

## *Retraction*

# **Retracted: Helmet Wearing Recognition of Construction Workers Using Convolutional Neural Network**

### **Wireless Communications and Mobile Computing**

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] H. Peng and Z. Zhang, "Helmet Wearing Recognition of Construction Workers Using Convolutional Neural Network," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4739897, 8 pages, 2022.

## Research Article

# Helmet Wearing Recognition of Construction Workers Using Convolutional Neural Network

Hong Peng<sup>1</sup> and Zhenwei Zhang<sup>2</sup>

<sup>1</sup>School of Computing, Huanggang Normal University, Huanggang 438000, China

<sup>2</sup>School of Physics and Telecommunications, Huanggang Normal University, Huanggang 438000, China

Correspondence should be addressed to Zhenwei Zhang; zhangzhenwei@hgnu.edu.cn

Received 9 January 2022; Revised 15 March 2022; Accepted 16 March 2022; Published 14 April 2022

Academic Editor: Xin Ning

Copyright © 2022 Hong Peng and Zhenwei Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the most effective ways to prevent construction workers from suffering a head injury is to wear a safety helmet. The use of a computer vision method to detect whether or not construction workers are wearing helmets can improve external construction worker supervision, reducing the number of head injury accidents. A helmet-wearing recognition method based on head recognition is proposed using CNN (convolutional neural network). The area of the head of constructors can be accurately located using crossvalidation of facial feature recognition and head recognition, which solves the problem of determining constructor head position under complex posture. The I-YOLOv3 target detection network is used to detect helmet wear, and the relative position relationship between the helmet and the human body is determined. The findings show that the helmet wearing identification system can be successfully applied to helmet wearing identification work in a complex environment construction site, and it offers a new research perspective and technical method for construction industry information-based safety monitoring.

## 1. Introduction

With the development of the construction industry, the scale of enterprises, the number of related employees, and the total economic output value have all increased to varying degrees. However, behind the prosperity, frequent construction accidents due to negligence pose a great threat to workers' safety [1]. Because China is such a large country when it comes to infrastructure development, construction-related accidents are always high. Every worker must wear a helmet as a vital safety precaution. However, because violations of not wearing a helmet while operating a vehicle occur from time to time, it is necessary to check whether or not to wear a helmet.

In recent years, the successful application of DL (deep learning), a research hotspot in machine field, in speech recognition, and image processing, has made it a new direction in machine learning [2–4]. Therefore, using artificial intelligence instead of traditional artificial video surveillance makes intelligent safety monitoring a new idea for construction workers' safety production. Traditional target detection algorithm [5]

based on machine learning sets feature extraction methods according to local conditions for different detection tasks and uses training classifiers to detect targets. Although this method is designed for feature extraction, it has limited generalization and is difficult to adapt to changes in lighting and background conditions. Due to the characteristics of CNN (convolutional neural network) [6, 7] and self-learning, the target detection algorithm based on DL is robust to illumination and background changes and performs very well in the field of target detection [8]. Using intelligent methods instead of traditional manual monitoring allows for real-time on-site monitoring, which saves labor costs while also improving on-site safety and lays the groundwork for China's development of "smart construction sites" [9].

The existing helmet wearing identification method is time-consuming and labor-intensive, and the demand for workers' safety supervision cannot be met. Compared with traditional full-time safety supervision, automatic identification method can realize real-time supervision and improve on-site safety, which is fast and convenient, and saves manpower, cost, and

time. This paper analyzes the research status and limitations of helmet wearing identification methods in the past, puts forward a semisupervised helmet wearing identification [10] algorithm based on I-YOLOv3, which is suitable for the complex environment of the construction site, and designs a helmet wearing identification system, which can effectively provide safety monitoring and early warning for construction workers without helmets.

## 2. Related Work

Under the background of Industry 4.0, it is inevitable to wear smart video surveillance helmet [11]. Compared with this article, some scholars have done research on the direction of video detection in recent years [12], but most of the traditional methods are used, the efficiency and stability need to be improved, and the research of tracking module is rarely involved.

Literature [13] identifies the face area using detailed information such as workers' facial features and skin color and compares different classifiers based on the color difference between hair and helmet to determine whether workers wear helmets. To identify helmets, literature [14] uses the entire head range and edge features. The combination of the foreground of a Gaussian mixture model and the human body edge is used in literature [15] to identify helmets. Although this method does not require attention to the color of the helmet, it does require the basic condition of collecting workers' facial features as samples, making it less stable than the color feature recognition method. To achieve the goal of identifying the target, the feature recognition method based on human body shape proposed in document [16] first creates a template by detecting the entire human body shape, then compares and matches the moving detection frame in the multilevel template, and finally selects the part that best matches the template by comparing the maximum likelihood ratio. A set of helmet wearing recognition algorithms with mean drift has been developed in the literature [17]. The main method is to use background difference to create a weighted time mean model, which is then combined with the best separation hyperplane model to fully realize the process from detection to tracking. Literature [18] combines Bluetooth technology with sensor network to identify construction workers in indoor environment and analyze the location information of indoor personnel. However, there are disadvantages. Bluetooth devices need to be charged regularly after being used for a period of time, and its applicability and applicability are poor; so, it is impossible to use this technology for a long time. Literature [19] locates the face area through the comprehensive information of facial features and skin color of construction workers and judges whether construction workers wear helmets according to the color difference between helmets and hair. At the same time, comparative experiments of different classifiers are also carried out.

The main work of the attitude estimation task at this stage is to reduce the complexity of the algorithm, be robust, and adapt to multiple scenes. Up to now, two DL methods of attitude estimation, top-down, and bottom-up have been developed [20]. Literature [21] constructed a cascade regression network to realize the estimation of human posture. In order

to better capture the context information in the whole picture, a deep neural network was used to make the regression unit of each joint use the information of the whole picture. A stacked hourglass model was proposed in the literature [22] to capture the information contained in pictures at different scales. Each hourglass network can use the previous output's thermal diagram as input and obtain the relationship of joint points to improve joint point prediction accuracy. Literature [23] proposed a method for estimating attitudes that did not require explicit graphical model inference. The sequential architecture composed of CNN runs directly on the response map of the previous stage, allowing for accurate estimation of the local position. The literature [24] proposes a CNN algorithm based on region extraction-R-CNN framework that improves target recognition accuracy. [25] cited CNN's excellent feature extraction and classification performance, introduced the selective search extraction area method, extracted 2000 bounding boxes from the image to be recognized, and extracted the features of each bounding box using the CNN method and classified using SVM (support vector machine). A Fast-RCNN-based computer vision method is presented that can be used to identify workers without helmets in remote monitoring images. CNN-based recognition, on the other hand, necessitates a large number of marked images.

To sum up, it can be seen that DL based on CNN has a good performance in target recognition and has been widely used in the civil engineering industry. It provides an effective technical support means for the identification of construction workers' safety helmets by monitoring video stream at the construction site. This paper will combine the video characteristics of real-time monitoring on the construction site and apply the identification method based on CNN to the practical application of whether the construction workers wear helmets or not.

## 3. Research Method

*3.1. Design of Helmet Wearing Identification System.* Safety helmet is an important device to protect the head. Construction and site management personnel must wear safety helmet correctly after entering the construction site; that is, "people" and "hats" are not separated. Therefore, in order to ensure the accuracy and reliability of the image detection results, first detect the area of the head wearing the helmet and then analyze and judge whether the person is wearing the helmet by extracting the characteristics of the helmet.

The helmet wearing monitoring system under intelligent video surveillance proposed in this paper includes a series of processes such as real-time video acquisition, helmet wearing identification, and personnel tracking and summarizes the functions and performance requirements of this system as follows:

When the detection module detects someone not wearing a helmet, the system should issue a timely alert. To avoid false alarms and repeated alarms, the system should track the detected personnel and grasp the personnel situation in real time, ensuring that those who do not wear helmets receive timely and accurate alarms. The system's accuracy is primarily divided into two parts. The first step is to

accurately detect whether or not workers are wearing helmets and not to miss or miss too much. Furthermore, giving an alarm to personnel who do not wear a safety helmet is accurate, as long as there are not too many missed alarms, repeated alarms, false alarms, etc.

Figure 1 is a flow chart of helmet wearing monitoring system. It is mainly divided into two modules: helmet wearing identification module and personnel tracking module.

The detection module's function is to train the detection module with a large amount of data, detect real-time surveillance video captured by the camera, analyze the situation of employees wearing helmets, and obtain target information for employees in the video for use by the follow-up tracking module. The personnel tracking module sets up the tracking set using the detection target box sent by the front detection module and then uses the target tracking algorithm to track the target set. The main function of this module is to track all personnel appearing in surveillance video and capture their position information in real time, ensuring that when warning personnel who do not wear helmets, and it can avoid missing alarm, repeated alarm, and false alarm to the greatest extent possible.

The overall design of the system consists of control part, background logging part, and detection part. Control part is as follows: the interface is developed by C++/QT, which is mainly used to provide the administrator's operation interface and be responsible for the interaction with the background system, so as to facilitate the query of the monitoring logs and other information saved on the server. Background logging part is as follows: it is built by the traditional MVC (model view controller) architecture, which is separated from the front end, reduces the coupling of the system, and mainly provides various APIs for data interaction. The background part is mainly responsible for managing all kinds of monitoring data detected and analyzed by the monitoring part, generating logs according to the time and date and sorting and saving them into the database, which is convenient for managers to view. This system is simple in architecture design, and its functions can be expanded by correspondingly expanding the interface according to requirements. It has maintainability and universality. The system can well monitor the wearing condition of safety helmet in real time and inform relevant responsible persons to take countermeasures.

### 3.2. Implementation of the CNN Recognition Algorithm

**3.2.1. CNN Principle.** CNN stands for "convolutional neural network." It is a more advanced algorithmic mathematical model based on an artificial neural network that mimics a biological neural network and can be used to simulate a variety of signals. The incoming signal value determines whether the next neuron is activated in a biological neural network. The neuron will be activated, and the signal will be output if the signal value exceeds a certain threshold [19]. The transmission of input and output signals by CNN is based on this principle. As shown in Figure 2, it is divided into three layers: IL (Input layer), HL (hidden layer), and OL (output layer).

CNN IL can process multidimensional data, which is generally preset as three-dimensional input in the field of

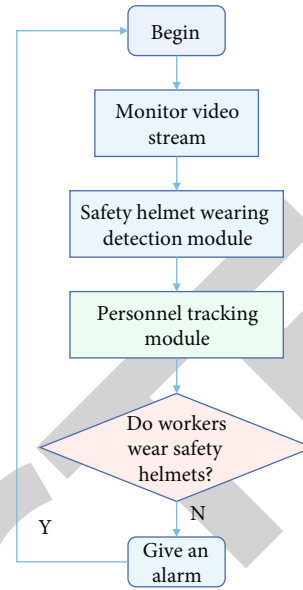


FIGURE 1: Comprehensive flow chart of helmet detection system.

computer vision, that is, pixels on the plane and RGB channels. Before the learning data enters CNN, the input data will be normalized in advance, and the original pixel values distributed on  $[0, 255]$  will be normalized to the  $[0, 1]$  interval, which will help improve the operation efficiency.

The HL in the structure includes CL (convolution layer), PL (pool layer), and FCL (full inception layer), and now it also includes incident module and RM (residual module). CL is used to extract features from input data, and the size of the extracted feature map is determined by convolution kernel size, compensation, and filling mode, as well as an excitation function to aid in the expression of complex features. Global mean pooling can sometimes replace the three-dimensional structure of the feature map by transferring it to other FCL via the excitation function. The upper layer of OL in a CNN is usually FCL, which is similar in structure and principle to the feed forward neural network OL. The logic function or normalization function is commonly used to output the classification label for the image classification part, and the output of the object recognition part also includes the size, coordinates, and category of the object.

**3.2.2. Realization of Identification of Construction Workers' Helmet Wearing.** The face detection algorithm based on DL can realize the end-to-end network structure, without manually extracting artificial features, automatically extracting image features by using multiple CL, and eliminating the influence of interference factors through deep network, which can greatly improve the recognition accuracy and is robust to the environment. Considering the timeliness, accuracy and applicability requirements of on-site safety monitoring in complex environment, the YOLOv3 algorithm is selected for face detection [10, 11].

The YOLOv3 algorithm can be used to identify the helmet wearing situation of construction workers in real time, and the image or video stream of the monitoring

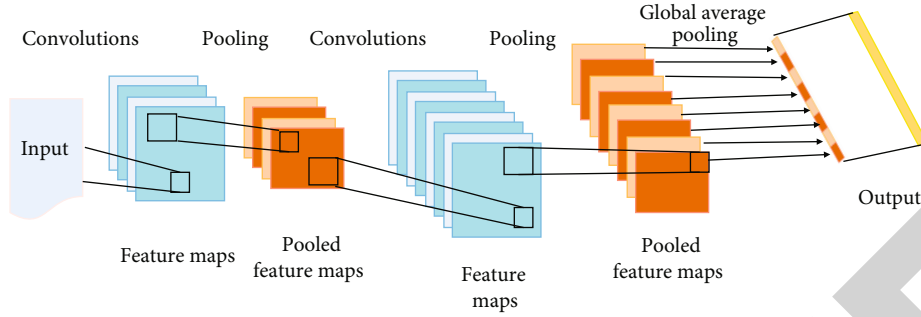


FIGURE 2: Schematic diagram of standard CNN structure.

system can be trained to the algorithm parameters. After successful training, the image or video stream can be identified in real time, and a good intelligent monitoring method is provided. Furthermore, using the YOLOv3 algorithm to identify the helmet wearing problem of construction workers can achieve all-weather real-time work without work and rest and without manual operation, and the work efficiency is extremely high. The YOLOv3 algorithm is used to identify construction workers who do not wear helmets. Actually, it is to mine and extract the category information and location information of helmets from video streams or images. The video characteristics of the construction site have a source influence on the algorithm recognition results. The YOLOv3 algorithm achieves the best balance between detection accuracy and speed through residual network feature fusion and multi-scale prediction. However, when the input is  $416 \times 416$ , the minimum feature map size used to extract features is  $13 \times 13$ , and the receptive field is too large compared with the actual target to be detected, resulting in poor detection effect for medium or small-sized objects, resulting in false detection, missed detection, or repeated detection.

In this paper, a new feature fusion algorithm, which is called I-YOLOv3 algorithm, is proposed by improving the YOLOv3 algorithm. The structure is shown in Figure 3.

Because of the complicated postures of constructors and various head postures, a single detection network cannot accurately detect the head area. The head positioning method proposed in this paper is used to accurately locate the head area of constructors and solve the problem that the head position of constructors is difficult to determine under complex postures.

In this paper, we use the  $C_r C_b$  range screening method based on  $YC_r C_b$  color space to identify skin color. The formula of cap transformation is shown in formula (1),

$$T_{\text{hat}}(f) = f - (f \circ b). \quad (1)$$

$f$  is a gray-scale image, and  $f \circ b$  is the on-off operation of structural element  $b$  on image  $f$ .

When the skin color area and the head area are compared and screened, the skin color areas with larger connected areas are not all face areas, but there are skin color areas such as arms, hands, and neck. Let the skin color area be  $F$ , the head area be  $H$ , and the evaluation value be  $x$ , and then the crossscreening formula of the head area is

$$\begin{cases} F \subset H, x = 10, \\ !\exists H, x = 8, \\ !\exists H, x = 5. \end{cases} \quad (2)$$

In the following training stage, continuously reduce the proportion of data samples with labeled information, so that the recognition network can get the information in the fixed target camera in the online learning framework, such as background, illumination, and viewing angle.

$D_L$  is defined as obtaining data samples with labeled information from the fixed target surveillance video, and this group of data samples with labels is  $N_L$ , while  $D_U$  is defined as the data sample without labeling information obtained from the fixed target surveillance video.

The control parameter  $m$  represents that proportion of tag data use in each training stage. The number of data samples extracted in each stage is  $N$ , in which the number of labeled data samples is  $N/m$ .

The output sample data of each cycle is as follows:

$$D_e^i = \left\{ D_{Le}^i = \Phi_{D_L} \left( \frac{N}{m} \right), D_{Ue}^i = \Phi_{D_U} \left( N \cdot \frac{N}{m} \right) \right\}, \quad (3)$$

where  $\Phi$  is a random sample selection operation,  $i$  is the  $i$ -th cycle, and  $D_e^i$  is the set of labeled information samples and unlabeled information samples output in the  $i$ -th cycle.

In the first training cycle, the value of  $m$  is the maximum, which means that the online learning framework starts training and needs to get more information from the target online surveillance video. The randomly obtained  $D_e^i$  data is transmitted forward and backward through the network, and the tentative network parameters of construction worker identification and helmet identification are obtained. The specific formula is as follows.

$$W_h^i = \arg \min L(N_h(D_e^i), T), \quad (4)$$

where  $\arg \min f(x)$  is the set of all independent variables  $x$  that make the function  $f(x)$  get its minimum value, and  $L$  is the regression loss function of YOLO.  $T$  is that real label information in the video image.

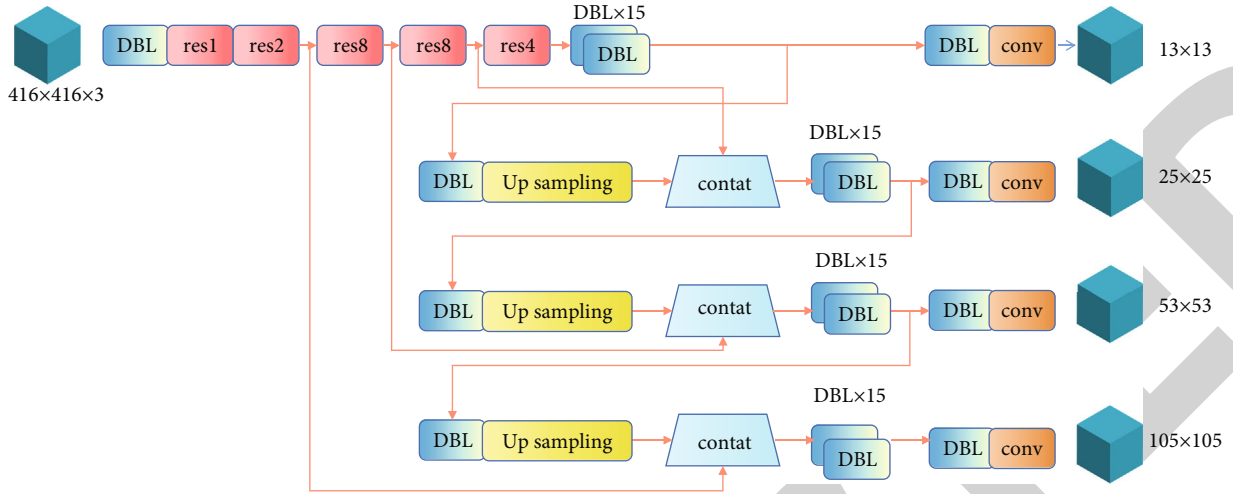


FIGURE 3: The model structure I-YOLOv3 of this paper.

Two residual blocks are stacked in sequence, and the corresponding formula is defined as formula (5):

$$x_{t+1} = F_t(x_t) + x_t, \quad (5)$$

where  $x_t, x_{t+1}$  is the input and output vector of the  $t$ -th residual block, respectively, and  $F_t(x_t)$  represents the conversion function, which corresponds to the residual differential branch composed of stacked layers. This deep residual network is easy for information flow and training.

The head position size information of employees will be directly predicted by I-YOLOv3, which is divided into four coordinate information  $t_x, t_y, t_w, t_h$ , which is defined as follows:

$$b_x = \sigma(t_x) + c_x, \quad (6)$$

$$b_y = \sigma(t_y) + c_y, \quad (7)$$

$$b_w = p_w e^{t_w}, \quad (8)$$

$$b_h = p_h e^{t_h}. \quad (9)$$

$c_x, c_y$  represents the horizontal and vertical distance between a grid and the upper left corner of the image, and  $p_w, p_h$  represents the width and height of the bounding box.

So far, the acceleration scheme of helmet wearing recognition model has been designed, and the model acceleration experiment in Chapter 4 of this paper will verify the scheme and prove its effectiveness.

#### 4. Results Analysis and Discussion

After logical regression analysis, the I-YOLOv3 framework obtains the classification score of each bounding box. The score is 1 if the predicted bounding box is close to the real bounding box; if the difference between the predicted bounding box and the real bounding box is large enough to fall below the threshold of 0.5, the score is 0, and the pre-

dicted bounding box is ignored. There are two basic requirements to control the quality of data samples and ensure the number of samples when collecting data from construction workers wearing helmets on the construction site, in order to ensure the algorithm training results of experimental pretraining. First, its real-time monitoring video system's shooting range should cover all construction sites and various construction site conditions; second, there must be sufficient image samples. Furthermore, two basic principles of construction site authenticity and noninterference with construction workers' behavior are followed during the collection of images for the helmet identification data set.

In this paper, I-YOLOv3 target detection network is used to detect the head area and helmet area. In this paper, the four indexes of accuracy, average accuracy, intersection-parallel ratio, and detection time are evaluated, and the comparative experimental results are shown in Figure 4.

It can be seen from Figure 4 that I-YOLOv3 target detection network is superior to the other three networks in terms of accuracy and detection time. As the latest detection network of YOLO series, I-YOLOv3 greatly improves the detection accuracy of the head area and helmet area in this paper.

In order to better express the experimental results, the 30-minute video stream captured 90 images at intervals of 20 seconds, counted the number of identified construction workers and helmets in the images, and then analyzed the accuracy of construction workers and helmets with broken lines, as shown in Figure 5.

As shown in the line chart of 5 recognition results, the integrity of the system's construction worker recognition performance is higher than that of helmet recognition accuracy, but other data also show that its algorithm overcomes the sunny and cloudy climate changes in the scene of the construction site, and the helmet wearing recognition system can still get better recognition results.

In a sunny climate, its shooting light is sufficient, and the collected video stream images are clear. Combining the above factors, the relativity of recognition accuracy is improved. In addition to these two special construction scene conditions,

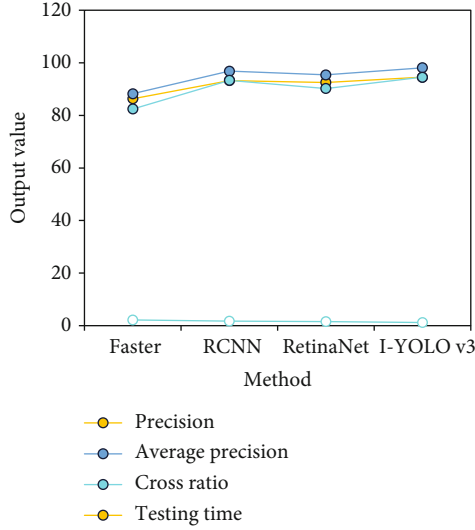


FIGURE 4: Comparison of experimental results of target detection network.

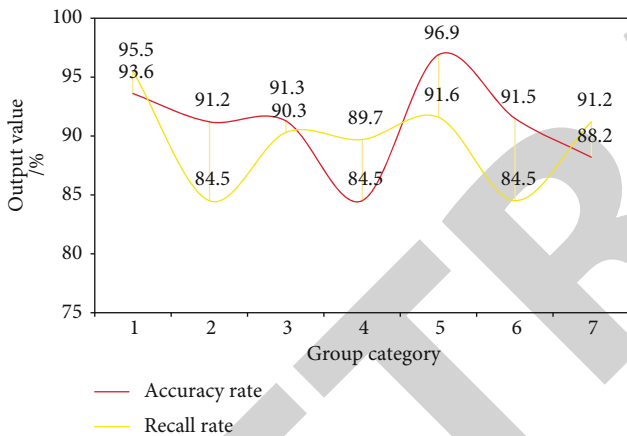


FIGURE 5: Identification result of helmet wearing identification system.

it can well meet the needs of identifying the wearing condition of safety helmet at the construction site.

In order to ensure that the system can accurately detect and track workers and give an alarm to those who do not wear helmets, the background algorithm of the system was tested. 100 pictures were extracted from three scenes, and the system algorithm was tested. The related data were recorded and made as shown in the following Figures 6 and 7:

From Figure 6, it can be seen that the system detection module has a very good effect on the target with a helmet. With a high recall rate, there are few false detections, but there are many missed detections in real scenes.

In Figure 7, the detection of targets without helmets is also a high recall rate, but the false detection and missed detection are slightly more than those with helmets. However, considering the actual application requirements of the system, to alert those who do not have helmets, the most important thing is to detect those without helmets, that is, to ensure a high recall

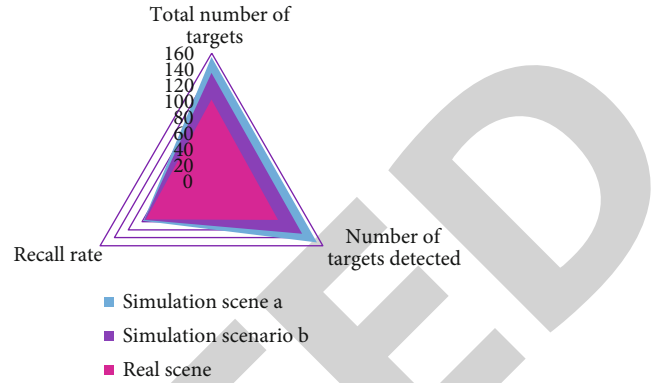


FIGURE 6: Target detection effect of wearing helmet.

rate of target detection. From Figures 6 and 7, the detection module of the system can guarantee a high recall rate in three scenarios.

The variation trend of the model loss value and mAP with the number of iterations in the training process is shown in Figure 8.

As can be seen from the Figure 8, the loss value of the model dropped rapidly during the first 100 iterations, then dropped gently, and finally stabilized around 0.75, indicating that the model has converged. The experimental results show that the mAP value of the model is the highest in the 200th batch, which shows that the mAP decreases by 0.25%, with slight fluctuation, but it has little influence on the judgment of the model.

Supervisors, construction managers, technicians, and field workers should wear blue, white, red, and yellow helmets, respectively, for ease of supervision. The method test in this paper selects the entrance to carry out the intelligent identification test of helmet wearing for people entering and leaving the construction site due to the limitations of actual engineering construction. Positive samples are people wearing various colored helmets, while negative samples are objects on the construction site that are similar in shape and color to helmets. Given the similarities between the shape of the constructors' hair and the shape of the helmet when they are not wearing it, various hairstyles and color samples are added to the negative samples, the positive and negative samples are classified and marked for training, and the training results are obtained.

Figures 9 and 10 show the variation scatter of the accuracy of the training and testing network with the number of iterations and the variation scatter of the loss rate with the number of iterations.

From the curves of accuracy  $_A$ , accuracy  $_B$ , and accuracy  $_C$  in Figure 9, it can be seen that increasing the network layer properly can improve the accuracy and convergence speed of the network. The number of iterations has a great influence on the accuracy and convergence of the network. Increasing the number of iterations can improve the accuracy and reduce the iteration time. The infinite increase of iteration makes the accuracy and loss rate of the network tend to be stable.

Through 1 000 tests, this system can accurately identify people who are not wearing helmets in sparse or crowded

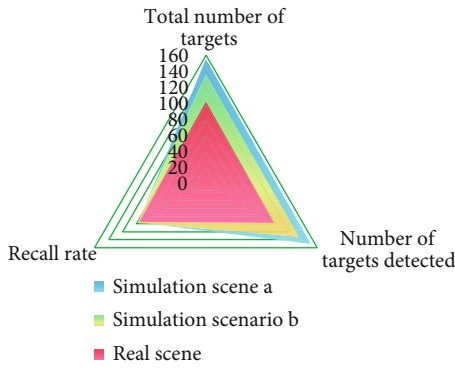


FIGURE 7: Effect of target detection without helmet.

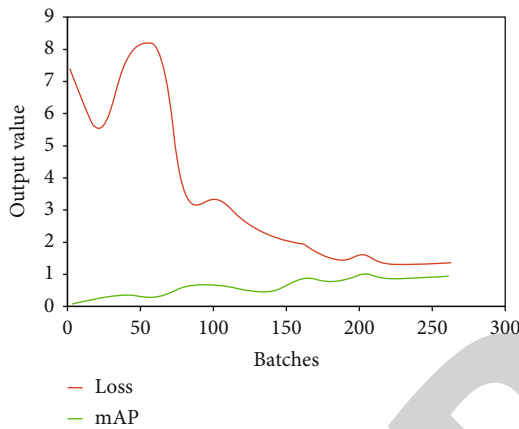


FIGURE 8: Loss value and mAP change curve of sparse training model.

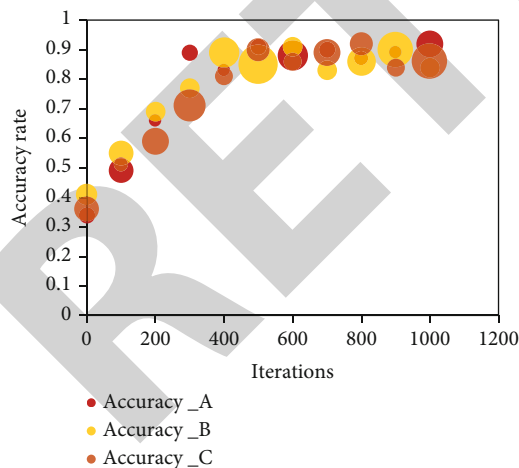


FIGURE 9: Relationship between correct rate and iteration times.

situations and can identify helmets of various colors. After the algorithm is optimized, it can also accurately identify the situation of partial occlusion when multiple people enter at the same time.

Comparing the loss rate  $_C$  curve in Figure 10, too little available data will lead to overfitting of the network, and the

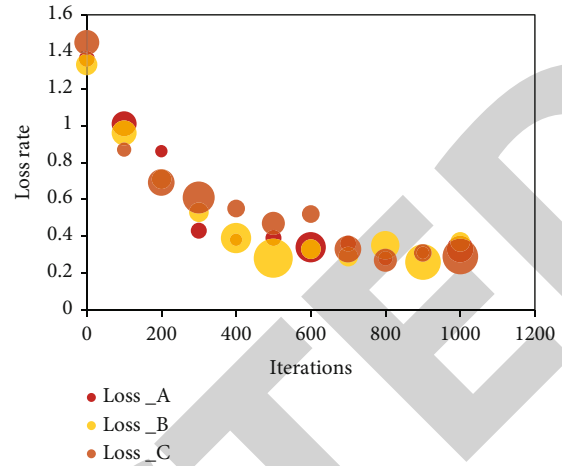


FIGURE 10: Relationship between loss rate and iteration times.

generalization ability of the network model is too poor. Increasing the data amount will help to enhance the accuracy of the test and the generalization ability of the network model.

The system records daily activities in real time and completes the safety management and resource management of the construction site. Because the construction site is in a very complicated open-air working environment, the surveillance video usually has changeable background and uncontrolled illumination. There are changes in weather and climate, shelter of site materials and machinery, and changes in visual range. According to the characteristics of surveillance video stream and construction site, this paper explores the influence of semisupervised real-time recognition algorithm based on I-YOLOv3 on the sensitivity of external conditions. The images in the data set will include multiclimatic conditions, multipose of construction workers, multidistance shooting, and multiocclusion of body parts and crowded degree.

The effect of real-time recognition is influenced by weather, lighting, personal posture, visual range, and occlusion. The robustness is used to verify whether the algorithm has good recognition performance in the face of various conditions and changes in order to explore the applicability of recognition to the construction site. The accuracy and recall rate can be a good indicator of the model's robustness in different situations.

## 5. Conclusion

In the monitoring of building construction environment, helmets are very small targets. For helmet wearing identification task, it is of great significance to solve the problem of low detection accuracy of small targets. A modified I-YOLOv3 algorithm which is deeply integrated with feature map is designed. In order to obtain more information of small and medium targets, a CL with larger feature map size is added after the three layers of CL in the main network, and four feature pyramids with different scales of CL are constructed together. At the same time, upsampling is performed with twice step size, and the detection performance of the system is improved by fusion with the front depth residual network. I-YOLOv3 target detection network is used to identify the



helmet, and the wearing condition of the helmet is judged by the positional relationship between the helmet area and the head area. Finally, the performance of other helmet wearing identification methods is compared and analyzed through experiments.

At present, our system mainly completes the design and implementation of real-time image acquisition and background algorithm, and there is no corresponding front-end alarm display interface; so, we cannot effectively grasp the real-time situation of the system, and we can design a complete front-end display and alarm system in the follow-up.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors do not have any possible conflicts of interest.

## Acknowledgments

This study was supported by the Huanggang Normal University High-level Cultivation Projects (No. 37), Huanggang Normal University First-Class Undergraduate Course Projects (No. 2021CK41), and Hubei Provincial Social Science Fund Prophase funding projects (No. 20ZD096).

## References

- [1] M. Ji, K. Zhang, Q. Wu, and Z. Deng, "Multi-label learning for crop leaf diseases recognition and severity estimation based on convolutional neural networks," *Soft Computing*, vol. 24, no. 20, pp. 15327–15340, 2020.
- [2] X. Jin, J. Zhang, J. Kong, T. Su, and Y. Bai, "A reversible automatic selection normalization (RASN) deep network for predicting in the smart agriculture system," *Agronomy*, vol. 12, no. 3, p. 591, 2022.
- [3] Y. Huang, L. Cheng, L. Xue et al., "Deep adversarial imitation reinforcement learning for QoS-aware cloud job scheduling," *IEEE Systems Journal*, vol. 1, pp. 1–11, 2021.
- [4] L. You, H. Jiang, J. Hu et al., "GPU-accelerated Faster Mean Shift with euclidean distance metrics," 2021, <https://arxiv.org/abs/2112.13891>.
- [5] Y. Hu, J. Zhan, G. Zhou et al., "Fast forest fire smoke detection using MVMNet," *Knowledge-Based Systems*, vol. 241, p. 108219, 2022.
- [6] Y. Ding, X. Zhao, Z. Zhang, W. Cai, N. Yang, and Y. Zhan, "Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.
- [7] D. Yao, Z. Zhi-li, Z. Xiao-feng et al., "Deep hybrid: Multi-graph neural network collaboration for hyperspectral image classification," *Defence Technology*, vol. 5, 2022.
- [8] F. Dingbang, F. Gui, C. Haiyan, Y. Hengjie, H. Xue, and Y. Yincheng, "Handwritten formula symbol recognition based on multi-feature convolutional neural network," *Laser & Optoelectronics Progress*, vol. 56, no. 7, article 072001, 2019.
- [9] Y. Guo, X. Wang, P. Xiao, and X. Xu, "An ensemble learning framework for convolutional neural network based on multiple classifiers," *Soft Computing*, vol. 24, no. 5, pp. 3727–3735, 2020.
- [10] Y. Li, H. Wei, Z. Han, J. Huang, and W. Wang, "Deep learning-based safety helmet detection in engineering management based on convolutional neural networks," *Advances in Civil Engineering*, vol. 2020, 10 pages, 2020.
- [11] S. Chen, X. Tan, B. Wang, H. Lu, X. Hu, and Y. Fu, "Reverse attention based residual network for salient object detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 3763–3776, 2020.
- [12] S. Kruthiventi, K. Ayush, and R. V. Babu, "DeepFix: a fully convolutional neural network for predicting human eye fixations," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4446–4456, 2017.
- [13] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.
- [14] A. T. Lopes, E. de Aguiar, A. F. de Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.
- [15] H. Chen, Y. Zhang, W. Zhang et al., "Low-dose CT via convolutional neural network," *Biomedical Optics Express*, vol. 8, no. 2, pp. 679–694, 2017.
- [16] M. Liu, J. Shi, Z. Li, C. Li, J. Zhu, and S. Liu, "Towards better analysis of deep convolutional neural networks," *IEEE Transactions on Visualization & Computer Graphics*, vol. 23, no. 1, pp. 91–100, 2017.
- [17] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, 2017.
- [18] L. Ying, H. Zhang, and S. Qiang, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.
- [19] H. Chen, Y. Zhang, M. K. Kalra et al., "Low-dose CT with a residual encoder-decoder convolutional neural network (RED-CNN)," *IEEE Transactions on Medical Imaging*, vol. 36, no. 99, pp. 2524–2535, 2017.
- [20] R. T. Schirrmester, J. T. Springenberg, L. D. J. Fiederer et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [21] Y. Qian, M. Bi, T. Tan, and K. Yu, "Very deep convolutional neural networks for noise robust speech recognition," *IEEE/ACM Transactions on Audio Speech & Language Processing*, vol. 24, no. 12, pp. 2263–2276, 2016.
- [22] W. Shen, M. Zhou, F. Yang et al., "Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification," *Pattern Recognition*, vol. 61, no. 61, pp. 663–673, 2017.
- [23] W. A. Hao and A. B. Jz, "Automated visual helmet identification based on deep convolutional neural networks," *Computer Aided Chemical Engineering*, vol. 44, pp. 2299–2304, 2018.
- [24] K. Ram and P. K. Bajpai, "FEM analysis of glass/epoxy composite based industrial safety helmet," *Iop Conference*, vol. 225, article 012174, 2017.
- [25] J. Wismans, I. Skogsmo, A. Nilsson-Ehle, A. Lie, M. Thynell, and G. Lindberg, "Commentary: status of road safety in Asia," *Traffic Injury Prevention*, vol. 17, no. 3, pp. 217–225, 2016.