 8, the improved algorithm has shown significant improvement in helmet detection. Firstly, for Figure 8 (a), the original detection algorithm did not detect safety helmets in complex backgrounds. The improved algorithm pays more attention to the characteristics of safety helmets and detects them when the safety helmet target is small. The detection results are shown in Figure 8 (b). Secondly, for Figure 8 (c), Under different backgrounds such as light and darkness, the original algorithm failed to detect the person wearing a safety helmet in the image. The improved algorithm can effectively detect it, as shown in Figure 8 (d). Secondly, for Figure 8 (e), when conducting helmet detection in dense environments, the original algorithm did not detect the partially occluded individuals wearing helmets in the figure. The improved algorithm can effectively detect them. At the same time, for each helmet, the confidence of the improved helmet is higher, as shown in Figure 8 (f).

5. Conclusion

This article proposes an improved YOLOv5 detection algorithm to address the issues of missing safety helmets and low detection accuracy in construction environments. Adding an attention mechanism to the YOLOv5 backbone network, adding a detection layer at the neck of the network, and integrating an ASFF module at the neck of the network have better detection performance when facing complex backgrounds and dense helmet detection; Meanwhile, the improved YOLOv5 algorithm has a faster detection speed compared to other network structures, with FPS only lower than the original algorithm, but still able to meet the real-time detection of safety helmets on construction sites.

The improved YOLOv5 algorithm in this article achieves more accurate detection of safety helmets, and will be applied to other unsafe behavior detection in the future, such as detecting seat belts with obscure features, smoking with smaller targets, etc. For mobile devices, reducing the number of model parameters while maintaining model accuracy so that the model can be effectively embedded into the device is also a research direction.

References

[1] LI Z Q, LIU H. Helmet wearing detection algorithm based on deep learning[J]. Computer Applications and Software, 2022, 39(6):194-202.

[2] DAI J F, LI Y, HE K M, SUN J. R-FCN: object detection via region-based fully convolutional networks[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona Spain, 5-12 Dec, 2016. New York: Curran Associates Inc, 2016: 379- 387.

[3] Ming Q L, Yao C H, X B L. Dilated Light-Head R-CNN using tri-center loss for driving behavior recognition[J]. Image and Vision Computing,2019,90(C):108-121.

[4] Li J , Liang X , Shen S , et al. Scale-Aware Fast R-CNN for Pedestrian Detection[J]. IEEE Transactions on Multimedia, 2017,20(4):985-996.

[5] HE K,GKIOXARI G,DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2020,42(2):386-397

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham,2016:21-37.

[7] GE Z，LIU S，WANG F, et al. YOLOX: Exceeding YOLO Series in 2021[J]. 10.48550/arXiv.2107.08430, 2021.

[8] DUAN K，BAI S，XIE L, et al. CenterNet: Keypoint Triplets for Object Detection[C]//Proceedings of the IEEE/ CVF international conference on computer vision. 2019: 6569-657

[9] BOCHKOVSKIY A, WANG C Y, LIAO H-Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J].In IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2020

[10] SUN G D，LI C，ZHANG H.Safety helmet wearing detection method fused with self-attention mechanism[J].Computer Engineering and Applications，2022，58（20）：300-304.

[11] FU D S, GAO L, HU T, et al. Research on safety helmet detection algorithm of power workers based on improved YOLOv5[J]. Journal of Physics(Conference Series),2022, 2171:012006.

[12] WANG Y L. Construction site helmet wearing monitoring algorithm based on YOLOv5[J]. Information Technology and Informatization，2022（7）：33-36.

[13] Huang, Ke and Zhang, Limin. ‘Helmet-wearing Detection with Intelligent Learning Approach’. Journal of Intelligent & Fuzzy Systems, 2023

Declare

The data provided in this manuscript can be used and the manuscript has no conflict of interest

Acknowledgment

This research was supported by Major Projects of Natural Science Foundation of Anhui Province Universities (KJ2019ZD35), Doctoral Scientific Research fund of Anhui Normal University, the Anhui Province Education Department General Project of China (AQKJ2015B005), the Key Projects of Natural Science Research in Colleges and Universities of Anhui Province in China (2022AH051042), the Key Teaching Research Project of Anqing Normal University in China (2020aqnujyxm21), and the Teaching Research Project of Anhui Provincial Department of Education (2021jyxm0882).