YOLOv5s-FCG: An Improved YOLOv5 Method for Inspecting Riders' Helmet Wearing

Pengfei Wang, Hanming Huang*, Mengqi Wang and Bingjun Li

School of Computer Science and Engineering, Guangxi Normal University, Guilin, Guangxi 541004, China

*Corresponding author’s e-mail: huanghm@gxnu.edu.cn

Abstract. Manual inspection of riders’ helmets is time-consuming and labor-intensive, low in efficiency and small in coverage. Aiming at this shortcoming, this paper proposed an improved YOLOv5s-FCG (FourLayers, CBAM attention, GhostBottleneck) helmet wearing detection method based on YOLOv5 (You Only Look Once). Based on the smallest volume of YOLOv5s in YOLOv5 series, the network was improved, the shallow feature detection layer was added, the three-scale feature detection was changed to four-scale feature detection, and the up sampling was increased by four times. Add the CBAM attention module; Use lightweight GhostBottleneck instead of Bottleneck structures. The results in our experiments show that YOLOv5s-FCG raises the average detection accuracy (mAP) by 2.0% compared with YOLOv5s on the ourselves’ riding safety helmet data set and 1.5% on the NWPU-VHR-10 public data set. The proposed algorithm not only ensured the detection rate, volume, computation and number of parameters, but also improved the detection accuracy. And it had good adaptability and generalization ability in complex road environments such as poor light and small targets.

1. Introduction

According to traffic casualties statistics, motorcycles and electric bicycles are among the most deadly vehicles in traffic accidents. About 80% of the deaths of riders are caused by head injury. Wearing safety helmets correctly can reduce the risk of death in related accidents by 60% to 70%. In order to increase the wearing rate of safety helmets and reduce the degree of casualties in traffic accidents, the Traffic Management Bureau of the China Ministry of Public Security stipulates that from June 1, 2020, strictly investigate and deal with traffic violations that do not wear safety helmets in accordance with regulations. At present, the detection of riders' helmet wearing is still dominated by traffic police inspections. The face-to-face inspection method is time-consuming, laborious, inefficient, and small-coverage. In recent years, with the rapid development of computer vision, it has gradually become a reality to replace people with intelligent machines and equipment to complete detection tasks.

Machine learning and image processing technology have been applied to the task of target detection by many domestic and foreign researchers. In the task of helmet wearing detection for riders, foreign researchers have conducted researches on motorcycle helmet detection, but none of domestic papers applied to helmet wearing detection for motorcycles and electric bicycle riders, only being found that the construction site helmet wearing detection in similar sites: Rattapoom Waranusast[1], Romuere Silva[2], Romuere Rodrigues Veloso e Silva[3], Kunal Dahiya[4], Liu Xiaohui[5], Jia Junsu[6], Li Qirui[7], etc. They used traditional methods such as directional gradient histogram, Hough transform descriptors, and local binary patterns to extract features, and then used a classifier to classify the...
extracted features, and achieved certain results in the detection of helmets and safety helmets, but these methods based on traditional feature extraction rely heavily on the designer's experience, the detection steps are cumbersome, the stability is insufficient, and the generalization ability is poor. With the rapid development of deep learning, there are researchers (Madhuchhanda Dasgupta\cite{8}, Xu Shoukun\cite{9}, Wang Bing\cite{10}, etc.) who used CNN (convolutional neural network) -based target detection algorithms to detect the wearing of helmets and safety helmets. The helmet detection experiment of Linu Shine\cite{11} showed that being compared with the target detection method based on traditional feature extraction methods, the detection speed of the CNN-based target detection algorithm is greatly improved in the detection accuracy although with slightly slower detecting speed.

CNN-based target detection algorithms mainly have two categories: two-stage and one-stage. The two-stage is a classification-based detection algorithm\cite{12}. The candidate area that may contain the target to be detected is first segmented, and then the segmented candidate area is classified and performed the position calibration, finally obtains the detection result. The representative algorithms mainly include RCNN\cite{13}, Fast-RCNN\cite{14}, Faster-RCNN\cite{15}; the one-stage is a regression-based target detection algorithm\cite{12}, which is end-to-end. The idea is that there is no need to segment candidate regions. The acquisition of anchors and the prediction of categories and positions are directly completed in the same CNN, and the final detection results are obtained from the original image. The representative algorithms mainly include YOLO\cite{16}, SSD\cite{17}, YOLOv2\cite{18}, YOLOv3\cite{19}, YOLOv4\cite{20}, YOLOv5. Before YOLOv4, the accuracy of the two-stage target detection algorithm was higher, but the real-time performance was inferior. The one-stage target detection algorithm had a great improvement in speed, but the detection accuracy was poor. However after being proposed of YOLOv4 and the following YOLOv5, the one-stage target detection algorithm greatly surpasses the two-stage target detection algorithm by detection accuracy and detection speed.

When the scale of the target to be detected is relatively small, the detection accuracy of the detector based on the anchor frame will drop sharply\cite{21}. The detection targets are mainly small in riders’ helmet wearing detection, inevitably incurring poor detection result. At present, small target detection still has problems such as few usable features, high positioning accuracy requirements, and easy aggregation of small targets, which makes it difficult to improve detection accuracy\cite{22}, and the helmet-wearing detection of motorcycles and electric bicycles are more complex than that of construction sites, with more very small targets in the distance. Moreover, it is difficult to locate the targets during riding and it is easier for people's heads to gather in the images, thus these bring greater challenges to the detection task. In order to ensure the speed and accuracy of detection, and to control the volume of the network model as small as possible, basing on YOLOv5s and its improved version, due to the shallow detection layer can detect smaller target, a layer of shallow feature detection layer is added, so the previous three-scale feature fusion is changed to four-scale; then the upsampling is increased by 4 times, thus more features are available. In order to further improve the detection accuracy and positioning accuracy and reducing small target aggregation, the CBAM\cite{23} attention module is embedded. The effects of the CBAM attention module embedded in the backbone, neck and detection head of YOLOv5s on the performance of the algorithm are explored. In order to further reduce the weight of the network while ensuring the detection accuracy, experimental exploration the effect of replacing the Bottleneck module of the network backbone part and neck part with GhostBottleneck module on network performance. The contribution of this paper is to propose a helmet-wearing detection method for riders based on improved YOLOv5s, YOLOv5s-FCG. Training and testing were carried out on the ourselves-made riding safety helmet dataset and the public remote sensing dataset NWPU-VHR-10 respectively, and compared with YOLOv4, YOLOv4-Tiny, SSD, YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x in the same environment. The experiments’ results show that the YOLOv5s-FCG algorithm is feasible and applicable.
2. Methods and Principles

2.1. YOLOv5

YOLOv5 was a target detection algorithm based on regression. At present, the author had only published the source code, but no corresponding paper had been published. The address of the source code was https://github.com/ultralytics/yolov5. This algorithm had brought together many advantages of deep learning target detection framework, and the network structure was shown in Figure 1. It can be known that it was mainly divided into four parts: Input end, Backbone, Neck and Output end.

Figure 1. YOLOv5 network structure
2.2. YOLOv5 improvement

2.2.1. Improved multi-scale detection. YOLOv5 experiment and test was mainly on the public data sets, COCO and VOC, most of these targets to be detected in the data was bigger, and this article targets to be detected for riding personnel safety helmet, detection area was mainly in the person's head position, were relatively small, the characteristics of image were extracted, multiple convolution operation would be lost a smaller target area information, As a result, the overall detection accuracy was reduced. In order to solve this problem, this paper, referring to the method in reference [25], proposed YOLOv5s-FourLayers (YOLOv5s-F) network to improve the multi-scale detection of YOLOv5s by adding a shallow detection layer. Under the premise of input size of 640×640, A 160×160 detection layer was added, and the original three-scale detection was changed into four-scale detection, so as to enhance the learning ability of the network for small targets. Moreover, in order to make full use of deep and shallow layers of information, realize feature fusion between different layers and further reduce the loss of feature information in the process of convolution, the original network of 2 times up sampling was changed into a combination of 2 times up sampling and 4 times up sampling to improve the detection accuracy of safety helmet. The YOLOv5s-F network structure was shown in Figure 2.

![Figure 2. YOLOv5s-F network structure](image)

2.2.2. Embedded design of CBAM attention module. The Convolutional Block Attention Module (CBAM) contained two independent submodules, namely, the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), which were responsible for attention on channels and space respectively. The structure was shown in Figure 3.
The CBAM attention module can enhance the important channel and spatial features in the feature map, that was, learn the "key points" of the target, thereby effectively improving the accuracy of target positioning in small target detection and reducing the problem of small target aggregation. Based on the YOLOv5s-F network proposed above, this paper explored the influence of CBAM attention module embedded in the Backbone, Neck and Head of the network on the network performance respectively through experiments. Since the CBAM attention module was to enhance important features, this paper embedded the CBAM attention module into the network structure after every feature fusion, that was, after “add”, “concat” and before the detection head, so as to generate four new network models: YOLOv5s-FC-A, YOLOv5s-FC-B, YOLOv5s-FC-C, and YOLOv5s-FC-D, as shown in Figure 4.

![Convolutional Block Attention Module](image)

**Figure 3. CBAM module network structure**

![CBAM module network structure](image)

**Figure 4. CBAM module network structure**

(a) Modified region of YOLOv5s-FC-A algorithm (b) Modified region of YOLOv5s-FC-B algorithm

(c) Modified region of YOLOv5s-FC-C algorithm (d) Modified region of YOLOv5s-FC-D algorithm

**Figure 4. Four kinds of YOLOv5s-FC models embedded in CBAM modules**

(In the figure, "⊕" represents the add operation, and "→" represents the conversion symbol)
In Figure 4 (a), after integrating CBAM into “add” and “concat” inside each C3 structure in the backbone, (b) after integrating CBAM into “add” and “concat” inside each C3 structure in the neck, and (c) after adding CBAM attention layer after “concat” in the neck, (d) Figure was to add a CBAM attention layer in front of each detection head.

### 2.2.3. GhostBottleneck lightweight

The core idea of GhostNet\[26\] was to design a phased convolution calculation module. On the basis of the feature graphs obtained from a small amount of nonlinear convolution (convolution and batch normalization and nonlinear activation), a linear convolution (general convolution) was carried out again to obtain more feature graphs. And the new feature map was called before the feature map "ghost", so as to achieve the elimination of redundant features, access to a more lightweight model.

GhostBottleneck was to use Ghost module to replace the existing Bottleneck modules in the network, which made the calculation amount was greatly reduced and the model volume was greatly reduced. In this paper, we used GhostBottleneck to replace the improved YOLOv5s-FC network backbone and neck congestion respectively, and explored the influence of different location replacement on algorithm performance, so we got two new network models: The network structures of YOLOv5s-FCG-A and YOLOv5s-FCG-B were shown in Figure 5.

Figure 5 shows (a) where Ghostbottleneck replaced every C3 structure with Bottleneck at the backbone and (b) where Ghostbottleneck replaced every C3 structure with Bottleneck at the neck.

### 3. Results & Discussion

#### 3.1. Experiment environment

Experimental environment of this paper: Intel(R) Core(TM) i7-10700KF CPU @ 3.80GHz 3.79GHz, 32G RAM, GPU NVIDIA GeForce RTX 3080, 10G video memory, Windows 10, 64-bit operating system, PyTorch deep learning framework with Python programming language and GPU acceleration software CUDA11.1 and CUDNN8.0.5.

#### 3.2. data set

The experimental data set consisted of three parts: one part was the pictures of "bicycle" and "motorcycle" extracted from the public data set MS COCO, one part was the pictures crawled from Baidu and Google, and the other part was other data sets downloaded from the Internet, a total of 10837 pictures. Labeling software was used for marking, and there were two kinds of labels: helmet wearing (helmet) and nohelmet wearing (nohelmet). In order to enlarge the data and improve the robustness of the data set, 1600 pictures were flipped, 1600 pictures were blurred, and 1600 pictures...
were reduced in brightness. The detection difficulty was increased but the data set was more in line with the practical application scenarios. The ratio of training set, verification set and test set was 8:1:1.

![Image](image.png)

(a) Before horizontal flipping  (b) After flipping horizontally  (c) Before blurring

(d) After blurring  (e) Before reducing brightness  (f) After reducing the brightness

Figure 6. Data pre-processing

| Data Set       | ImageNum |
|----------------|----------|
| Training set   | 8671     |
| Validation set | 1083     |
| Test set       | 1083     |
| Total          | 10837    |

Table 1. DataSet.

Figure 6 was an example of data preprocessing, and Table 1 was the specific division of the data set.

3.3. Network training and testing

In this paper, Precision (P), Recall (R), mean Average Precision (mAP) of multiple categories, Floating Point Operations (FLOPS) and Frames Per Second (FPS) were used to evaluate the network performance. Detection accuracy P, recall rate R and mAP were respectively expressed as:

\[
P = \frac{TP}{TP + FP} \quad (1)
\]

\[
R = \frac{TP}{TP + FN} \quad (2)
\]

\[
mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \quad (3)
\]
In the above formula, TP (True Positives) refers to the number of correctly detected targets, FP (False Positives) refers to the number of wrongly detected targets, FN (False Negatives) refers to the number of undetected targets, and $n$ refers to the number of categories that need to be classified. AP (Average Precision) represented the Average Precision of a target class. The experimental setting basically adopted the official recommended parameter setting of YOLOv5, which adopted adaptive anchor and Mosaic data enhancement. The size of the input image was $640 \times 640$, the batchsize was 16, the epoch was 300, the initial learning rate was 0.01, and the learning rate momentum was 0.937.

3.4. Results and analysis

3.4.1. Comparison results of improved multi-scale detection. The experiment first compared the changes in the detection performance of the original YOLOv5s network and the improved YOLOv5s-F network on the data set of riders' safety helmets in this paper. The specific results were shown in Table 2 and Figure 7.

| Network model | P% | R% | mAP@0.5/% |
|---------------|----|----|-----------|
| YOLOv5s       | 94.0 | 91.1 | 94.7      |
| YOLOv5s-F     | 94.0 | 92.5 | 96.0      |

Table 2. Comparison results of improved multi-scale detection

![Figure 7. Changes in the training process mAP](image)

Table 2 compared and analyzed the impact of improved multi-scale detection on the performance of the network model from the three indexes of P, R and mAP@0.5 (a fixed crossover ratio of IOU threshold value 0.5 is used to calculate AP value). As can be seen from the comparison results in Table 2, the improved four-scale detection can significantly improve the detection accuracy of the network, mAP@0.5 increased by 1.4%.

3.4.2. Embedding experiment of CBAM attention module. In this experiment, YOLOv5s-F proposed in Section 2.1 of this paper was used as the baseline. The four network models proposed in Section 2.2, YOLOv5s-FC-A, YOLOv5s-FC-B, YOLOv5s-FC-C, and YOLOv5s-FC-D, were used for training on the training set and verification set of safety helmet data set. The performance of the model was compared and evaluated on the test set. The experimental results were shown in Table 3.
Table 3. Performance comparison of different network models on the test set

| Network model    | Volume /MB | Parameters /ten thousand | Operand /GFLOPS | mAP@0.5/% | FPS |
|------------------|------------|--------------------------|-----------------|-----------|-----|
| YOLOv5s-F        | 22.5       | 1102                     | 40.3            | 96.0      | 286 |
| YOLOv5s-FC-A     | 22.7       | 1108                     | 40.5            | 95.7      | 243 |
| YOLOv5s-FC-B     | 22.8       | 1116                     | 40.6            | 96.4      | 213 |
| YOLOv5s-FC-C     | 22.9       | 1119                     | 40.7            | 96.0      | 213 |
| YOLOv5s-FC-D     | 22.7       | 1112                     | 40.5            | 96.6      | 222 |
| YOLOv5s-FC(B+D)  | 23.0       | 1125                     | 40.8            | 96.8      | 196 |

Table 3 showed the comparison of the performance results of the newly proposed different network models on the test set after training. It can be seen that the performance of the network can be effectively improved after integrating the CBAM attention module into "add" and "concat" in the C3_F structure of the neck and adding the CBAM attention layer before the detection head. mAP@0.5 was 0.4% and 0.6% higher than YOLOv5s-F, respectively; When the CBAM attention module was incorporated into the backbone, mAP@0.5 had a %0.3 decrease. We believed that there would be different results when the CBAM attention module was embedded in different parts of the network, because the main feature extraction effect of the backbone part of YOLOv5s was already very good, and the integration of attention module would bias the "key points" of model learning. The neck of the network belonged to the enhanced feature extraction network. The addition of CBAM attention module can further enhance the feature extraction, reduce the loss of information, and alleviate the problem of small target aggregation. The head had rich semantic information, and the important channel and spatial features can be effectively utilized by CBAM attention to further highlight the "key points" and improve the positioning accuracy, so as to effectively improve the detection accuracy of the network.

By comparing the performance of YOLOv5s-FC-A, YOLOv5s-FC-B, YOLOv5s-FC-C and YOLOv5s-FC-D network models, we combined YOLOv5s-FC-B and YOLOv5s-FC-D, which can improve the network performance. A YOLOv5s-FC network model was proposed. mAP@0.5 improved by 0.8% compared with YOLOv5s-F and 2.1% compared with the original YOLOv5s under the condition of guaranteed real-time performance.

3.4.3 GhostBottleneck lightweight tests. In this experiment, the two models in section 2.2.3 of this article and the combination of the two models were trained on the training set and validation set of the safety helmet data set, and the test was compared on the test set. The results were shown in table 4.

| Network model    | Volume /MB | Parameters /ten thousand | Operand /GFLOPS | mAP@0.5/% | FPS |
|------------------|------------|--------------------------|-----------------|-----------|-----|
| YOLOv5s-FC       | 23.0       | 1125                     | 40.8            | 96.8      | 196 |
| YOLOv5s-FCG-A    | 20.7       | 1005                     | 36.9            | 96.5      | 200 |
| YOLOv5s-FCG-B    | 19.0       | 922                      | 30.6            | 96.7      | 196 |
| YOLOv5s-FCG-(A+B)| 16.6       | 802                      | 26.8            | 96.5      | 204 |
Table 4 told us that Ghostbottleneck structure for different parts of the network can effectively reduce the number of parameters and computation, and reduce the volume of the model with the same accuracy. YOLOv5s-FCG-B, by contrast, can achieve considerable lightweight without sacrificing FPS and only reducing %0.1mAP. Therefore, we replaced the Bottleneck structure of YOLOv5s-FCG network with YOLOv5s-FCG-B, which was our final proposed YOLOv5s-FCG network model. The network structure was shown in Figure 8.

Figure 8. YOLOv5s-FCG network structure

3.4.4. Performance test and comparison of YOLOv5s-FCG network. In this experiment, the YOLOv5s-FCG network proposed in this paper was compared with the original YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, SSD, YOLOv4 and YOLOv4-tiny detection network on the safety helmet data set at first. The results were shown in Table 5.

| Network model | Volume /MB | P%   | R%   | mAP@0.5/% |
|---------------|------------|------|------|-----------|
| YOLOv5s       | 14.4       | 94.0 | 91.1 | 94.7      |
| YOLOv5m       | 42.5       | 94.7 | 92.1 | 95.3      |
| YOLOv5l       | 93.7       | 94.0 | 92.5 | 95.9      |
| YOLOv5x       | 175.1      | 93.8 | 94.3 | 96.4      |
| SSD           | 100.0      | 93.1 | 74.2 | 80.6      |
| YOLOv4        | 244.4      | 94.5 | 93.0 | 95.7      |
| YOLOv4-Tiny   | 22.5       | 94.1 | 75.7 | 82.8      |
| YOLOv5s-FCG   | 19.0       | 94.0 | 94.4 | 96.7      |

By comparing the test results of different detection networks in Table 5, the YOLOv5s-FCG network model proposed in this paper showed the best detection performance, which proved the effectiveness of the algorithm presented in this paper.
Figure 9 showed the comparison of partial detection effects of YOLOv5s before improvement and YOLOv5s-FCG after improvement. Through comparison, it can be found that the YOLOv5s-FCG proposed in this paper had been improved in both false detection and error detection, and the detection accuracy of small targets had also been improved. At the same time, this experiment also verified the effectiveness of the algorithm proposed in this paper on other data sets. We selected NWPU-VHR-10, an open remote sensing data set with many small targets, and divided the image data into 6:2:2 according to the partition method in literature\(^{[27]}\). The experimental model proposed in this paper was trained on the training set and verification set of this data set, and the test comparison was made on the test set. The results were shown in Table 6.

| Network model               | mAP@0.5/\% | FPS |
|-----------------------------|------------|-----|
| YOLOv5s                     | 91.8       | 278 |
| YOLOv5s-F                   | 92.3       | 200 |
| YOLOv5s-FC-A                | 91.5       | 189 |
| YOLOv5s-FC-B                | 92.9       | 175 |
| YOLOv5s-FC-C                | 92.2       | 164 |
| YOLOv5s-FC-D                | 92.5       | 179 |
| YOLOv5s-FC                 | 93.2       | 159 |
| YOLOv5s-FCG-A              | 90.9       | 159 |
| YOLOv5s-FCG-B(ours)         | 93.3       | 161 |
| YOLOv5s-FCG(A+B)            | 90.7       | 161 |

As shown in Table 6, on the NWPU-VHR-10 data set, the algorithm proposed in this paper improved by 1.5% compared with the original YOLOv5 network, and by 2.8% compared with the results in literature\(^{[27]}\), on the premise of guaranteeing the real-time performance of the network, showing a certain generalization.
4. Conclusions
The riders’ helmet wearing detection was a kind of small targets detection, and the usable features of small targets detection were obvious fewer. Firstly, in order to improve detection precision, this paper improved the multi-scale detection: increased a shallow detection layer, increased 4 times on sampling, increased the available features, enhanced the deep and shallow features fusion, improved the detection accuracy effectively. Then, in order to solve the problems of higher positioning accuracy requirement and small targets clustering in small targets detection, basing on the improved YOLOv5s-F network, attempted to embed CBAM attention module into different positions in the the network were carried out in several experiments. It had been found that while the CBAM module was embedded only after the "add" and "concat" of the C3 structure in the neck of YOLOv5s-F network and before the detection head, the positioning accuracy could be improved and the clustering of small targets were reduced, resulting the improvement of the detection accuracy. Finally, the GhostBottleneck was used to replace the Bottleneck in the improved YOLOv5s-FC network’s backbone and neck parts for network performance evaluation, the results showed that replacing neck Bottleneck by GhostBottleneck could effectively reduce the volume of the model, and reduce the number of parameters and computation. Several experiments were carried out on the proposed YOLOv5s-FCG network detection model. The results of these experiments showed that the proposed network model could significantly improve the detection accuracy of small target data sets. On the ourselves’ data set of rider safety helmet, mAP@0.5 had improved by 2.0% while compared with YOLOv5s and for the open remote sensing data set NWPU-VHR-10, mAP@0.5 increased by 1.5%.

Although the proposed YOLOv5s-FCG network model improved the detection accuracy and well controlled the model's volume, calculation amount and number of parameters, however, the detection speed suffered decrease. In further study, we will consider other lightweight methods to improve our YOLOv5s-FCG and further improve the detection speed of the network. We will also consider to apply the algorithm to devices such as raspberry and drones applications.

Acknowledgments
Financially supported by Key R & D plan in Guangxi Zhuang Autonomous Region on contract No. AB18126045.

References
[1] R. Waranusast, N. Bundon, V. Timtong, C. Tangnoi and P. Pattananathaburt, "Machine vision techniques for motorcycle safety helmet detection," 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), 2013, pp. 35-40, doi: 10.1109/IVCNZ.2013.6726989.
[2] R. Silva, K. Aires, T. Santos, K. Abdala, R. Veras and A. Soares, "Automatic detection of motorcyclists without helmet," 2013 XXXIX Latin American Computing Conference (CLEI), 2013, pp. 1-7, doi: 10.1109/CLEI2013.6670613.
[3] R. R. V. e. Silva, K. R. T. Aires and R. d. M. S. Veras, "Helmet Detection on Motorcyclists Using Image Descriptors and Classifiers," 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, 2014, pp. 141-148, doi: 10.1109/SIBGRAPI.2014.28.
[4] K. Dahiya, D. Singh and C. K. Mohan, "Automatic detection of bike-riders without helmet using surveillance videos in real-time," 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 3046-3051, doi: 10.1109/IJCNN.2016.7727586.
[5] Liu Xiaohui, Ye Xining. Application of Skin Color Detection and HU Moment in Safety Hat Recognition [J]. Journal of East China University of Science and Technology (Natural Science Edition),2014,40(03):365-370.
[6] Jia Junsu, Bao Qingjie, Tang Huiming. Safety Helmet Wear Detection Based on Deformable Part Model [J]. Application Research of Computers,2016,33(03):953-956.
[7] LI Qirui. Research and Implementation of Safety Hat Video Detection System Based on Human Body Recognition [D]. University of Electronic Science and Technology of China, 2017.

[8] M. Dasgupta, O. Bandyopadhyay and S. Chatterji, "Automated Helmet Detection for Multiple Motorcycle Riders using CNN," 2019 IEEE Conference on Information and Communication Technology, 2019, pp. 1-4, doi: 10.1109/CICT48419.2019.9066191.

[9] XU Shoukun, WANG Yaru, GU Yuwan, et al. Research on Safety Hat Wear Detection Based on Improved Faster R-CNN [J]. Application Research of Computers, 2020, 37(03): 901-905.

[10] Wang Bing, Li Wenjing, Tang Huan. Improved YOLO V3 algorithm and its application in safety helmet detection [J]. Computer Engineering and Applications, 2020, 56(09): 33-40.

[11] Shine, L., C. V., J. Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN. Multimed Tools Appl 79, 14179–14199 (2020). https://doi.org/10.1007/s11042-020-08627-w

[12] Cao Yan, Li Huan, Wang Tianbao. Target detection calculation based on deep learning A review of method research [J]. Computer and Modernization, 2020(05): 63-69.

[13] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[14] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[15] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. arXiv preprint arXiv:1506.01497, 2015.

[16] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[17] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[18] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[19] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[20] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[21] Huang J, Rathod V, Sun C, et al. Speed/accuracy trade-offs for modern convolutional object detectors[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[22] Gao Xinbo, Mo Mengjingcheng, Wang Haitao, Leng Jiaxu. Research progress of small target detection [J]. Data Acquisition and Processing, 201,36(03): 391-417.

[23] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[24] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[25] XIAO Ti-gang, CAI Le-cai, GAO Xiang, et al. Improve YOLOv3 helmet wearing test [J/O/L]. Computer engineering and application: 1-13 [2021-06-14]. http://kns.cnki.net/kcms/detail/11.2127.tp.20210207.1519.014.html.

[26] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1580-1589.
[27] LI Yufeng, GU Manxuan, ZHAO Liang. Faster R-CNN for target detection in remote sensing images [J]. Signal Processing, 20, 36(08): 1363-1373.