

Research Article

Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks

Yange Li,¹ Han Wei,¹ Zheng Han ,¹ Jianling Huang,¹ and Weidong Wang^{1,2}

¹School of Civil Engineering, Central South University, Changsha 410075, China

²The Key Laboratory of Engineering Structures of Heavy Haul Railway, Ministry of Education, Changsha 410075, China

Correspondence should be addressed to Zheng Han; zheng_han@csu.edu.cn

Received 8 August 2019; Revised 14 May 2020; Accepted 10 September 2020; Published 19 September 2020

Academic Editor: Antonio Formisano

Copyright © 2020 Yange Li et al. 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.

Visual examination of the workplace and in-time reminder to the failure of wearing a safety helmet is of particular importance to avoid injuries of workers at the construction site. Video monitoring systems provide a large amount of unstructured image data on-site for this purpose, however, requiring a computer vision-based automatic solution for real-time detection. Although a growing body of literature has developed many deep learning-based models to detect helmet for the traffic surveillance aspect, an appropriate solution for the industry application is less discussed in view of the complex scene on the construction site. In this regard, we develop a deep learning-based method for the real-time detection of a safety helmet at the construction site. The presented method uses the SSD-MobileNet algorithm that is based on convolutional neural networks. A dataset containing 3261 images of safety helmets collected from two sources, i.e., manual capture from the video monitoring system at the workplace and open images obtained using web crawler technology, is established and released to the public. The image set is divided into a training set, validation set, and test set, with a sampling ratio of nearly 8:1:1. The experiment results demonstrate that the presented deep learning-based model using the SSD-MobileNet algorithm is capable of detecting the unsafe operation of failure of wearing a helmet at the construction site, with satisfactory accuracy and efficiency.

1. Introduction

Construction is a high-risk industry where construction workers tend to be hurt in the work process. Head injuries are very serious and often fatal. According to the accident statistics released by the state administration of work safety from 2015 to 2018, among the recorded 78 construction accidents, 53 events happened owing to the fact that the workers did not wear safety helmets properly, accounting for 67.95% of the total number of accidents [1].

In safety management at the construction site, it is essential to supervise the safety protective equipment wearing condition of the construction workers. Safety helmets can bear and disperse the hit of falling objects and alleviate the damage of workers falling from heights. Construction workers tend to ignore safety helmets because of weak safety awareness. At the construction site, workers that wear safety helmets improperly are much more likely to be injured.

Traditional supervision of the workers wearing safety helmets on construction sites often requires manual work [2]. There are problems such as a wide range of operations and difficult management of site workers. These factors make manual supervision difficult and inefficient and it is difficult to track and manage the whole workers at the construction sites accurately in real time [3]. Hence, it is hard to satisfy the modern requirement of construction safety management only relying on the traditional manual supervision. In this context, it remains a significant issue to study on the automatic detection and recognition of safety helmets wearing conditions.

The automatic monitoring method can contribute to monitoring the construction workers and confirm the safety helmet wearing conditions at the construction site. In particular, considering that the traditional manual supervision of the workers is often costly, time-consuming, error-prone, and not sufficient to satisfy the modern requirements

of construction safety management, the automatic supervision method can be beneficial to real-time on-site monitoring.

In this paper, based on the previous studies on computer vision-based object detection, we develop a deep learning-based method for the real-time detection of safety helmet at the construction site. The major contributions are as follows: (1) a dataset containing 3261 images of safety helmets collected from two sources, i.e., manual capture from the video monitoring system at the workplace and open images obtained using web crawler technology, is established and released to the public. (2) The SSD-MobileNet algorithm that is based on convolutional neural networks is used to train the model, which is verified in our study as an alternative solution to detect the unsafe operation of failure of wearing a helmet at the construction site. The article is organized as follows. Section 2 gives a brief description of the related work. Section 3 describes the methodology of the research. Section 4 introduces the construction of the database. Section 5 reports the experiment results of the study. Sections 6 and 7 discuss the pros and cons of the study and conclude the paper.

2. Literature Review

2.1. Related Research into the Safety Helmets Detection. At present, previous studies of safety helmets detection can be divided into three parts, sensor-based detection, machine learning-based detection, and deep learning-based detection. Sensor-based detection usually locates the safety helmets and workers (Kelm et al. [4], Torres et al. [5]). The methods usually use the RFID tags and readers to locate the helmets and workers and monitor how personal protective equipment is worn by workers in real time. Kelm et al. [4] designed a mobile Radio Frequency Identification (RFID) portal for checking personal protective equipment (PPE) compliance of personnel. However, the working range of the RFID readers is limited and the RFID readers can only suggest that the safety helmets are close to the workers but unable to confirm that the safety helmets are being properly worn.

Up to date, machine learning-based object detection technologies are widely used in many domains for its powerful object detection and classification capacity (e.g., Rubaiyat et al. [6], Shrestha et al. [7], Waranusast et al. [8], Dounghala et al. [9], Jia et al. [10], and Li et al. [11]). Remarkable studies are made by Rubaiyat et al. [6], who proposed an automatic detection method to obtain the features of construction workers and safety helmets and detect safety helmets. The method combines the frequency domain information of the image with the histogram of oriented gradient (HOG) and the circle Hough transform (CHT) extractive technique to detect the workers and the helmets in two steps. The detection methods based on machine learning can detect safety helmets accurately and precisely under various scenarios but also have some drawbacks. Sometimes the method can only detect safety helmets with a specific color and it is difficult to distinguish the hats with similar color and shape to the safety helmets.

Moreover, the method cannot detect faces and safety helmets thoroughly under some circumstances; for example, some workers do not turn their faces towards the camera at the construction site.

2.2. Deep Learning-Based Object Detection. The above-mentioned methods are commonly based on traditional machine learning to detect and classify the helmets and choose features artificially with a strong subjectivity, a complex design process, and poor generalization ability. In recent years, with the rapid development of deep learning technology, the object detection algorithm turns to the one based on convolutional neural networks with a great promotion of speed and accuracy (e.g., Wu et al. [12]).

The methods construct convolutional neural networks with different depths to detect safety helmets. Some other strategies such as multiscale training, increasing the number of anchors and introducing the online hard example mining, are added to increase the detection accuracy (e.g., Xu et al. [13]). However, these methods have some limitations in the preprocessing aspects of image sharpness, object proportion, and the color difference between background and foreground.

Deep learning-based methods are very potential for the purpose of people's unsafe behavior identification. Many previous studies have presented a solution to this topic. Remarkable studies include the following: Ding et al. [14] developed a hybrid deep learning model that integrates a convolution neural network (CNN) and long short-term memory (LSTM) that automatically recognizes workers' unsafe actions. The results demonstrated that the model can precisely detect safe and unsafe actions conducted by workers on-site. However, some behaviors cannot be recognized owing to the lack of data, the small sample size used for training, and the limited number of unsafe actions that were considered. Fang et al. [15] proposed a novel deep learning-based framework to check whether a site worker is working within the constraints of their certification. The framework includes key video clips extraction, trade recognition, and worker competency evaluation. Results demonstrate that the proposed framework offers an effective and feasible solution to detect noncertified work. However, some workers cannot be detected when the workers' faces hardly appear or are obstructed by the safety helmets or other equipment. Also, the worker close to the camera failed to be recognized. Fang et al. [16] integrated a Faster R-CNN and a deep CNN to detect the presence of a worker from images and the harness, respectively, which can identify whether workers wear safety harness while working at heights or not. The research is limited by the restricted activities working at heights and the dataset size. Fang et al. [17] developed a computer vision-based approach which uses a Mask R-CNN to detect people and recognize the relationship between people and concrete supports to identify unsafe behaviors. The study has some restrictions: it focuses on a limited number of activities related to the construction of deep foundation-pits. Luo et al. [18] proposed an increased CNN that integrates Red-Green-Blue,

optical flow, and gray stream CNNs to monitor and assess workers' activities associated with installing reinforcement at the construction site. The research is limited by occlusions, insufficient knowledge of a time series of actions definition, and lack of a large-scale database.

Considering its excellent ability to extract features, in the paper, we use the convolutional neural network (CNN) to build a safety helmet detection model. Automatic detection of safety helmets worn by construction workers at the construction site and timely warning of workers without helmets can largely avoid accidents caused by workers wearing safety helmets improperly. The designed CNN is trained using the TensorFlow framework. The contributions of the research include a deep learning-based safety helmet detection model and a safety helmet image dataset for further research. The model provides an opportunity to detect the helmets and improve safety management.

Deep learning-based methods are commonly used to detect unsafe behaviors on-site. Nevertheless, many traditional measures of safety helmet detection are commonly sensor-based and machine-based, thus limited by problems such as sensor failure over long distances, the manual and subjective features choice, and the chaotic scene interference. Based on the previous studies, we present a deep learning-based method to detect the safety helmets in the workplace, which is supposed to avoid the abovementioned limitations.

3. Methodology

A convolutional neural network (CNN) is a multilayer neural network. It is a deep learning method designed for image recognition and classification tasks. It can solve the problems of too many parameters and difficult training of the deep neural networks and can get better classification effects. The structure of most CNNs consists of input layer-convolutional layer (Conv layer)-activation function-pooling layer-fully connected layer (FC layer). The main characteristics of CNNs are local connectivity and parameter sharing in order to reduce the number of parameters and increase the efficiency of detection.

The Conv layer and the pooling layer are the core parts, and they can extract the object features. Often, the convolutional layer and the pooling layer may occur alternately. The Conv layers can extract and reinforce the object features. The pooling layers can filter multiple features, remove the unimportant features, and compress the features. The activation layers use nonlinear activation functions to enhance the expression ability of the neural network models and can solve the nonlinear problems effectively. The FC layers combine the data features of objects and output the feature values. By this means the CNNs can transfer the original input images from the original pixel values to the final classification confidence layer by layer.

In order to better extract the object features and classify the objects more precisely, Hinton et al. [19] proposed the concept of deep learning which is to learn object features from vast amounts of data using deep neural networks and then classify new objects according to the learned features. Deep

learning algorithm based on convolutional neural networks has achieved great results in object detection, image recognition, and image segmentation. Girshick et al. [20] proposed R-CNN detection framework (region with CNN features) in 2014. Many models based on R-CNN were proposed after that including SPP-net (spatial pyramid pooling network) [21], Fast R-CNN (fast region with CNN features) [22], and Faster R-CNN (faster region with CNN features) [23].

Classification-based CNN object detection algorithms such as Faster R-CNN are widely used methods. However, the detection speed is slow and cannot detect in real time. Regression-based detection algorithms are becoming increasingly important. Redmon et al. [24] proposed YOLO (You Only Look Once) algorithm in 2016. At the end of 2016, Liu et al. [25] combined the anchor box of Faster R-CNN with the bounding box regression of YOLO and proposed a new algorithm SSD (Single Shot MultiBox Detector) with higher detection accuracy and faster speed.

Although the SSD algorithm is not capable of the highest accuracy, the detection speed of the SSD algorithm is much faster and comparable to the YOLO algorithm and the precision can be higher than that of the YOLO algorithm when the sizes of the input images are smaller. While the Faster R-CNN algorithm tends to lead to more accurate models, it is much slower and requires at least 100 ms per image [26]. Therefore, considering the real-time detection requirements, the SSD algorithm is chosen in the research. In order to reduce greatly the calculation amount and model thickness, the MobileNet [27] model is added. Therefore, in the paper, the SSD-MobileNet model is selected to detect safety helmets worn by the workers.

The SSD algorithm is based on a feed-forward convolutional network to produce bounding boxes of fixed sizes and generate scores for the object class examples in the boxes. A nonmaximum suppression method is used to predict the final results.

The early network layers of the SSD model are called the base network, based on a standard framework to classify the image. The base network is truncated before the classification layers, and the convolutional layers are added at the end of the truncated base network. The sizes of the convolutional feature maps decrease progressively to predict the detections at multiple scales.

The SSD algorithm sets a series of fixed and different size default boxes on the cell of each feature map as shown in Figure 1. Each default box predicts two kinds of detections. One is the location of bounding boxes including 4 offsets (cx, cy, w, h) , which represent, respectively, x and y coordinates of the center of the bounding box and the width and height of the bounding box; the other is the score of each class. If there are C classes of the objects, the SSD algorithm predicts a total of $C+1$ score including the score of the background.

The setting of default boxes can be divided into two aspects: size and aspect ratio. The sizes of the default boxes in every feature map will be calculated as follows:

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m-1}k - 1, \quad k \in [1, m]. \quad (1)$$

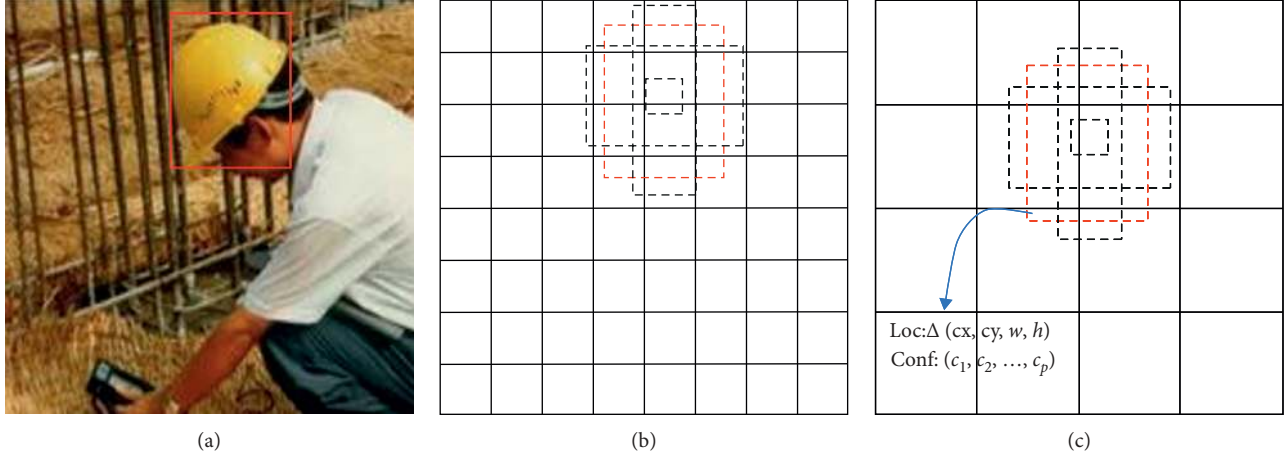


FIGURE 1: Detection process. (a) Image with GT box. (b) 8×8 feature map. (c) 4×4 feature map.

In the formula, S_{\min} is 0.2, S_{\max} is 0.95. The aspect ratios are set different for the default boxes, expressed as $a_r \in \{1, 2, 3, (1/2), (1/3)\}$. The width of the default boxes is calculated as follows:

$$w_k^a = S_k \sqrt{a_r}. \quad (2)$$

The height of the default boxes is calculated as follows:

$$h_k^a = \frac{S_k}{\sqrt{a_r}}. \quad (3)$$

When the aspect ratio is 1, a default box size is added: $S'_k = \sqrt{S_k S_{k+1}}$. Therefore, there are six default boxes of different sizes for each feature cell.

The default boxes will be matched to the ground truth boxes. Each ground truth box can choose default boxes of different locations, aspect ratios, and sizes to match. The ground truth box will be matched to the default box with the best Jaccard overlap. The Jaccard overlap is also called the IoU (Intersection over Union), or the Jaccard similarity coefficient. The IoU is the ratio of the intersection and the union of the default box to the ground truth box. The schematic illustration of IoU is shown in Figure 2:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \quad (4)$$

After the match process, most default boxes are negative examples which do not match the objects but the background. Therefore, the SSD algorithm uses the hard negative mining strategy to avoid the significant imbalance between the positive and negative training examples. The default boxes are ranked in the descending order according to the confidence error and the top ones are chosen to be the negative examples so the ratio between the negative and positive examples is almost 3:1.

The SSD algorithm defines the total loss function as the weighted sum between localization loss and confidence loss:

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)). \quad (5)$$

In the prediction process, the object classes and confidence scores will be confirmed according to the maximum class confidence score and the prediction box that belongs to the background will be filtered out. The prediction boxes with confidence scores below 0.5 are also removed. As for the left boxes, the location will be obtained according to the default boxes. The prediction boxes are ranked in the descending order according to the confidence score and the top ones are retained. Finally, the nonmaximum suppression algorithm is used to filter out the prediction boxes with higher but not the highest IOU and the left prediction boxes are the results.

Although the SSD algorithm performs well in the speed and the precision, the large model and a large amount of calculation make the training speed a bit slow. Therefore, the base network of the SSD model is replaced by the MobileNet model to reduce the calculation amount and the model thickness. In the paper, the SSD-MobileNet model is chosen to detect the safety helmets worn by the workers.

The core concept of the MobileNet model is the factorization of the filters. The main function is to reduce the calculation amount and the network parameters. The model is used to factorize a standard convolution into a depthwise convolution and a pointwise convolution. The model is shown in Figure 3.

The model also introduces two hyperparameters: width multiplier and resolution multiplier to reduce the channel numbers and reduce the image resolutions, respectively. The network model with less calculation amount can be built. Hence, using the SSD-MobileNet model can reduce the thickness of the SSD model effectively.

4. Database

The data required for the experiment were collected by the author. Since there are few object detection applications of safety helmets using deep learning and there is no off-the-shelf safety helmets dataset available, part of the experimental data was collected using web crawler technology, making full use of network resources. By using several

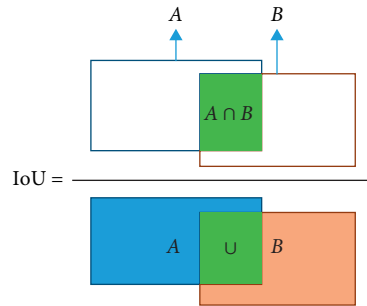


FIGURE 2: Schematic illustration of IoU.

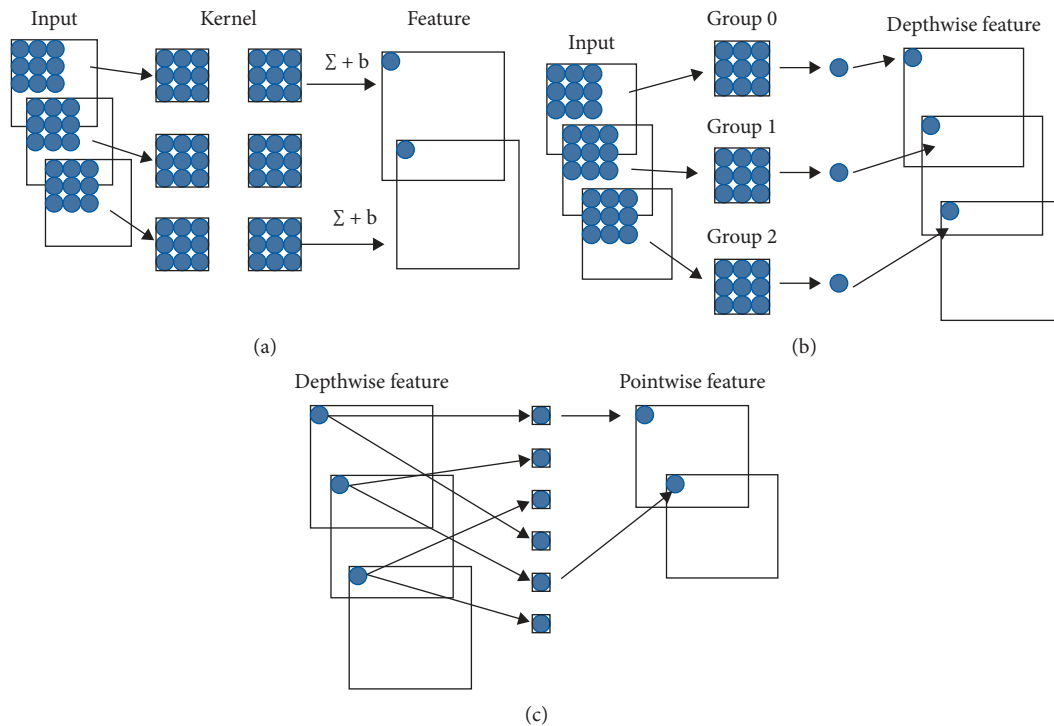


FIGURE 3: Schematic illustration of MobileNet separate convolution. (a) Standard convolution. (b) Depthwise convolution. (c) Pointwise convolution.

keywords, such as “workers wear safety helmets” and “workers on the construction site,” python language is used to crawl relevant pictures on the Internet.

However, the quality of the crawled images varies greatly. There are problems that there is an only background and no objects in some images, the size of the safety helmet is small, and the shape is blurred. Therefore, images were also collected manually besides web crawling. 3500 images were collected in total. The images that did not contain safety helmets, duplicate images, and the images that are not in the RGB three-channel format were eliminated and 3261 images were left, forming the safety helmet detection dataset. Some images in the dataset are shown in Figure 4. To increase the detection effect of the safety helmet detection model in detecting helmets with different directions and brightness in images, the image dataset was preprocessed such as rotation, cutting, and zooming.

Then, the samples in the dataset are divided into three parts randomly: training set, validation set, and test set. Commonly,

a ratio of 6:2:2 is suggested for dividing the training set, validation set, and test set in the previous machine learning studies, such as the course of Andrew Ng from deeplearning.ai. In deep learning, the dataset scale is much larger and the validation and test sets tend to be a smaller percentage of the total data which are commonly less than 20% or 10%. In this sense, an adequate ratio of 8:1:1 according to the previous experience is adopted in our study. The numbers of the three sets are 2769, 339, and 153, respectively. All the images that contained safety helmets were manually pre-labeled, using the open-source tool LabelImage (available in <https://github.com/tzutalin/labelImg>). In each labeled image, the sizes and the locations of the object are recorded (Figure 5).

5. Results

In the paper, the open-source TensorFlow framework is chosen to train the model. The pretrained

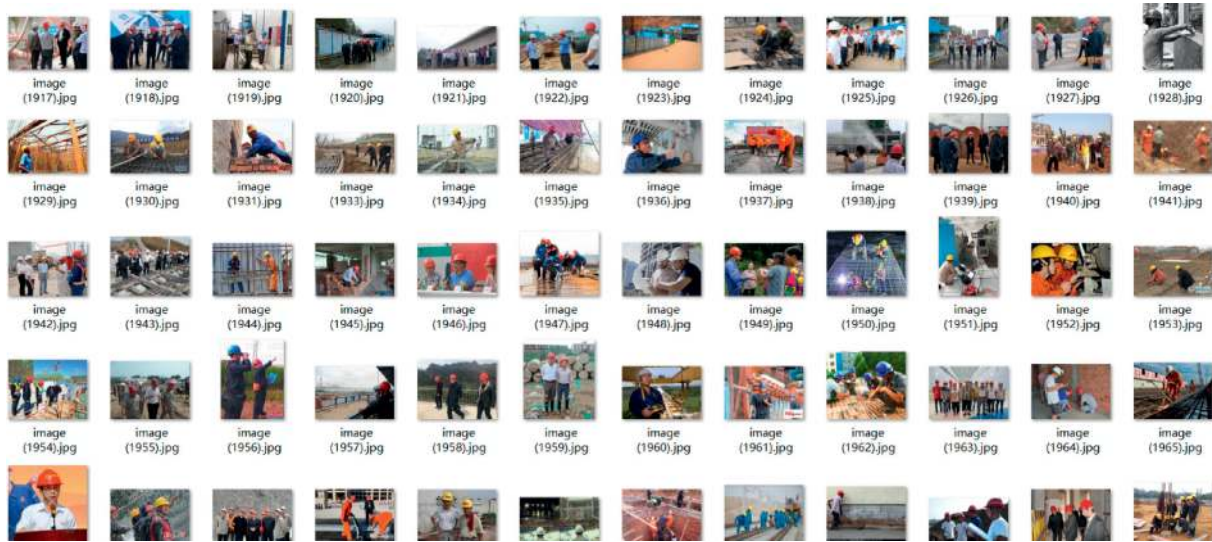


FIGURE 4: Image dataset.



(a)

Filename	Width	Height	Class	Xmin	Ymin	Xmax	Ymax
Image (1).jpg	315	207	Helmet	100	6	152	51
Image (1).jpg	315	207	Helmet	162	35	200	61
Image (1).jpg	315	207	Helmet	206	53	229	69
Image (1).jpg	315	207	Helmet	238	45	258	60
Image (1).jpg	315	207	Helmet	1	30	33	68
Image (1).jpg	315	207	Helmet	46	45	83	74

(b)

FIGURE 5: (a) Manual labeling. (b). Data recording.

SSD_mobilenet_v1_COCO model with the COCO dataset is used to learn the characteristics of the safety helmet in the built dataset to reduce the training time and save the computing resources. The initial weights and the parameter values of our own model are the same as the SSD_mobilenet_v1_COCO model. Finally, the weights and the parameter values of the safety helmet detection model are trained and obtained through the training process.

Among the 3261 images, 2769 images were divided into the training set, 339 images were divided into the validation set, and 153 images were divided into the test set. The training set is used to train the model or to determine the parameters of the model. The validation set is used to adjust the hyperparameters of the model and to evaluate the capacity of the model preliminarily. The test set is used to evaluate the generalization ability of the final model [28].

In the course of training, the change of the mean average precision (mAP) and the loss function during training was recorded by TensorBoard. As a measure index, the mean average precision (mAP) [29] is generally used in the field of object detection. Figure 6 illustrated that the mean Average

Precision shows an overall upward trend, and the trend has ups and downs and is not a steady rise. When training rounds up to 50,000, the mean average precision of the detection model is 36.82%. Figure 7 shows that the total loss values decrease slowly at the beginning of the training and converge at the end of the training. The values of the loss function are the differences between the true value and the predicted value in general speaking. The change in the values of the loss function represents the training process of the model. The smaller the values are, the better the model is trained. The convergence of the loss functions demonstrates that the training of the model is completed. Hence, loss functions mainly influence the training process but not detection. Figures 8(a) to 8(c) show the variation of the classification loss function, localization loss function, and regularization loss function against the steps. Figures 8(a) and 8(b) demonstrate that the values of the classification loss function decrease slowly at first and, then, decrease rapidly when training rounds up to nearly 7,000; the values of the localization loss function decrease rapidly at first and converge at the end of the training. The convergence of the loss functions demonstrates that the training of the model is completed.

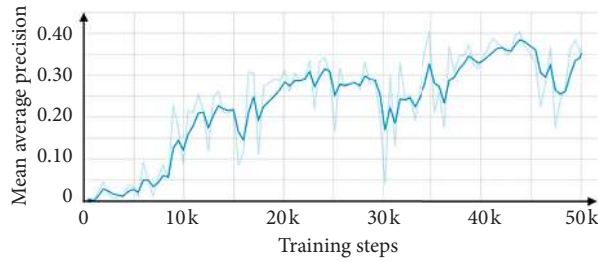


FIGURE 6: Mean average precision change during training.

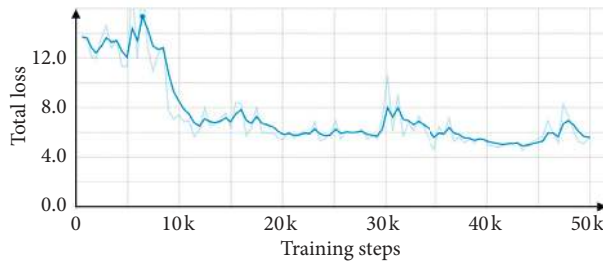


FIGURE 7: Total loss change during training.

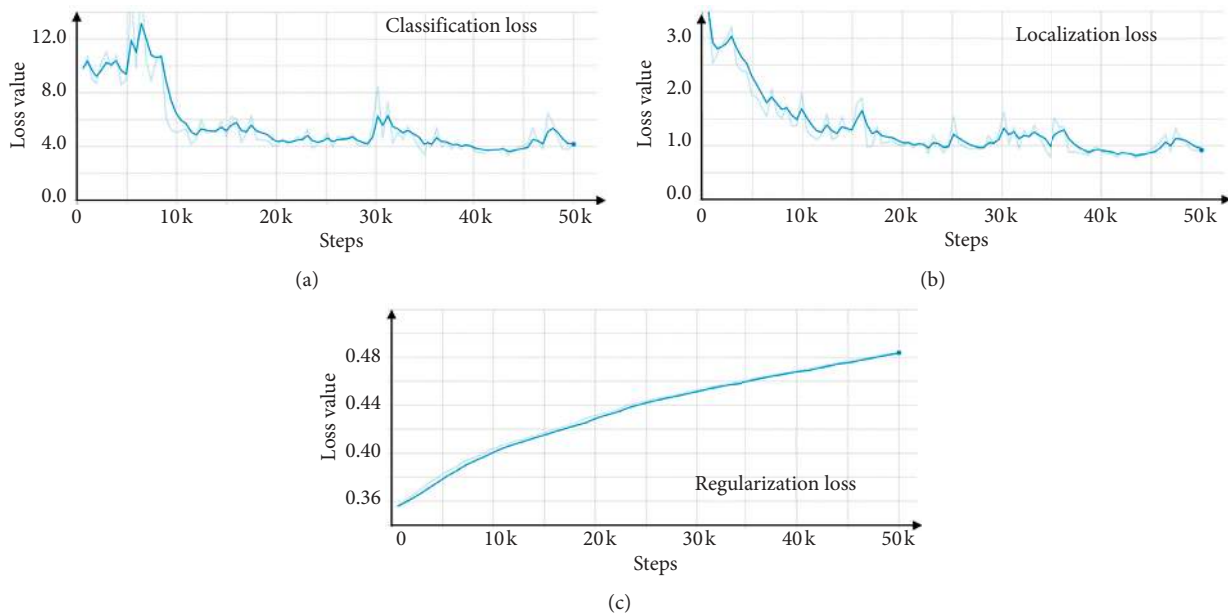


FIGURE 8: Partial loss function value change. (a) Classification loss. (b) Localization loss. (c) Regularization loss.

After the model was trained, it was used to validate the collected validation set by using the Spyder software. The 153 images of the validation set were input into the model and the detected images were output. The output images show the predicted labels and the confidence scores of safety helmets. Some validation results are shown in Figure 9.

The precision and recall are the commonly used metrics to evaluate the performance and reliability of the trained model. Precision is the ratio of true positive (TP) to true positive and false positive (TP + FP). TP + FP is the number of helmets detected. Recall is the ratio of true positive (TP) to true positive and false negative (TP + FN). TP + FN means

the actual number of helmets. There are 250 true positive objects, 12 false positive objects, and 73 false negative objects in the detected images. The precision of the trained model is 95% and the recall is 77%, which demonstrates that the proposed method performs well in safety helmet detection.

As the pictures above show, the probabilities of recognizing the safety helmets worn by workers as safety helmets are more than 80%. However, the output images of the model demonstrate some errors in the detection model. For example, it is hard for the model to detect the safety helmets of small sizes or large rotation angles. It is possible to recognize the objects of the same colors in the images as the



FIGURE 9: Validation results.

safety helmets. When the illumination intensity of the construction site in the images and the objects are not clear, the safety helmets are difficult to be recognized. That suggests the detection model established in the paper is not accurate enough.

As shown in Figure 10(a), the probability predicted by the model is 98%, but the probability of recognizing the background as safety helmets is 78%. This fake detection generates false positive. This is a case of false detection which predicted the false object as correct. The case of Figure 10(c) is the same as the first one. In Figure 10(b), the red helmet is missed and this is a case of false negative. The errors occur because of the interference of the complex background, the limitation of the number of the image dataset, and the safety helmets proportion in the images. In order to improve the performance of the model, some measures must be taken such as increasing the number of the image dataset and adding the pre-processing operations of the images. Besides the above measures, ameliorating the nonmaximum suppression algorithm, adjusting the parameters and weights, and so forth can also be a great solution to reduce the false positives.

In summary, there are several detection errors of the model. (1) The hats with the same shapes and colors or the background are recognized mistakenly as the safety helmets. (2) The safety helmets of incomplete shapes and small sizes are hard to be recognized. (3) The two or more helmets that are very close to each other are often recognized as a safety helmet.

6. Discussion

6.1. Effect of the Presented Method. The proposed automatic detection method based on deep learning to detect safety helmets worn by workers provides an effective opportunity to improve safety management on construction sites. Previous studies have demonstrated the effectiveness of locating the safety helmets and workers and detecting the helmets.

However, most of the studies have limitations in practical application. Sensor-based detection methods have a limited read range of readers and cannot be able to confirm the position relationship between the helmets and the workers. The machine learning-based detection methods choose features artificially with a strong subjectivity, a complex design process, and poor generalization ability. Therefore, the study proposed a method based on deep learning to detect safety helmets automatically using convolutional neural networks. The experimental results have suggested the effectiveness of the proposed method.

In the paper, the SSD-MobileNet algorithm is used to build the model. A dataset of 3261 images containing various helmets is trained and tested on the model. The experimental results demonstrate the feasibility of the model. And the model does not require the selection of handcraft features and has a good capacity of extracting features in the images. The high precision and recall show the great performance of the model. The proposed model provides an opportunity to detect the helmets and improve construction safety management on-site.

6.2. Limitations. However, the detection model has a poor performance when the images are not very clear, the safety helmets are too small and obscure, and the background is too complex as shown in Figure 10. Moreover, the presented model is limited by the problems that some images of the dataset are less in quantity; the preprocessing operations of the images are confined to rotation, cutting, and zooming; the manual labeling is not comprehensive and may miss some objects. In some extreme cases, for example, only part of the head is visible and the safety helmet is obstructed, the model cannot detect the helmets accurately. This is the common limitation of the-state-of-art algorithms. Due to the above reasons, the detection performance is not good enough and there are some detection errors.



FIGURE 10: Detection errors.

The algorithm we use emphasizes the real-time detection and fast speed. However, the accuracy of the detection is also quite important and the performance needs to be improved. Hence, in the ongoing studies, we are working at the expansion and improvement of the dataset in order to solve the problems of inadequate data with poor quality. More comprehensive preprocessing operations should be done to improve the performance of the model.

7. Conclusions

The paper proposed a method for detecting the wearing of safety helmets by the workers based on convolutional neural networks. The model uses the SSD-MobileNet algorithm to detect safety helmets. Then, a dataset of 3261 images containing various helmets is built and divided into three parts to train and test the model. The TensorFlow framework is chosen to train the model. After the training and testing process, the mean average precision (mAP) of the detection model is stable and the helmet detection model is built. The experiment results demonstrate that the method can be used to detect the safety helmets worn by the construction workers at the construction site. The presented method offers an alternative solution to detect the safety helmets and improve the safety management of the construction workers at the construction site.

Data Availability

The database used to train the CNN of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Professor Y.G. Li collected the dataset and wrote the manuscript. Professor Z. Han designed the study. H. Wei trained and optimized the model. Professor J.L. Huang and Professor W.D. Wang participated in the analysis of the results. All the authors discussed the results and commented on the manuscript.

Acknowledgments

This study was financially supported by the National Key R&D Program of China (Grant no. 2018YFC1505401); the National Natural Science Foundation of China (Grant no. 52078493); the Natural Science Foundation of Hunan (Grant no. 2018JJ3638); and the Innovation Driven Program of Central South University (Grant no. 2019CX011). These financial supports are gratefully acknowledged.

References

- [1] X. Chang and X. M. Liu, "Fault tree analysis of unreasonably wearing helmets for builders," *Journal of Jilin Jianzhu University*, vol. 35, no. 6, pp. 67–71, 2018.
- [2] Z. Y. Wang, *Design and Implementation of Detection System of Warning Helmets Based on Intelligent Video Surveillance*, Beijing University of Posts and Telecommunications, Beijing, China, 2018.
- [3] H. Zeng, *Research on Intelligent Helmets System for Engineering Construction*, Harbin Institute of Technology, Harbin, China, 2017.
- [4] A. Kelm, L. Laußat, A. Meins-Becker et al., "Mobile passive radio frequency identification (RFID) portal for automated and rapid control of personal protective equipment (PPE) on construction sites," *Automation in Construction*, vol. 36, pp. 38–52, 2013.
- [5] S. Barro-Torres, T. M. Fernández-Caramés, H. J. Pérez-Iglesias, and C. J. Escudero, "Real-time personal protective equipment monitoring system," *Computer Communications*, vol. 36, no. 1, pp. 42–50, 2012.
- [6] A. H. M. Rubaiyat, T. T. Toma, M. Kalantari-Khandani et al., "Automatic detection of helmet uses for construction safety," in *Proceedings of the 2016 IEEE ACM International Conference on Web Intelligence Workshops (WIW)*, ACM, Omaha, NE, USA, October 2016.
- [7] K. Shrestha, P. P. Shrestha, D. Bajracharya, and E. A. Yfantis, "Hard-hat detection for construction safety visualization," *Journal of Construction Engineering*, vol. 2015, Article ID 721380, 8 pages, 2015.
- [8] R. Waranusast, N. Bundon, V. Timtong, C. Tangnoi, and P. Pattanathaburt, "Machine vision techniques for motorcycle safety helmet detection," in *Proceedings of the Image & Vision Computing New Zealand*, IEEE, Wellington, New Zealand, November 2013.
- [9] P. Doungmala and K. Klubsuwan, "Helmet wearing detection in Thailand using haar like feature and circle hough transform

- on image processing,” in *Proceedings of the IEEE International Conference on Computer & Information Technology*, IEEE, Nadi, Fiji, December 2016.
- [10] J. S. Jia, Q. J. Bao, and H. M. Tang, “Method for detecting safety helmet based on deformable part model,” *Application Research of Computers*, vol. 33, no. 3, pp. 953–956, 2016.
- [11] K. Li, X. G. Zhao, J. Bian, and M. Tan, “Automatic safety helmet wearing detection,” 2018, <https://arxiv.org/abs/1802.00264>.
- [12] H. Wu and J. Zhao, “Automated visual helmet identification based on deep convolutional neural networks,” in *Proceedings of the 13th International Symposium on Process Systems Engineering (PSE 2018)*, vol. 44, pp. 2299–2304, San Diego, CA, USA, July 2018.
- [13] S. Xu, Y. Wang, Y. Gu, N. Li, L. Zhuang, and L. Shi, “Safety helmet wearing detection study based on improved faster RCNN,” *Application Research of Computers*, vol. 37, no. 3, pp. 901–905, 2019.
- [14] L. Ding, W. Fang, H. Luo, P. E. D. Love, B. Zhong, and X. Ouyang, “A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory,” *Automation in Construction*, vol. 86, pp. 118–124, 2018.
- [15] Q. Fang, H. Li, X. Luo et al., “A deep learning-based method for detecting non-certified work on construction sites,” *Advanced Engineering Informatics*, vol. 35, pp. 56–68, 2018.
- [16] W. Fang, L. Ding, H. Luo, and P. E. D. Love, “Falls from heights: a computer vision-based approach for safety harness detection,” *Automation in Construction*, vol. 91, pp. 53–61, 2018.
- [17] W. Fang, B. Zhong, N. Zhao et al., “A deep learning based approach for mitigating falls from height with computer vision: convolutional neural network,” *Advanced Engineering Informatics*, vol. 39, pp. 179–177, 2019.
- [18] H. Luo, C. Xiong, W. Fang, P. E. D. Love, B. Zhang, and X. Ouyang, “Convolutional neural networks: computer vision-based workforce activity assessment in construction,” *Automation in Construction*, vol. 94, pp. 282–289, 2018.
- [19] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2014.
- [22] R. Girshick, “Fast R-CNN. computer science,” in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, December 2015.
- [23] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal network,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 6, p. 1137, 2016.
- [24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: unified, real-time object detection,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.
- [25] W. Liu, D. Anguelov, D. Erhan et al., “Single shot multibox detector,” in *Proceedings of the ECCV 2016: Computer Vision- ECCV 2016*, vol. 9905, pp. 21–37, Springer, Amsterdam, The Netherlands, October 2016.
- [26] J. Huang, V. Rathod, C. Sun et al., “Speed/accuracy trade-offs for modern convolutional object detectors,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3296–3297, Honolulu, HI, USA, July 2017.
- [27] A. G. Howard, M. Zhu, B. Chen et al., “MobileNets: efficient convolutional neural networks for mobile vision applications,” 2017, <https://arxiv.org/abs/1704.04861>.
- [28] J. Grum, “Book review: pattern recognition and neural networks by B.D. Ripley,” *International Journal of Microstructure and Materials Properties*, vol. 4, no. 1, p. 146, 2009.
- [29] M. Everingham, L. V. Gool, and J. Winn, “The pascal visual object classes (VOC) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.