Harness-Wearing Detection of Construction Workers Based on Deep Learning

Sensen Fan, Jinshan Liu, and Yujie Lu(✉)

College of Civil Engineering, Tongji University, Shanghai, China
Lu6@tongji.edu.cn

Abstract. The death and injury rate of the construction industry is higher than the average level of other industries, and falls from heights account for a large share of the accidents. The automatic monitor of the harness-wearing condition of construction workers can alleviate this problem, but the traditional method such as wearing sensor equipment has many disadvantages, and previous research which used the computer vision methods rarely discussed the automatic monitor of harness-wearing under a specific dangerous scene. In this research, we attempted to analyze the effect of the automatic monitor of the harness-wearing condition using the latest computer vision technology and the feasibility of applying it in a specific scene. First, we set a scene in construction that the construction workers working on the mobile lifting platform (mlp) are detected to need to wear a harness, and we created a dataset about the worker, mlp, and harness for this research. Then we used an objects detection algorithm (YOLOv5) as a technical tool for experimental study, which showed that the mAP of the model was greater than 0.97, and the detection speed was between 9 ms/fps and 15 ms/fps, which met the real-time detection needs in a construction site. Besides, we added conditional detection to detect whether the worker needs to wear a harness and whether they are wearing a harness based on the position relation output on the images. The research in this paper presents a method to detect harness-wearing automatically in a specific scene of construction and shows that applying computer vision technology in specific construction activities has been feasible and valuable.

Keywords: Construction safety · Computer vision · Deep learning · Harness-wearing detection

1 Introduction

The construction industry is essential in the world, but it is also known for its dangers. Riza et al. [1] stated that the construction industry employs about 7% of the world's workforce but is responsible for 30–40% of fatal injuries. Moreover, in many countries, the mortality and incidence rate of the construction industry is far greater than the average level of the whole industry. These facts show that special attention needs to be paid to the safety of the construction industry. Meanwhile, the Ministry of Housing and Urban-Rural Development of China reported that [2], there were 773 production safety accidents of housing and municipal engineering in 2019, with 904 deaths, an increase

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
H. Guo et al. (Eds.): CRIOCM 2021, LNOR, pp. 147–156, 2022.
https://doi.org/10.1007/978-981-19-5256-2_13

of 39 accidents, and 64 deaths compared with the number of accidents in 2018. The number of safety accidents is significant and has an upward trend, and the situation is still grim. And falls from heights occupy the largest share of accident types, accounting for 53.69%. An important reason for falls from heights is that workers are not wearing a harness when they need it.

Heinrich [3] analyzed a large number of data and concluded that 88% of the accidents were caused by people's unsafe behaviors, 10% by dangerous environmental conditions, and 2% by other factors of force majeure. A certain degree of supervision and training of construction workers can help workers get used to correctly dealing with unsafe behaviors and help to reduce the occurrence of accidents. Traditional supervision measures include security patrol, video surveillance, and other measures, but such supervision measures cost more time and workforce. The supervision efficiency is low, and the scope of supervision is limited. In recent years, with the development of new technologies, especially computer vision technology, it is possible to automate the supervision of unsafe behavior of workers. At the same time, the monitoring camera widely used in the construction site can provide a large number of low-cost information resources, convenient for the application of computer vision technology in the construction site.

To reduce the occurrence of falls from heights, it is necessary to inspect and urge workers to wear a harness when they are in a dangerous condition. Therefore, this paper conducted an experimental study on the use of computer vision technology based on deep learning to automatically detect the wearing condition of harness in a certain scene, and the scene was set that the construction workers working on the mobile lifting platform (mlp) are detected to need to wear a harness.

2 Literature Review

A harness is a kind of PPE (personal protective equipment). The automatic monitoring of PPE wearing condition of construction workers is mainly realized by two methods. The first is wearing sensor equipment. Analyzing the sensor's signal to determine whether the workers are wearing the PPE, such as Kelm A [4] used on automatic identification (ID) of the existence of business and information technology (IT) to design the mobile radio frequency identification (RFID) portal, to test whether staff wearing PPE, place the door at the entrance or the construction site, And by embedding or pasting a low cost passive RFID tag on the PPE, automated site access, time logging, and compliance control can be performed. The limitation of such studies is that they can only detect whether workers are carrying PPE when they enter the site and cannot monitor in real-time whether workers are wearing PPE at work. In addition, Barro-Torres S et al. [5] constructed a network architecture for real-time transmission of information from PPE monitoring sensors worn by workers. Unlike Kelm A et al., who set detectors at the site entrance, they integrated sensors on workers' clothing to achieve continuous monitoring. However, this method cannot determine whether workers are wearing PPE or just beside the PPE, which still has significant limitations. Dong S et al. [6] determined whether the workers were correctly wearing the PPE by installing a pressure sensor and monitoring its pressure information through Bluetooth. In general, existing technology makes it difficult to determine whether a worker is wearing a PPE device accurately. The need to set up a large number of sensor devices increases the cost of automatic monitoring.

The second method is using the technology of computer vision. Many cameras used in construction sites provide abundant and low-cost information resources for the application of computer vision technology in this field. Han S U et al. [7] analyzed workers' unsafe behaviors through deption-based cameras such as Kinect and Vichon. However, such RGB-D cameras have a limited field of vision, and ordinary RGB cameras are more suitable for construction sites. Du S et al. [8] and Shrestha K et al. [9] used the edge detection algorithm to detect the edges of objects in the upper region of the head, but such studies relied on facial features. The camera cannot take a positive image when the worker is facing down. In this case, the program cannot detect the face and the helmet. In addition, when the helmet detection program is executed, the edge detection program cannot give a clear outline of the helmet when the contrast between the helmet and the background color is not high enough. Sometimes the program fails to detect a worker when they are moving quickly. Park M W et al. [10] used background subtracting and HOG features to simultaneously detect the human body and helmet in each video frame and realize the identification of workers without a helmet. However, this method relies on the spatial and geometric relationship between the recognition window of the human and the helmet to carry out the human-helmet matching. When it is applied to the recognition of the upright or walking posture if it is squatting or other posture, the matching parameters need to be adjusted. Secondly, workers cannot be identified if they are shaded; In addition, if the worker is stationary, it will be filtered out as background.

Besides, in recent years, deep learning methods have attracted a great deal of attention in computer vision due to their ability to learn valuable functions from large-scale annotated training data. Compared with traditional computer vision, it can process more complex data, and is more flexible, and often has higher accuracy. Some related technologies have been gradually applied in the field of construction. In terms of PPE detection, Fang Q et al. [11] used Faster R-CNN to test the performance of workers without safety helmets under various possible visual conditions (visual range, weather, posture, occlusion, and lighting) at the construction site, and the accuracy and recall rates of the test results were both over 90%, indicating good performance and robustness. In addition, Wu J et al. [12] used the framework of Single Shot Multibox Detector (SSD) and the aggregation features of Reverse Progressive Attention (RPA) to detect the situation without helmet wearing, and it has better performance and lower computing costs. Nath N D et al. [13] introduced and used YOLOv3 to test the detection performance of PPE, and mainly designed three kinds of tests: the first was to detect helmet, vest, and workers respectively; next, ML classifier (such as NN and DT) was used to check whether the detected hat or vest was indeed worn by the detected workers; The second directly classifies each worker into those who wear the corresponding PPE and those who do not wear it. The third method first detects all workers in the input image and then applies a CNN-based classifier model (such as VGG-16, ResNet-50, Xception, or Bayesian) to the cut images of workers to detect workers as either wearing or not wearing the appropriate PPE. Fang W et al. [14] used Faster R-CNN to identify workers and used a CNN model to identify the harness attached to workers so as to detect the workers working at a height without wearing a harness. However, this study only identified whether workers working at a certain height were attached to harness. It cannot be applied to a broader

scene and to judge whether the scene requires workers to wear a harness or whether the harness is worn correctly.

3 Method

The flowchart of this research is shown in Fig. 1 below. The aim of this research is to test the performance of the detection of the workers who are not wearing a harness in a mobile lifting platform (mlp) using a kind of object detection algorithm named YOLOv5 [15]. For this purpose, we collected RGB images from the Internet and construction sit about mlp, workers, and harness and established a dataset based on these images. Then YOLOv5 was used to train this dataset, and the results are evaluated. Based on this, a condition judgment link is added to determine whether workers need to wear a harness (if the worker is in the mlp) and whether they are wearing harnesses according to the position relation output on the picture.

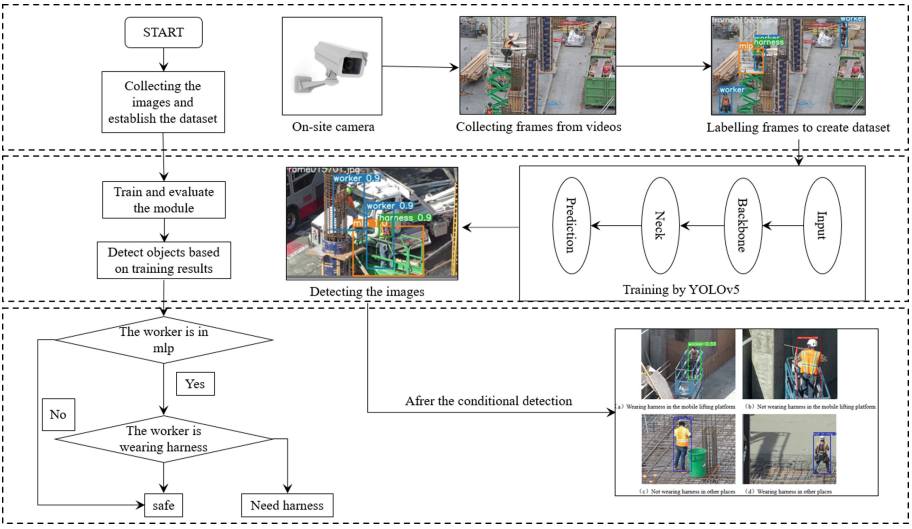


Fig. 1. The flowchart of this research

As mentioned above, in building construction, falling from height is a large part of the various safety accidents that are prone to occur.

Unlike other safety equipment, such as helmets and vests, which need to be worn anytime and anywhere in construction scenarios, harnesses are only used in specific scene where a high fall is likely to occur. Considering that there are a variety of specific aerial work scenes, such as climbing ladders, roof work, aerial edge work, etc., this paper only selects a scene for the test, which is the scene in which workers work on the mobile lifting platform (mlp).

3.1 The Establishment of the Dataset

The dataset contains the training set of 929 images and the verification set of 254 images.

3.1.1 The Collection of the Data

From the types of images, there are mainly three categories: propaganda pictures on the Internet; pictures of the actual construction site; the researchers took pictures of a simulated construction scene. Data sources in this paper are mainly from the first two kinds, and one is from the data on the Internet, the other is from the shooting on the construction site.

In terms of data acquisition methods, data can be obtained in three main ways: open-source data sets on the Internet, pictures by crawlers and other technical means on the Internet, and pictures captured by shooting videos on the construction site. This article mainly uses the first two methods to obtain the required images.

3.1.2 The Labeling of the Data

The labeling tool used in this paper is Labelme [16]. It is installed under Anaconda, Python version 3.7, detailed installation procedures can be found in the GitHub project description. It can be used for annotation of data sets such as instance segmentation, semantic segmentation and target detection. The annotated text is saved in JSON format. Since the file format required by YOLOv5 for training is txt format, data format conversion of JSON formatted annotated files is needed.

An object of the txt format contains five numbers. The first number is the category of the label, usually represented by an integer starting from 0. The second number is the x-direction coordinate of the center point of the annotation box, the third is the y-direction coordinate of the center point of the annotation box, the fourth is the width of the annotation box, and the fifth is the height of the annotation box. All are normalized coordinate data.

3.2 The Process of the Experiment

The dataset was trained under three network structures proposed in YOLOv5: s, m, and l (which means the structure is small, middle, or large). To facilitate understanding of the subsequent test process, these important parameters of this experiment are explained as follows:

The parameter weights have four choices, including s, m, l, and x, in the form of .pt file.

The parameter epochs represent the training rounds. According to the test requirements and attempts in this paper, it is found that the performance of the model tends to be stable in all aspects after 20 rounds, and 50 rounds are used in all subsequent tests in this paper.

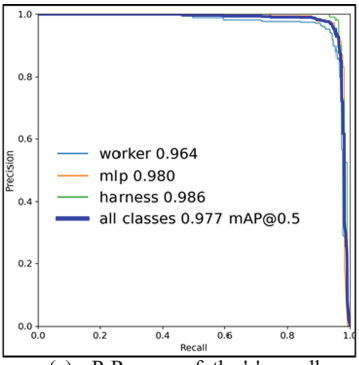
BatchSize refers to the number of samples selected in one training, which affects the use of GPU memory. The value of BatchSize is 16 for all subsequent experiments in this paper.

The img_size also naturally affects the training results because images with higher resolution have richer features. In YOLOv5, this value must be a multiple of 32, and its value is 960 in all subsequent experiments in this paper.

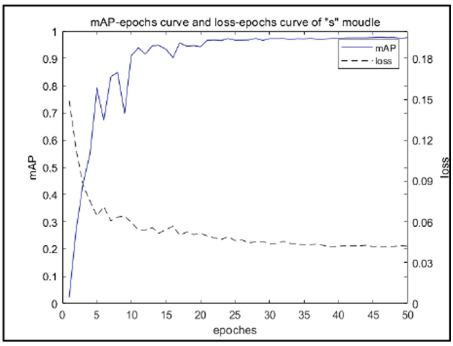
4 Results and Discussion

4.1 The Results of the Experiments

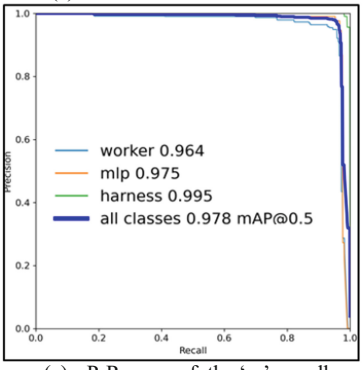
The result of the experiments is shown as following Fig. 2:



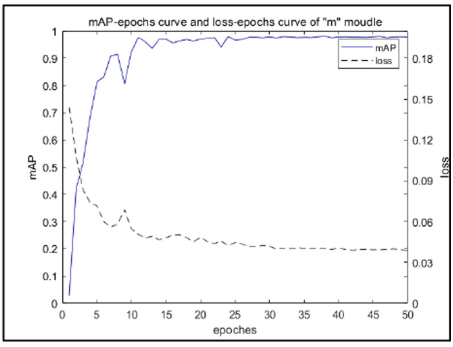
(a) P-R curve of the 's' moudle



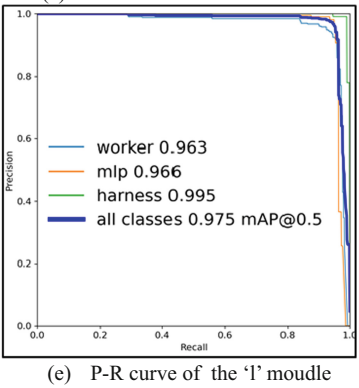
(b) mAP, loss-epochs curve of the 's' moudle



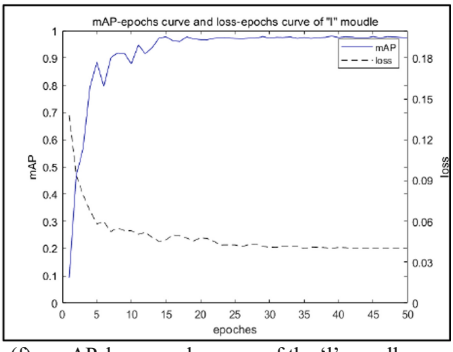
(c) P-R curve of the 'm' moudle



(d) mAP, loss-epochs curve of the 'm' moudle



(e) P-R curve of the 'l' moudle

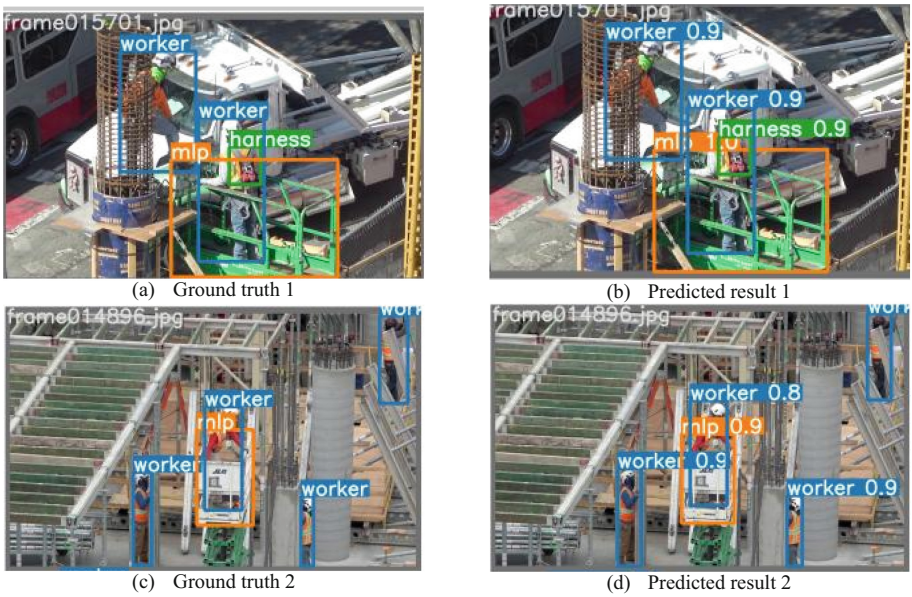


(f) mAP, loss-epochs curve of the 'l' moudle

Fig. 2. The training results of s, m, and l network structures. (The lower area of the P-R curve is the average precision (AP) of this category)

Precision refers to the proportion of the actual positive samples in all the predicted positive samples, while recall refers to the proportion of the predicted positive samples in all the actual positive samples. After giving a limit value for IoU (Intersection over Union), only the sample whose IoU is more significant than this value is called a positive sample, and the sample whose IoU is less than this value is called a negative sample. The test results were arranged according to their score, and the number of selected prediction boxes was increased successively from high to low. Each prediction box was selected, and a set of precision and recall rates could be calculated according to the IoU threshold value. By putting the precision and recall rates of each group into an image, a precision-recall curve could be obtained. The mean value of the precision rate (that is, the area under the curve) for different recall rates is called AP, and the mean value of the AP for each detection category is called mAP. And the mAP with an IoU threshold of 0.5 is denoted as mAP@0.5.

The comparison between the original labeling results of the verification set and the detection results generated by the model is shown in Fig. 3 below:



(a) Ground truth 1 (b) Predicted result 1

(c) Ground truth 2 (d) Predicted result 2

Fig. 3. Comparison between the original labeling and the predicted results of the verification set in the experiment (the figure (a) and (c) is the ground truth, and the figure (b) and (d) is the predicted results of the model, and the predicted results will also give a confidence in the upper right corner)

4.2 The Conditional Detection

Different from safety equipment such as safety helmets that need to be worn at all times, harnesses need to be worn in certain scenes. Therefore, a scene judgment part is set up

in this paper to judge whether workers need to wear harnesses according to the position relationship between workers and the lifting platform. If the center of the worker's box falls within the range of the detection box of the lifting platform, it will be judged that the worker needs to wear a harness. The second step is to judge whether the worker is wearing the harness according to the position relationship between the harness and the worker. If the center of the harness's box falls within the range of the worker's detection box, it will be judged that the worker is wearing the harness.

There are four output modes in total. If workers need to wear a harness and are wearing it, the output displays 'worker'. If the worker is on the lifting platform but not wearing a harness, the output shows 'need harness'; if the worker is not on the lifting platform, the output shows 'not in mlp'. In addition, if the worker's share in the picture is less than 0.005%, then the output displays 'too far' to indicate that this distance may be vague, and no judgment will be made on its state. The examples of the test results are shown in Fig. 4 below:

（a）Wearing harness in the mobile lifting platform

（b）Not wearing harness in the mobile lifting platform

（c）Not wearing harness in other places

（d）Wearing harness in other places

Fig. 4. Examples of the harness wearing condition detection output

4.3 Discussion

Considering that falls from height account for a large share of construction safety accidents, this paper discusses the application of computer vision technology in the automatic detection of harness-wearing of construction workers and sets up a specific scene for experimental research and analysis of its effect. According to the results of the experiment, the following conclusions can be drawn: 1) The mAP of the model training results are all greater than 0.97, and the detection speed is between 9 ms/fps and 15 ms/fps,

which indicates that the application of YOLOv5 in the automatic detection of a harness is efficient; 2) In this paper, the training is carried out under s, m and l network structures. Compared with the network structure of s, the training results of m and l are not significantly improved. This indicates that large models are not needed. And according to the mAP-epochs curve, training epochs do not need to be too high; 3) the precision-recall curve is not very smooth, which may need to increase the dataset.

5 Conclusion

In order to improve the safety management level of the construction site, to reduce the unsafe behavior of workers, and reduce the risk of accidents, this paper attempts to study the feasibility of automatic detection of the harness-wearing of construction workers. In this paper, a deep learning-based computer vision method was used to conduct experimental research. With YOLOv5 as the tool, a scene was set for workers to wear a harness when working on a mobile lifting platform, and a dataset about the harness, lifting platform, and workers was created for training. The results show that it has good performance. The research in this paper presents a method to detect harness-wearing automatically in a specific scene of construction, and shows that the application of computer vision technology in specific construction activities has strong feasibility, which is worth future research. But the work of this paper can only detect whether workers need to wear a harness in the lifting platform and can only detect whether workers are wearing the belt of the harness but not whether their hooks are correctly hanging. Future research can enrich more application scenes or find a new way to detect whether workers are working in a dangerous height or scene, such as climbing or tilting, by summarizing various movements and postures of workers.

Acknowledgments. This work was in part supported by National Natural Science Foundation of China (52078374), Fundamental Research Funds for the Central Universities (22120210288).

References

1. Sunindijo, R.Y., Zou, P.X.W.: Political skill for developing construction safety climate. J. Constr. Eng. Manag. **138**(5), 605–612 (2012)
2. Ministry of Housing and Urban-Rural Construction of the People's Republic of China: Ministry of Housing and Urban-Rural Development General Office on 2019 Notification of production safety accidents in housing and municipal engineering, 19 June 2020. http://www.mohurd.gov.cn/wjfb/202006/t20200624_246031.html
3. Heinrich, H.W.: Industrial Accident Prevention. A Scientific Approach, 2nd edn. McGraw-Hill, New York (1941)
4. Kelm, A., et al.: Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites. Autom. Constr. **36**, 38–52 (2013)
5. Barro-Torres, S., Fernández-Caramés, T.M., Pérez-Iglesias, H.J., Escudero, C.J.: Real-time personal protective equipment monitoring system. Comput. Commun. **36**(1), 42–50 (2012)
6. Dong, S., He, Q., Li, H., Yin, Q.: Automated PPE misuse identification and assessment for safety performance enhancement. In: ICCREM 2015, pp. 204–214 (2015)

7. Du, S., Shehata, M., Badawy, W.: Hard hat detection in video sequences based on face features, motion and color information. In 2011 3rd International Conference on Computer Research and Development, vol. 4, pp. 25–29 (2011)
8. Shrestha, K., Shrestha, P.P., Bajracharya, D., Yfantis, E.A.: Hard-hat detection for construction safety visualization. J. Constr. Eng. **2015**, 1–8 (2015)
9. Park, M.-W., Elsafty, N., Zhu, Z.: Hardhat-wearing detection for enhancing on-site safety of construction workers. J. Constr. Eng. Manag. **141**(9), 04015024 (2015)
10. Fang, Q., et al.: Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Autom. Constr. **85**, 1–9 (2018)
11. Wu, J., Cai, N., Chen, W., Wang, H., Wang, G.: Automatic detection of hardhats worn by construction personnel: a deep learning approach and benchmark dataset. Autom. Constr. **106**, 102894 (2019)
12. Nath, N.D., Behzadan, A.H., Paal, S.G.: Deep learning for site safety: real-time detection of personal protective equipment. Autom. Constr. **112**, 103085 (2020)
13. Fang, W., Ding, L., Luo, H., Love, P.E.D.: Falls from heights: a computer vision-based approach for safety harness detection. Autom. Constr. **91**, 53–61 (2018)
14. Jocher, G.: YOLOv5, 5 January 2021. https://github.com/ultralytics/yolov5
15. Wada, K.: Labelme, 25 January 2021. https://github.com/wkentaro/labelme