Incorporate Online Hard Example Mining and Multi-Part Combination Into Automatic Safety Helmet Wearing Detection

NING LI [1,2], **XIN LYU** [1], **SHOUKUN XU** [2], **YARU WANG** [2],
YUSHENG WANG [2], **AND YUWAN GU** [2]
[1] School of Computer and Information Engineering, Hohai University, Nanjing 210098, China
[2] Aliyun School of Big Data, Changzhou University, Changzhou 213164, China

Corresponding authors: Ning Li (lncczu@cczu.edu.cn) and Xin Lyu (lvxin@hhu.edu.cn)

ABSTRACT Automatic detection of workers wearing safety helmets at the construction site is essential for safe production. Aiming at the problem of low recognition rate caused by factors such as background and light in the automatic detection of safety helmets using traditional machine learning methods, this paper proposes an object detection framework that combines Online Hard Example Mining (OHEM) and multi-part combination. In our framework, we first use the multi-scale training and the increasing anchors strategies to enhance the robustness of the original Faster RCNN algorithm to detect different scales and small object. Then, the OHEM is to optimize the model to prevent the imbalance of positive and negative samples. Finally, the person wearing the helmet and its parts (helmet and person) are detected by improved Faster RCNN. The multi-part combination method uses the geometric information of the detection objects to determine if a worker is wearing a helmet. Experiments show that compared with the original Faster RCNN, the detection accuracy is increased by 7%. It also has better detection performance for partial occlusion and different-size objects, showing good generalization and robustness.

INDEX TERMS Safety helmet wearing detection, multi-part combination, faster RCNN, multi-scale training, online hard example mining.

I. INTRODUCTION

Various risk factors safety of workers due to complex environment in chemical plants, power substations and construction sites. The causes of injury and fatality include falls, slips, being corroded by chemicals, being struck by objects and electrocution, etc. Struck by falling object and falls to lower level are the leading hazards. According to the Occupational Safety and Health Administration (OSHA) statistics, about 5-6% of fatal accidents in the United States are caused by falling objects [1]. There is one third of the deaths because of falling to lower level [2]. Therefore, people working in such places must wear safety helmets to protect them from being struck by falling objects and falling to lower level [3]. Automatically detecting workers wearing safety helmets at the construction

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa Rahimi Azghadi [ID].

site and making corresponding feedback in the monitoring system is crucial for safety production.

With the development of computer technology, automatic visual detection has been widely used in industrial applications. Many related studies have been conducted for helmet wearing detection [4], [5]. Wua and Zhaoa [4] divided the entire helmet wearing detection process into two parts. First, workers were detected by combining the frequency domain information and Histogram of Oriented Gradient (HOG) of the image. Then, the color and Circle Hough Transform (CHT) features were combined for safety helmet detection. The method achieved a certain detection effect; however, the overall accuracy of the method is low and only a specific color safety helmet can be detected. Rubaiyat *et al.* [5] utilized the Local Binary Patterns(LBP), Hu Moment Invariant(HMI) and Color Histogram(C.H.) of the image to extract the feature of different-color helmets,

then hierarchical Support Vector Machine (SVM) is used to recognize safety helmets.

Above methods are based on traditional machine learning methods for object detection. These methods are mostly based on subjective feature selection, which required a solid professional foundation and rich experience. Moreover, feature selection is time-consuming, and its generalization ability is poor, hard to adapt to changes in conditions, such as lighting.

With the rapid development of deep learning in recent years, more and more researchers have applied deep learning methods to many complex tasks such as image classification [6], object recognition [7], image segmentation and detection [8], etc. Object detection algorithms based on deep learning are mainly divided into two categories: one is RCNN series, such as Fast RCNN [9], Faster RCNN [10], and R-FCN [11]. Faster RCNN modularized (region proposals generation, feature extraction, object classification, location refinement) the object detection into a deep network framework and fully implementing an end-to-end object detection. The detection results of such algorithms are more accurate, but the speed is slower. Another type of method converts detection problems into regression problems, such as YOLO3 [12], SSD [13], and RetinaNet [14], etc. Such algorithms run faster, but object detection accuracy is lower, especially for small objects.

Due to the challenge of detecting small helmet targets on the construction site, this paper proposes an object detection framework based on combining Online Hard Example Mining (OHEM) and multi-part combination. In the framework, the OHEN strategy is employed to extract the personnel wearing safety helmets and their safety helmets for coarse detection. Then the multi-component combination method is utilized to calculate the belonging relationship of the components to detect the wearing of the safety helmet accurately. The main contributions of this paper are as follows:

(1) To solve the problem that hard negative samples are difficult to learn, an OHEM learning strategy is proposed to initially select hard negative samples, and then input them into the network again for retraining, so that the network pays more attention to these hard negative samples.

(2) A multi-part model is proposed to determine whether there is a component in the corresponding position in regions of interest(ROIs), which can further eliminate the false object and improve the detection accuracy.

(3) The framework can automatically detect the wearing of safety helmets in different construction site scenarios. This method can obtain better detection accuracy and robustness than the other state-of-the-art methods.

The rest of this paper is organized as follows. Section II introduces the existing safety helmet wearing detection methods. The detailed description of our proposed method is then presented in Section III. In Section IV, the experimental results on datasets are reported. Finally, the conclusion is provided in Section V.

II. RELATED WORK

Safety helmet wearing detection has been extensively studied in the literature of computer vision. The helmet wearing detection is the basis for analyzing production safety on the construction site, and provides essential technical support for enterprise intelligent video surveillance. Safety helmet wearing detection methods are mostly based on traditional machine learning methods. For example, in 2013, Lin *et al.* [14] designed a helmet wearing detection system for traffic scenarios. The method detected whether to wear a helmet based on the detection of the motorcycle driver, and used the upper 1/5 of the object area of the motorcycle driver as a potential area of the helmet, and extracted local binary pattern features and HOG features of the object. Then three types of classifiers, including Naive Bayes, Random Forest, and SVM, were trained for comparative experiments. The results showed that the trained random forest classifier had the best detection performance with an accuracy of 93.08%. In 2014, Silva *et al.* [15] first used the Adaptive Mixture of Gaussians (AMG) to extract moving objects and then detected motorcycle drivers. Through the calculation of the sub-window, the human head is framed, converted into a grayscale image, and the mean filtering is performed to denoise, and the binarization conversion and the hough transform are performed to find the circular region operation. Then the LBP, HOG and W.T. features of the head region were extracted, and any pairs of two features were combined. The experimental results showed that the detection based on the combination of HOG and LBP features was the best, and the detection accuracy was as high as 94.04%. In 2018, [16] proposed a method for helmet recognition based on feature fusion. First, a head image was extracted based on the acquired video. Then, the LBP (texture), H.U. moment invariant (geometry), and color histogram (color) feature vectors of the head image were extracted. Finally, the head image was divided into four categories (red hard hat, yellow hard hat, blue hard hat, and no hard hat) using a hierarchical SVM (HSVM). The method in [16] not only monitored whether the worker weared a helmet, but also further recognized the color of the helmet. Based on the traditional machine learning object detection algorithms, relevant researchers are required to conduct in-depth research on detection fields for different detection tasks by designing specific and adaptable features. Such methods are individually optimized during the feature extraction and classifier training phases and do not affect each other. But it is susceptible to environmental changes.

In recent years, we have witnessed advances in object detection using deep learning, which often outperforms traditional computer vision methods significantly. For example, in 2017, Wu and Zhao [4] developed a system for automatic detection of helmet driver wearing in a traffic scene. First, the adaptive image subtraction method was used to acquire dynamic objects for video images. Then two different convolutional neural networks were used to perform motorcycle driver detection and helmet detection. The experiment used two data sets: one containing one single object per image

without small object and fuzzy object and the other one consisting of multiple objects per image with occlusion and small objects. The average detection accuracy of the experiment was as high as 92.87%. This method is a helmet wearing test in a traffic scene, and the background, category and posture of the picture are relatively simple. Also, it uses two deep convolutional neural networks, which are cumbersome and increase the computational complexity. Inspired by the wide application of Faster RCNN in the field of object detection. In this study, we propose a framework for safety helmet wearing detection by improving the Faster RCNN.

III. PROPOSED METHOD

This paper proposes an object detection framework that combines OHEM [17]–[19] and multi-part combination [20]–[22], which adopt the Faster RCNN as the backbone. Firstly, multi-scale strategy is adopted in the network training stage to enhance the robustness of the object's size and increase the number of anchors to improve the detection accuracy for small objects. Then, the OHEM method is used to automatically select hard samples, which are fed into the network again to retrain. During the training process, the samples of wearing safety helmets are often consider hard negative samples. The retraining of these samples can make the network pay more attention to these samples with safety helmets. Finally, according to the geometric information between workers and safety helmets, a multi-part model is proposed to eliminate false detection objects and identify missed objects to improve the detection accuracy. In the following, we discuss our framework in detail.

A. MULTI-SCALE TRAINING

At the actual construction site, the difference in size between different targets such as helmet workers and helmets is large, and the sizes of similar targets in the same image are also different. To detect different-sizes safety helmets, we utilize the pyramid method to extract multi-scale images semantics.

The original Faster RCNN network sets the short side of the image to 600 based on the premise that the original image scale of the input image is unchanged. There is only one scale, which makes the network have poor generalization performance for different-sizes object. In this paper, the multi-scale strategy is adopted in the network training process to make the Faster RCNN network learn and extract the different-scale features of the object. During the network training process, the input image is randomly resized under the premise of ensuring the original proportion of the image, so that the shorter side takes the pixel size of one of 480, 600 and 750. Then one of the three scales is randomly selected and sent to the network for training. Experiments show that multi-scale training enables the network to learn various-dimension objects, making the network robust to the object size.

To improve the ability of the network to detect small targets, we have modified anchor parameters of the network. Based on the default parameters, a set of 64×64 anchors (smaller than the default setting) allow the network to detect

more small targets. In the training process, the RPN part uses 12 anchor points, the size of which is 64×64, 128×128, 256×256, 512×512, and the three aspect ratios are 1:1, 1:2 and 2:1, respectively. Experiments show that the increased scale of 64×64 can detect smaller targets.

B. OHEM

The hard sample is the sample where the wrong object is classified as correct and the confidence threshold is high. In the training process of Faster RCNN, many ROIs will be generated randomly in the RPN. Due to the small proportion of the object in the image, there is a huge imbalance between the number of positive samples and negative samples, and the network training model tends to be negative samples.

To make the network pay more attention to those hard samples, OHEM is incorporated into the backbone network for safety helmet wearing detection, which can simultaneously select hard samples without setting the positive and negative ratio of samples. The structure of Faster RCNN with OHEM is shown in Fig.1.
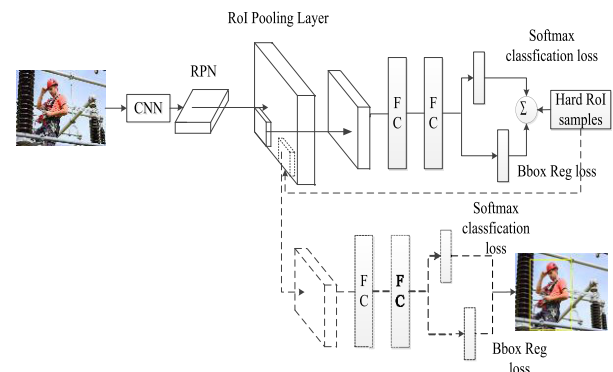

FIGURE 1. The structure of faster RCNN with OHEM.

The latter part of the ROIs pooling layer of Faster RCNN was called the ROIs network. Integrating OHEM method, the original ROIs network is expanded into two ROI networks, which share net-work parameters. One of them is read-only. In the read-only ROIs network, all operations are forward. Its main functions include calculating and sorting the loss values of all region proposals, selecting 128 region proposals with large loss values. Another ROIs network is the standard ROIs network, which contains forward and backward operations. The input is the hard sample selected by the first ROIs network. The output is the predicted classification result and the coordinates of the bounding box. In conclusion, an extra ROIs is added to select hard examples, which are then used for the standard ROIs network training. This algorithm does not need to set the ratio between positive and negative samples to solve the imbalanced problem. It improves the accuracy of object detection. The experiments show that the OHEM strategy can enhance the discrimination ability of the algorithm and improve the detection accuracy of network.

C. MULTI-PART COMBINATION

Whether the worker wears a helmet is mainly determined by whether there is a helmet in the head area of the worker. To mark the image of the helmet, the helmet worker and the helmet worker without wearing the helmet, and then use the optimized Faster RCNN network for model training. Since the helmet area is relatively small in the region of the helmet wearing area, the network will confuse the worker who does not wear the helmet with the worker who wears the helmet, resulting in a wrong inspection.

According to the geometric position relationship between the helmet and the worker, the positional relationship between the object and the component is calculated to eliminate the false detection object. We propose a multi-part combination method to detect the helmet on the worker's head, as shown in Fig.2.
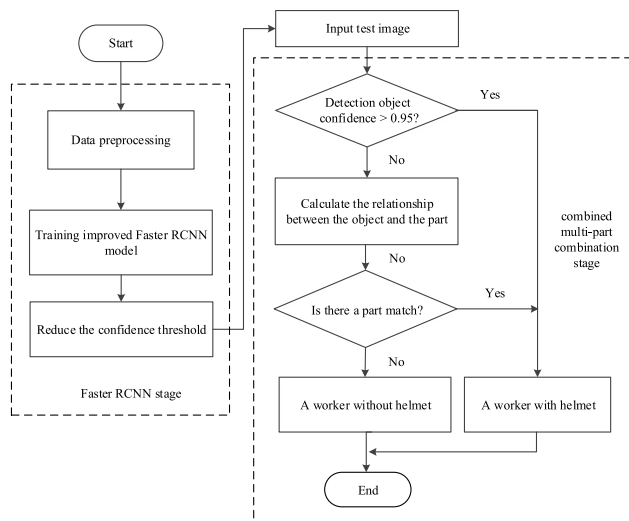

FIGURE 2. Multi-part combination method.

The training data set labeling these types of targets are input into the optimized Faster RCNN network for training. After the initial target detection using the optimized Faster RCNN framework to reduce the network detection confidence threshold to achieve more goals (wearing helmet workers and not wearing helmet workers) and components (helmets).

For the detected wearing helmet worker, the relationship between the component and the worker is judged by calculating the overlapping ratio of the component. The relative positional relationship between the component and the worker is calculated to determine the object category. We examine the upper 1/3 part of the target as a potential area for the helmet. If the target and the target overlap rate are the highest and the relative positional relationship is correct, it is judged to be wearing a helmet worker, otherwise, it is a wrong check. For the detected un-wearing helmet worker target, check if there is a helmet at the top of the target, and if it exists, it is a wrong check.

The functions of Faster RCNN part are data set format conversion and model optimization. The multi-part combination method is used to determine whether there is a corresponding part in the object area, such as a safety helmet. The whole process of our framework is as follows.

(a) Get datasets from the VOC2012 dataset [23]–[25], Internet, and other ways. Then, labeled helmet, worker and the worker with helmet in the datasets are converted to the VOC2007 dataset format.

(b) The processed data set was imported into the improved Faster RCNN model to train.

(c) Reduce the model detection confidence threshold, detecting worker wearing helmets and safety helmets, etc. Based on the above optimized model. Then, eliminate the isolated parts.

(d) For the remaining pending objects, calculate whether there is a matching part. If there is, it is our object. Else, remove it.

After testing with the improved Faster RCNN, the detected objects fall into two categories: workers wearing helmets and related parts such as safety helmets. If the confidence of our object area is less than 0.95 [26], [27] the relative positional relationship and the overlap rate between the part and our detected object are calculated to judge their affiliation. If the overlap rate is the largest and the relative positional relationship is correct (For example, the safety helmet at the top 1/3 of our object area). From this, we can ensure that this is our object.

The overlap rate formula [28] is shown in Eq.(1).

$$IoU = \frac{PartArea \cap OverallArea}{PartArea \cup OverallArea} \tag{1}$$

where PartArea is the area of the helmet and other parts after the detection. OverallArea is our object area. The IoU is the overlap rate between the part and the object. Finally, the isolated test results are removed, and the rest is our object, i.e., worker wearing helmet.

IV. EXPERIMENT ANALYSIS

The data set we created, comparative methods, and performance metrics used to validate our approach are presented in this section. The proposed algorithm is compared with the current typical object detection algorithms on our data set. The experiment used the Caffe (Convolution Architecture For Feature Extraction) deep learning framework for related codes and parameters training. The network framework of Faster RCNN uses the VGG 16 network. Experimental environment configuration: GPU: GeForce GTX 1080Ti, CUDA8.0, Ubuntu16.04, memory 12GB.

A. DATASETS AND EVALUATION METRICS

The image of the worker's work image at the construction site is the basis for studying the wearing of the worker's helmet for the construction site. At present, there is no publicly available image data set of workers' work images at the construction site. Image data is an indispensable element of

image processing tasks. The quality and quantity of image data have a significant impact on the results of helmet wear detection.

This paper refers to the establishment criteria of PASCAL VOC dataset, combined with the requirements of this method and the characteristics of detection targets, to establish a more standardized construction site worker image dataset containing multiple scenarios and multiple objectives. There is no public dataset in the research on safety helmet wearing detection. The data used in this experiment were collected from VOC2012 dataset, self-collection and online collection. A total of 7000 images were collected, including monitoring pictures with different quality under various background scenes in construction sites and substations. Some example images are shown in Fig.3.

FIGURE 3. Datasets for safety helmet detection.

According to the experimental requirements, the datasets were converted into VOC2007 datasets format. The example is shown in Fig. 4. It manually labels each part. In addition, an extra 200 monitoring images in actual work scenes are collected for testing to verify the effectiveness of the proposed method.
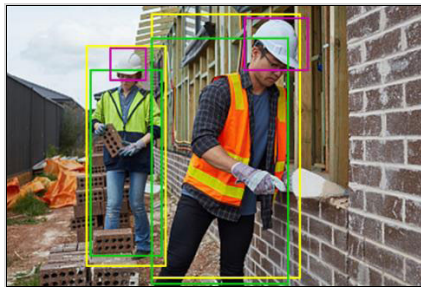

FIGURE 4. Annotated image data.

To evaluate the effectiveness of the proposed method for object detection, the experiment uss precision and recall [28] for evaluations. The calculation formula are shown in Eq. (2) and Eq. (3).

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

where TP (True Positive) represents a positive sample predicted to be positive by the model. FP (False Positive) represents a positive sample predicted to be negative by the model. FN (False Negative) represents a negative sample predicted to be positive by the model.

B. COMPARE THE DETECTION EFFECTS OF THE ORIGINAL FASTER RCNN AND THE IMPROVED FASTER RCNN IN THE SAME SAMPLES

In order to verify the effectiveness of the improved Faster RCNN, 7000 pictures of VOC2007 format are used as the training set. The original Faster RCNN network and the improved Faster RCNN are trained through multi-scale training to increase anchor points and OHEM. Two models were tested using 200 actual scene monitoring images (including 377 objects). The results of the two models were shown in Table 1.

TABLE 1. Test results using two models.

Method	TP	FP	FN	Recall/%	Precision/%
Faster RCNN	310	44	67	82.22	87.57
Improved Faster RCNN	341	32	36	90.45	91.42

Table 1 shows that the improved Faster RCNN improves test accuracy by 3.85% and recall by 8.23%. Compared with the original Faster RCNN network, the accuracy and recall rate of the optimized Faster RCNN is greatly improved, whereas the false detection target and the missed detection target are reduced. For helmet workers, reducing false detection targets can reduce false positives, whereas reducing false positives can increase real alert rates. It can be seen that the optimized Faster RCNN network has strong robustness for complex scenes including chemical plants, substations, and building construction.

Fig. 5 shows the detection effect using two algorithms on the actual picture. The green, purple, and yellow boxes are the worker, the safety helmet, and the worker wearing the safety helmet, respectively. The category name and confidence value are displayed above the bounding box. As can be seen from Fig. 5, the improved Faster RCNN is significantly better than the original Faster RCNN. Fig. 5(b) is able to detect more occludded targets and small targets, compared to Fig. 5(a). The detected target confidence values are also higher and the positions are more accurate. Experiments show that the improved Faster RCNN network can effectively optimize the model.
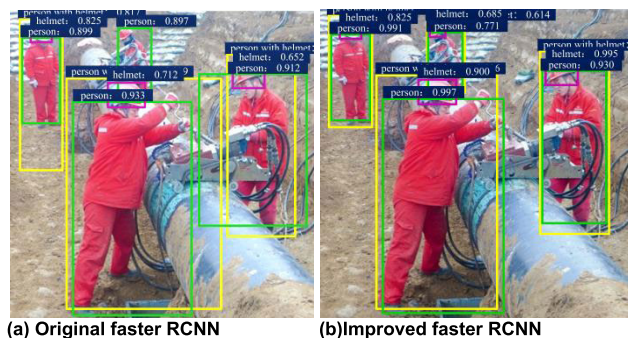

(a) Original faster RCNN   (b)Improved faster RCNN

FIGURE 5. Comparison of detection effects between two models.

C. THE EFFECT OF TRAINING THE NETWORK USING DIFFERENT STRATEGIES

To verify the effectiveness of different strategies, different strategies were used to train and test. The detection performance is shown in Table 2.

TABLE 2. Detection accuracy under different strategies.

Strategies	Anchors	Multi-scale	OHEM	Recall/%	Precision/%
1	9	no	no	82.22	87.57
2	12	no	no	83.96	88.36
3	12	yes	no	86.31	89.51
4	12	no	yes	87.59	89.88
5	12	yes	yes	90.45	91.42

Compared with strategy 1 and strategy 2, the detection accuracy is increased by 0.79%, which is because a set of (64×64)-scale anchors are added into network. In the experiment, the number of anchor points is increased from 9 to 12, so that the network can detect small objects. Compared with strategy 2 and strategy 3, the test precision of the network model is improved by about 1.15%, which is because the network model adopts the multi-scale training strategy in the training stage, which makes the network have certain robustness to objects with different sizes. Compared with strategy 2 and strategy 4, the detection accuracy of strategy 4 was improved by about 1.52% due to OHEM mechanism. This way can solve the problem of too large negative sample space in the training process and enhance the network resolution. In conclusion, all three strategies improve the detection performance of the network model.

D. DETERMINE THE CONFIDENCE THRESHOLD

To obtain more object parts such as wearing safety helmet worker and safety helmet in the initial detection stage, a lower confidence threshold value should be set. The experiment is based on the improved Faster RCNN framework to discuss the confidence threshold. Table 3 shows the object detection results under different confidence thresholds.

TABLE 3. Object detection results under different confidence thresholds.

Confidence thresholds	False detection rate/%	Miss rate/%
0.1	19.1	3.4
0.2	16.5	3.4
0.3	12.6	7.5
0.4	8.3	13.3

As shown in Table 3, when the confidence threshold is 0.2, the object that workers wearing safety helmet have the lowest miss rate, while the false detection rate is lower compared with the threshold of 0.1.

To improve the accuracy of detection to the greatest extent, confidence threshold is selected as 0.2. As shown in Fig. 6, when the confidence threshold value is 0.2, the worker wearing safety helmet is detected in Fig. 6(a), in which confidence
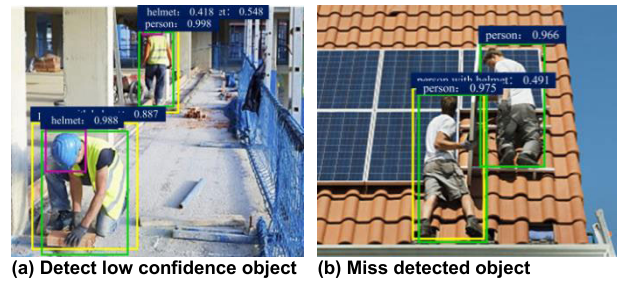

(a) Detect low confidence object   (b) Miss detected object

FIGURE 6. Test results with lower confidence threshold.

is 0.548. Also, the parts with lower confidence are detected, so that more objects and parts can be detected in the initial detection stage. However, the wrong objects will also be detected when the confidence threshold is lower. As shown in Fig.6 (b), the left worker is wrongly detected as a worker wearing a safety helmet. Therefore, the detected object needs to be filtered by multi-part combination method.

E. MULTI-PART COMBINATION METHOD

After reducing the detection confidence threshold, the multi-part combination method will be employed to detect the object more accurately. Some examples of detection results are shown in Fig.7.

After reducing the confidence threshold, the isolated parts are filtered out, and the rest is the object that is expected to be detected. Fig. 7(a) is the detection result when the confidence threshold is high, and Fig. 7(b) is the detection result after reducing the confidence threshold. Fig. 7(b) detected more objects and parts that are missed in Fig. 7(a). Using the multi-part combination method, the overlap rate and relative position relationship between the parts and the object are calculated. If the relationship is correct, the object will be judged to be the worker wearing safety helmets. Therefore, the three objects on Fig. 7(a) are detected as worker wearing a safety helmet.

When the confidence threshold decreases, a misdetection target appears in Fig. 7(c). We calculate the relative positional relationship of the parts and find that it is incorrect (the position of the safety helmet) using multi-part combination method. Therefore, the misdetection object is removed, as shown in Fig. 7(d). After the multi-part combination method is performed, the misdetection object is removed, resulting in the final detection object.

F. COMPARED WITH OTHER AUTOMATIC HELMET DETECTION METHODS

To further verify the effectiveness of the proposed method, in addition to the original Faster RCNN model, some representative target detection methods are selected for analysis and comparison. The chosen method is HOG+SVM based on traditional machine learning methods.

Experiments are performed using the training dataset and test datasets herein. Compare the test accuracy and recall rate for each method. The experimental results are shown
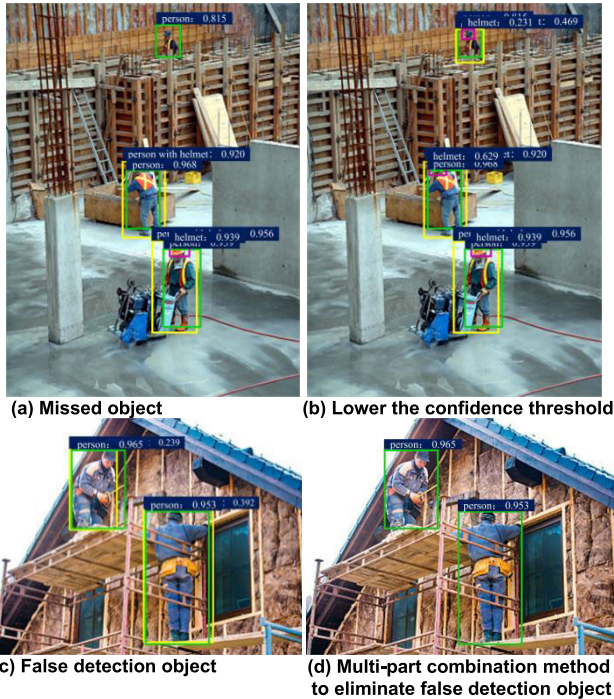
(a) Missed object (b) Lower the confidence threshold

(c) False detection object (d) Multi-part combination method to eliminate false detection object

FIGURE 7. Multi-part combination.

in Table 4. It can be seen that the proposed method is better than the traditional machine learning feature extraction method HOG+SVM, SSD, YOLOv3, RetinaNet and the optimized Faster RCNN algorithm. Compared with the HOG+SVM method, the test accuracy and recall rate of this method are greatly improved. This is because the worker's movements are varied, and the working environment is complex. If there is occlusion and deformation, the HOG gradient feature will be weak, and the target cannot be extracted.

TABLE 4. Comparison of experimental results of different methods.

Methods	Recall/%	Precision/%
HOG+SVM	71.95	79.83
YOLOv3	83.11	85.43
SSD	79.25	83.54
RetinaNet	84.36	86.82
Improved Faster RCNN	90.45	91.42
Our method	91.51	94.56

Compared with the optimized Faster RCNN network, the proposed method integrated a target geometric position information, so the detection accuracy is improved by about 3%. Detecting the relationship between the target and the position of the helmet is feasible for accurate detection of the wearing of the worker's helmet.

G. . DETECTION OF DIFFERENT SCENES AND DIFFERENT IMAGE QUALITY

The testing sets contain images of different scenes and qualities, as shown in Fig. 8.

(a) Multi-objec detection (b) Poor light

(c) Multi-scale detection (d) Partial occlusion

FIGURE 8. Safety helmet wearing detection effect in different scenes.

In Fig. 8(a), our method detected multiple objects with various sizes. Our method has better detection robustness for poor light environment, multi-object and partial occlusion, as shown in Fig. 8(b), (c), and (d). Our method can automatically detect workers wearing safety helmets in different scenarios, showing its robustness.

V. CONCLUSION

To solve small safety helmet objects during the detection process of workers wearing safety helmets, we propose a deep learning object detection framework that integrates the HOEM mechanism and multi-part method. The pyramid method is used to obtain the multi-scale features of the image, and the OHEM mechanism is introduced to select hard samples, and then send them to the network retraining so that the network can learn the hard samples. Through the multi-part combination method, the helmet and the worker's position information are combined to detect the worker wearing the helmet. Experimental results show that the method proposed in this paper effectively improves the accuracy of helmet automatic detection. Besides, it is still robust to low light environments and occlusion images. However, the posture of the workers is different, and our method can only roughly select the relative position of the safety helmet and other parts. In the future, we will solve this problem by attitude estimation.

REFERENCES

[1] O. N. Aneziris, I. A. Papazoglou, M. Mud, M. Damen, L. J. Bellamy, H. J. Manuel, and J. Oh, "Occupational risk quantification owing to falling objects," *Saf. Sci.*, vol. 69, pp. 57–70, Nov. 2014.

[2] *Center for Construction Research and Training. The Construction Chart Book: The U.S. Construction Industry and Its Workers*. 5th ed, CPWR, Silver Spring, MD, USA, 2013.

[3] I. W. H. Fung, Y. Y. Lee, V. W. Y. Tam, and H. W. Fung, "A feasibility study of introducing chin straps of safety helmets as a statutory requirement in hong kong construction industry," *Saf. Sci.*, vol. 65, pp. 70–78, Jun. 2014.

[4] H. Wu and J. Zhao, "An intelligent vision-based approach for helmet identification for work safety," *Comput. Ind.*, vol. 100, pp. 267–277, Sep. 2018.

[5] A. H. M. Rubaiyat, T. T. Toma, M. Kalantari-Khandani, S. A. Rahman, L. Chen, Y. Ye, and C. S. Pan, "Automatic detection of helmet uses for construction safety," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Workshops (WIW)*, Oct. 2016, pp. 135–142.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[7] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Int. Conf. Mach. Learn.*, 2014, p. 647.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 580–587.

[9] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[11] N. Sang and Z. Ni, "Gesture recognition based on R-FCN in complex scenes," *Huazhong Keji Daxue Xuebao*, vol. 45, no. 10, pp. 54–58, 2017.

[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, Oct. 2017, pp. 2999–3007.

[15] R. Silva, K. Aires, T. Santos, K. Abdala, R. Veras, and A. Soares, "Automatic detection of motorcyclists without helmet," in *Proc. 39th Latin Amer. Comput. Conf. (CLEI)*, Oct. 2013, pp. 1–7.

[16] R. R. V. E. Silva, K. R. T. Aires, and R. D. M. S. Veras, "Helmet detection on motorcy-clists using image descriptors and classifiers," in *Proc. 27th SIBGRAPI Conf. Graph., Patterns Images*, Aug. 2014, pp. 141–148.

[17] C. Vishnu, D. Singh, C. K. Mohan, and S. Babu, "Detection of motorcyclists without helmet in videos using convolutional neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 3036–3041.

[18] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[19] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[20] S. Ceri, A. Bozzon, M. Brambilla, E. D. Valle, P. Fraternali, and S. Quarteroni, "An introduction to information retrieval," *Pharmacogenomics J.*, vol. 2, no. 2, pp. 96–102, 2013.

[21] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 761–769.

[22] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional neural networks based fire detection in surveillance videos," *IEEE Access*, vol. 6, pp. 18174–18183, 2018.

[23] F. Yates, "The influence of statistical methods for research workers on the development of the science of statistics," *Amer. Stat. Assoc.*, vol. 46, no. 253, pp. 19–34, 2012.

[24] Y. Wang, W. Song, G. Fortino, L.-Z. Qi, W. Zhang, and A. Liotta, "An experimental-based review of image enhancement and image restoration methods for underwater imaging," *IEEE Access*, vol. 7, pp. 140233–140251, 2019.

[25] J. Hung and A. Carpenter, "Applying faster R-CNN for object detection on malaria images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2018, pp. 1–7.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, vol. 2015, pp. 1–14.

[27] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science, vol. 8689. Cham, Switzerland: Springer, 2014, pp. 818–833.

[28] Y.-H. Byeon and K.-C. Kwak, "A performance comparison of pedestrian detection using faster RCNN and ACF," in *Proc. 6th IIAI Int. Congr. Adv. Appl. Informat. (IIAI-AAI)*, Jul. 2017, pp. 858–863.

NING LI is currently pursuing the Ph.D. degree in computer science and technology with Hohai University. He is also an Associate Professor of software engineering with Changzhou University. His research interests include deep learning, image segmentation, object recognition, image annotation, and so on.

XIN LYU is currently a Lecturer with the College of Computer and Information, Hohai University. He has published more than 60 articles. His research interests include cryptography, network information security, and privacy-preserving theory and technology.

SHOUKUN XU is currently a Professor of software engineering with Changzhou University. His research interests include deep learning and image processing in chemical production.

YARU WANG is currently pursuing the master's degree with Changzhou University, Changzhou, China. Her research interests include deep learning and image processing in chemical production.

YUSHENG WANG is currently pursuing the master's degree with Changzhou University, Changzhou, China. His research interests include deep learning and image processing in chemical production.

YUWAN GU received the Ph.D. degree in agricultural engineering from Jiangsu University, Zhenjiang, China, in 2016. She was a Lecturer of computer science and technology with the School of Information Science and Engineering, Changzhou University. Her research interests include machine learning and image processing.

● ● ●