

Article

Workers' Unsafe Actions When Working at Heights: Detecting from Images

Qijun Hu ^{1,*}, Yu Bai ¹, Leping He ², Jie Huang ³, Haoyu Wang ⁴ and Guangran Cheng ³

¹ School of Mechatronic Engineering, Southwest Petroleum University, Chengdu 610500, China; by_duo@163.com

² School of Civil Engineering and Geomatics, Southwest Petroleum University, Chengdu 610500, China; 201231010028@swpu.edu.cn

³ CECEP Construction Engineering Design Institute Limited, Chengdu 610052, China; hi_huangjie@163.com (J.H.); gr-cheng@live.cn (G.C.)

⁴ College of Physics, Chongqing University, Chongqing 400044, China; wanghaoyu200007@163.com

* Correspondence: huqijunswpu@163.com; Tel.: +86-130-7283-2168

Abstract: Working at heights causes heavy casualties among workers during construction activities. Workers' unsafe action detection could play a vital role in strengthening the supervision of workers to avoid them falling from heights. Existing methods for managing workers' unsafe actions commonly rely on managers' observation, which consumes a lot of human resources and impossibly covers a whole construction site. In this research, we propose an automatic identification method for detecting workers' unsafe actions, considering a heights working environment, based on an improved Faster Regions with CNN features (Faster R-CNN) algorithm. We designed and carried out a series of experiments involving five types of unsafe actions to examine their efficiency and accuracy. The results illustrate and verify the method's feasibility for improving safety inspection and supervision, as well as its limitations.

Keywords: unsafe actions; working at heights; intelligent recognition; deep learning; construction site



Citation: Hu, Q.; Bai, Y.; He, L.; Huang, J.; Wang, H.; Cheng, G. Workers' Unsafe Actions When Working at Heights: Detecting from Images. *Sustainability* **2022**, *14*, 6126. <https://doi.org/10.3390/su14106126>

Academic Editors: Srinath Perera, Albert P. C. Chan, Xiaohua Jin, Dilanthi Amarantunga, Makarand Hastak, Patrizia Lombardi, Sepani Senaratne and Anil Sawhney

Received: 25 March 2022

Accepted: 11 May 2022

Published: 18 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Working at heights is associated with frequent injuries and the deaths of workers during construction activities [1]. Fatal injuries due to construction accidents exceed 60,000 injuries every year all over the world [2]. According to the Occupational Safety and Health Administration, a similar trend occurs in developed countries, even though infrastructure construction has almost been completed in these countries [3]. Statistics from the Ministry of Housing and Urban-Rural Development of China show that, from 2010 to 2019, there were an average of 603 production safety accidents per year, resulting in approximately 730 worker deaths per year [4]. Among these accidents, fall-from-height accidents accounted for at least 52.10%, followed by struck-by-object accidents (13.90%, Figure 1). Many researchers have noted that the root causes of safety accidents are workers' unsafe behaviors [5–7]. Heinrich's accident causation theory states that more than 80% of safety accidents are caused by workers' unsafe behaviors [8]. Therefore, management to minimize workers' unsafe behaviors is important to construction safety.

Behavior-based safety (BBS) plays an influential role in the supervision and management of workers' activities [5,9,10], in which workers' activities are recorded and their behaviors are analyzed through observation, interview, and survey. Most BBS studies have involved four necessary steps [11]: (1) create a list of workers' unsafe behaviors; (2) observe and record the frequency of unsafe behaviors; (3) educate and intervene in workers' behavior; (4) provide feedback and perform follow-up observations. BBS has gained its status in construction management because it has been more successful than other methods for solving problems caused by unsafe behaviors. Furthermore, researchers

have recognized unsafe behaviors as the most important problem. The purpose of normal science is neither to discover new types of phenomena nor to invent new theories [12]. On the contrary, normal science research is to continuously improve phenomena and theories provided by existing paradigms, which are eternal challenges to researchers' skills and imagination [13]. BBS observation is a traditional form of worker behavior measurement, which has certain limitations in practical applications [14].

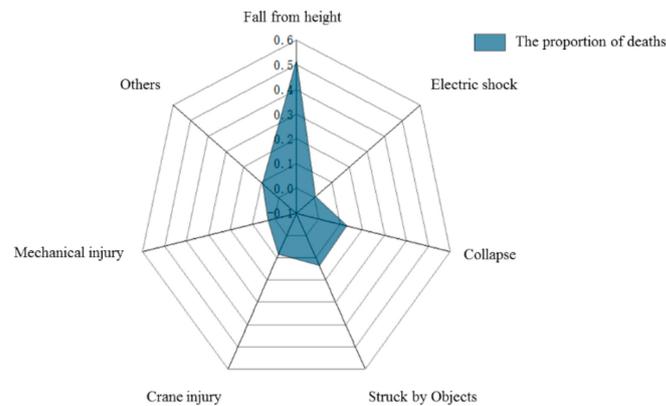


Figure 1. Distribution map of the causes of death during construction from 2010 to 2019 [4].

Observation—the second stage—is important because it can provide more data for the analysis of patterns. Ref. [15] puts forward a safety assessment method of leading indicators based on jobsite safety inspection (JSI) through a lot of accident data analyzing. Traditional unsafe behavior observation mainly relies on safety managers' manual observation and recording, which not only consumes a lot of time and cost, but it is also difficult to cover the whole construction site, or all workers. On the one hand, many human resources are needed for data acquisition due to large sample data requirements [16]. On the other hand, excessive reliance on workers' observations can easily cause personal impact since different people have different feelings about the same thing [17]. Therefore, an automated and reliable method that could efficiently measure unsafe behavior is needed to support BBS observation. Automation technology is already making its mark in the observations of workers' behaviors [18]. Proposed real-time positioning systems based on different types of sensors and the Internet of Things (IoT) have played considerable roles in workers' safety observations [19,20]. However, sensors can sometimes affect workers' normal work [21]. Computer vision technology can also be used for collecting and processing workers' safety information [22]. Its ability to provide a wide range of visual information at a low cost has attracted a lot of attention [23–26].

Construction workers' unsafe actions are a type of unsafe behavior that could be the main reason leading to construction accidents, primarily occurring when working at heights. Most unsafe actions are instantaneous, and therefore, it is difficult for safety supervisors to observe them in real time. Furthermore, detecting workers' unsafe actions is critical to the observation process of BBS. Computer vision technology for the automatic recognition and detection of workers' unsafe actions could tentatively replace manual observations of BBS. The gap in research could be significant for specific groups of construction workers. To date, there is no automatic method of detecting the unsafe behavior of workers in a high working environment. Therefore, the present study is to help improve the observation method of workers' unsafe behavior considering five unsafe actions that mostly appeared in the high working environment on construction sites. A series of experiments involving over 30 testees were implemented to verify the proposed method. This paper is structured as follows: First of all, the research method is presented, involving unsafe actions lists, dataset construction, and the Convolutional Neural Networks (CNN) model built. Following this, the results are presented and discussed. Finally, a conclusion is drawn.

2. Related Works

2.1. Safety Management in High Places

Falls from height in construction sites have earned science mapping research to reveal the existing research gaps [27]. The common causes that lead to falling accidents are defects in protective devices, poor work organization [28], and workers' unsafe actions, such as sleeping on the baseboard. Ref. [29] built a database that dissected the mechanics of workers falling off the baseboard. An IoT infrastructure, combined with the fuzzy markup language for falling objects on the construction site, could greatly help safety managers [30].

Deep learning enhances the automation capabilities of computer vision in safety monitoring [14,23,31]. CNN has shown exceptionally superior performance in high-dimensional data with intricate structures processing. Related research on using computer vision for worker safety management under working-at-height conditions can be described from the following three aspects:

Aspect 1: To automatically check a worker's safety equipment, such as their helmet and seat belt [31–33]. The detection of safety equipment originated from early feature engineering research, such as the histogram of oriented gradients (HOG) [34]. It has been proposed that increasing the setting of the color threshold of the helmet and the upper body detection could improve detection accuracy [35]. Deep learning has been used to develop multiple processing layers to extract unknown information, without the need to set the image features artificially [36]. In addition, a regional convolutional neural network (R-CNN) has been used to identify a helmet and has achieved good results [37].

Aspect 2: To automatically identify hazardous areas. This research is generally based on object recognition, including openings, rims, and groove edges in high places [38–40]. Computer vision has been used to detect whether workers pass through a support structure [39], since workers walking on support structures have a risk of falling.

Aspect 3: To monitor non-compliance of safety regulations [41], in particular, climbing scaffolding and carrying workers with a tower crane [42]. There has been a study on intelligent assessment for interactive work of workers and machines [24]. Research in this area needs to combine computer vision and safety assessment methods so that safety status can be explained using computer semantics [38,41].

2.2. Computer Vision in Construction

Scientists designed the CNN to describe the primary visual cortex (V1) of the brain by studying how neurons in a cat's brain responded to images projected precisely in front of the cat [43,44]. There are three fundamental properties to a CNN imitating the V1 [45]:

Property 1: The V1 can perform spatial mapping, and CNN describes this property through two-dimensional mapping.

Property 2: The V1 includes many simple cells, and the convolution kernel is used to simulate these simple cells' activities; that is, the linear function of the image in a particular receptive field.

Property 3: The V1 includes many complex cells, which inspire the pooling unit of CNN.

On all accounts, there are still two factors that determine the deep learning target detection model quality:

Factor 1: The dataset that is used to train the workers' unsafe actions recognition model should have unique unsafe action features that can be identified by a computer. There are many open-source datasets available for deep learning research, such as ImageNet [46] and COCO [47]; however, not all of them are suitable for object detection on construction sites. Many researchers have also established datasets related to construction engineering. For object recognition and detection on construction sites, there are datasets about workers [31]; construction machinery [31]; and on-site structures, such as railings [48]. For workers' activities on the construction sites, a dataset dedicated to steelworkers engaged in steel processing activities has been established [49]. Scholars have even enhanced datasets by preprocessing Red-Green-Blue (RGB) images to optical and gray images [49], which has provided novel ideas for dataset acquisition.

Factor 2: Deep learning algorithms and models as mathematical methods to find optimal solutions are important, as they affect the detection results. Le-Net is the foundation of deep learning models [50], which contains the basic modules of deep learning, convolutional layer, pooling layer, and fully connected layer. AlexNet came first in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC2012) [36]. Since then, various neural network models have shown their accuracy and efficiency in feature extraction, including ZF-net [51], VGG-net [52], Res-net [53], Inception-net [54], etc. After obtaining the feature map, additional algorithms are needed to classify and locate the object. Faster R-CNN, YOLO, and SSD are deep learning algorithms that are widely used in many applications.

In this research, we propose an improved faster R-CNN utilizing a special dataset composed of unsafe actions when working at heights, which is based on ZF-net. This work's application is that it could provide construction managers with information on workers' unsafe actions, and therefore, assist with interventions. In addition, it could become a new procedure for BBS observation.

3. Methods

3.1. Research Framework

Feature engineering needs to set the transcendental extracted features artificially. Deep learning training is a process of finding the optimal solution based on a dataset using representation. Unlike feature engineering, representation learning is a group of machine learning methods that can automatically find useful input data features through a general-purpose learning procedure [55–57]. The representation learning is similar to a “black box”, and therefore, it is difficult to understand how the internal nonlinear function works. As the dataset and the algorithm are the critical factors in applied research related to deep learning, the research process is designed by improving design science research [58]. In this study, this is more suitable for research on the automatic detection of workers' unsafe actions, as shown in Figure 2.

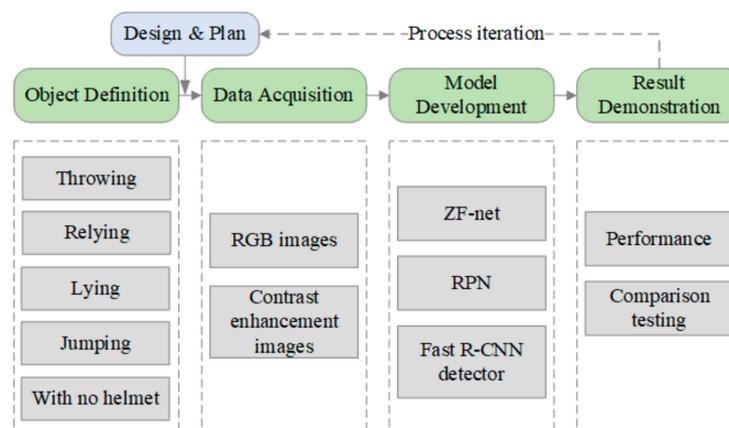


Figure 2. The research process of intelligent recognition for workers' unsafe action.

The method is structured as follows:

- (1) Before building the automatic recognition model, the workers' unsafe actions that are to be detected should be defined. This research lists five kinds of worker action types that were likely to cause safety accidents when working at heights.
- (2) The second stage is data acquisition, mainly to acquire images with features of workers' unsafe actions that could be used for deep learning training, validation, and testing. In this step, Red-Green-Blue (RGB) and contrast enhancement images are integrated to reinforce the dataset performance.
- (3) Then, the model development stage involves deep learning training and testing.

Therefore, the test results demonstrate the model's performance and are the basis of the design and plan modifications.

3.2. Definition of Workers' Unsafe Actions

Fall-from-height and struck-by-object accidents are prone to happen when workers are working at heights; therefore, in this research, we analyzed workers' unsafe actions through investigating the regulations and reports on these two accident types [1,59,60]. Simultaneously, it should be considered whether the unsafe actions have features that could be identified through computer vision. The five main unsafe actions are summarized in Table 1, namely, throwing objects downwards, relying on railings, lying on scaffold boards and operating platforms, jumping up and down levels, and not wearing a helmet.

Table 1. Unsafe actions categories and descriptions.

No.	Categories	Unsafe Actions Descriptions
1	Throwing	1.1 Throw waste and leftover materials at will. 1.2 Throw fragments down when working on building exterior walls. 1.3 Throw tools and materials up and down. 1.4 Throw rubbish from windows. 1.5 Throw dismantled objects and remaining materials arbitrarily. 1.6 Throw broken glass downwards when installing skylights.
2	Relying	2.1 Relying on the protective railing. 2.2 Rely or ride on the window rails when painting windows.
3	Lying	3.1 Lie on scaffold boards and operating platforms.
4	Jumping	4.1 Jump up and down shelves.
5	With no helmet	5.1 Workers fail to use safety protection equipment correctly when entering a dangerous site of falling objects.

All of the unsafe actions are momentary actions, except for not wearing a helmet, and therefore, it is difficult for security managers to observe these actions instantly. Whether or not to wear a helmet is a relatively stable state of a person's action, which security managers can clearly observe and prevent in time. However, wearing a helmet is very important to the safety of workers, especially for working at heights; therefore, it is worthwhile including it in the research content.

3.3. Data Acquisition

The quality and quantity of the dataset is the decisive factor affecting detection accuracy. Many researchers have established datasets for construction safety management, such as datasets used to identify construction machines and workers and to identify the activities of steel processing workers. To date, there is no dataset for the characteristics of workers' unsafe actions when working at heights on construction sites; therefore, even if multiple publicly annotated open-source datasets could be used for deep learning detection model training and testing, none of them could be used as experimental data for this study directly. The dataset used in this research needs to include the five types of workers' unsafe actions that occur when working at heights, including throwing, relying, lying, jumping, and with no helmet. A total of 2200 original sample images were collected. The five unsafe actions sample distribution is shown in Table 2.

Table 2. Sample distribution.

Unsafe Actions	Distribution
Jumping	11.12%
Throwing	21.41%
Relying	22.12%
Lying	15.82%
Helmet	29.53%

3.4. Model Development

The detection algorithm is another decisive factor affecting the detection accuracy. Deep learning algorithms have received wide attention for their potential to improve construction safety and production efficiency. R-CNN and fast R-CNN have been proposed successively and have dramatically improved the accuracy of target recognition. Faster R-CNN, a deep learning algorithm with an “attention” mechanism, introduces the region proposal network (RPN), further shortening the model’s training time and the detection network’s running time. Based on the convolutional neural network framework of convolutional architecture for fast feature embedding (Caffe), a faster R-CNN is mainly composed of two modules. One of the modules is the RPN, which generates a more accurate, high-quality candidate frame position by premarking the targets’ possible positions. The other module is the fast R-CNN detector, which aims to improve the target recognition area based on the RPN’s candidate frame.

The entire training process works by alternate training of the RPN network and the fast R-CNN detector, and both use the Zeiler and Fergus network (ZF-net) [51]. Figure 3 shows the flowchart of the model training for workers’ unsafe actions. The procedure to implement the workers’ unsafe action detection model is described as follows:

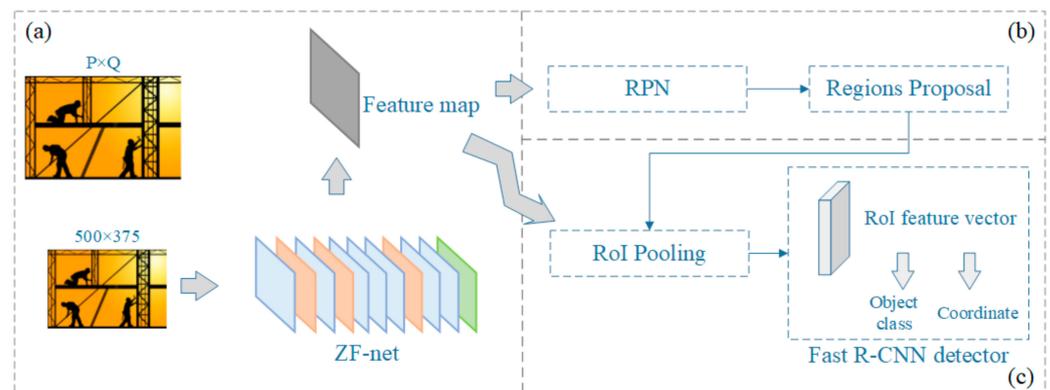


Figure 3. Flowchart of the model training. (a) Input the training samples to the ZF-net; (b) Generate region proposals using the RPN network; (c) Integrate feature maps and proposals and sent to the Faster R-CNN detector to determine the target category.

Stage 1: Input the training samples to the ZF-net for pretraining. Then, conduct alternate training of the model to acquire the first stage RPN and the fast R-CNN detector. Alternate training means to obtain the first stage ZF-net and RPN through the first round of training, and then the first round of the fast R-CNN detector training uses the training samples and the first stage RPN.

Stage 2: The second stage of training is almost the same as the first stage, except the input parameters are the results obtained in the first stage of training. After the second stage of training ends, the obtained ZF-net is saved as the final training network.

Stage 3: The validation sample is entered, and the final ZF-net is used to adjust and update the RPN network and the fast R-CNN. Finally, the proposed detection model is obtained.

4. Experiment and Results

4.1. Experiment

The data were mostly acquired from graduate students, while a small amount of data were acquired from workers on the construction site. College students and workers are adults, and their movements are similar. Since the dataset construction needed to consider image quality factors, such as illumination conditions and different shooting angles, using students as experimental subjects facilitated the collection of a large number of images. Therefore, we selected students’ images as the training set for the model. There were

31 graduate students, with heights ranging from 158 cm to 181 cm, who participated in the experiments. The actions of throwing, relying, lying, jumping, and not wearing a helmet were examples of unsafe actions. The image sample collection followed 6 principles:

- (1) Each student must perform the 5 actions, i.e., throwing, lying, relying, jumping, and without helmets, with their usual manner of behavior.
- (2) Each type of unsafe action would be taken in different scenarios, with different shooting angles and lighting conditions.
- (3) For each class of unsafe actions, 3–5 sequential images as a group were collected to reflect a continuously varying action.
- (4) Images of poor quality were filtered out and deleted, such as indistinct images and targets with a small proportion of images.
- (5) The originally collected images were preprocessed through contrast enhancement, and then included as part of the samples. Adding preprocessed images into the dataset increased the sample size of the dataset and improved the deep learning convolutional neural network model [31,49].
- (6) Samples were reshaped in the dataset to a resolution of 375×500 .

The data collected from construction workers were mainly about wearing a helmet. The total number of participants was approximately 50. Figure 4 shows examples of the 5 unsafe actions image samples. The samples were labeled using the sample labeling tool LabelMag, and the labeled annotation files were saved in the xml format file. The samples were finally divided into 3 groups, i.e., the training data, the validation data, and the testing data. Finally, the dataset was ready, which included 5 action types of images, annotation files, and image sets.

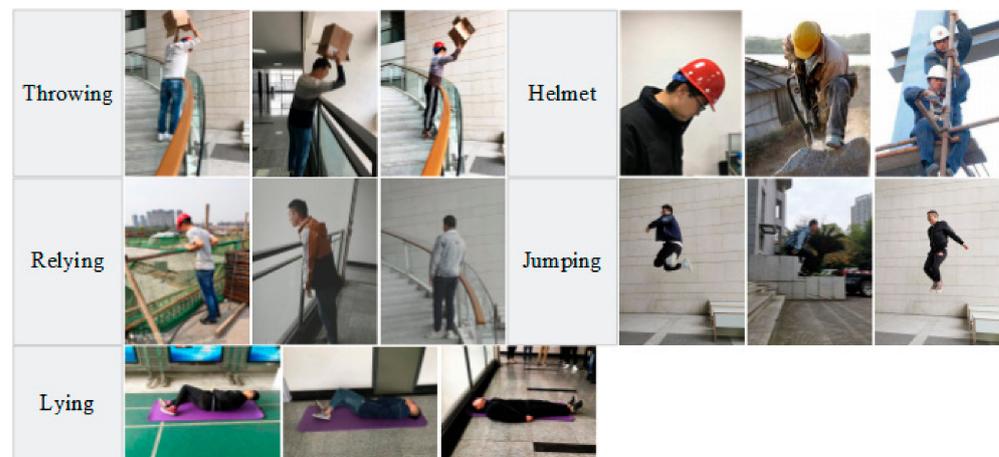


Figure 4. Sample examples.

At the implementation stage, the Faster-RCNN algorithm was trained using 20,000 iterations with a learning rate of 0.001. The dataset was the improved VOC 2007, which has been built in Section 3.3. The proposed method was implemented in MATLAB. For the hardware configuration, the model was tested on a computer with Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz, memory card of 16.0 GB, GPU of NVIDIA GeForce GTX 1080 Ti, and Windows 10 64-bit OS.

4.2. Results

After training, the ZF-net model file for detecting unsafe actions was obtained. The remaining samples were used to test the final model, and the testing results are shown in Figure 5. The average detection time of the sample test is 0.042 s.



Figure 5. Testing results.

Some concepts of target detection need to be explained as follows: True positive (TP), the input images are positive samples, and the detection results are also positive samples; false-positive (FP), the input images are negative samples, but the detection results are positive samples; true negative (TN), the input images are negative samples, and the detection results are negative samples; false-negative (FN), the input images are positive samples, but the detection results are negative. Take the action of relying on a railing as an example. If the testing result is “relying”, it is recorded as TP; if the testing result is without relying, an FN is recorded. However, if no worker is relying on the test image’s railing, but a “relying” is detected, FP is recorded. Figure 6 shows examples of TP, FN, and FP samples.



Figure 6. Examples of (a) TP, (b) FP, and (c) FN samples.

The analysis of TP, TN, FP, and FN’s test results with the fourfold table are shown in Figure 7, which helps to understand the distribution of test results.

Positive (detection)	93.5%	0.3%
Negative (detection)	6.2%	0
	Positive (actually)	Negative (actually)

Figure 7. Test results with the fourfold table.

This experiment used four key performance indicators (KPIs) to assess the performance of the model for detecting unsafe actions: (1) accuracy, (2) precision, (3) recall, and (4) F1 measures.

Accuracy, i.e., the ratio of the TPs and TNs to the total detections, is generally used to evaluate the global accuracy of a training model. Precision is an indicator of the accuracy of prediction and recall is a measurement of the coverage area.

The F1 measure is the weighted harmonic average of precision and recall, which evaluates the model's quality. The higher the F1 measure, the more ideal the model.

The overall performance of the model is good. The test results were 93.46% for accuracy, 99.71% for precision, 93.72% for recall, and 96.63% for the F1 measure.

4.2.1. Sample Source Analysis

For the sample source analysis, we established a dataset of workers' unsafe actions and trained a CNN model for workers' unsafe action detection. The model's performance based on four indicators was good; the results for all of the indicators were above 90%. However, the model's performance indicators were based on the examinees added into the dataset. The images in the dataset had a fairly high degree of similarity. The results could be understood as the model mainly aimed at workers in the same project and construction scenes. However, unsafe actions outside of the dataset should be considered, for example, different scenarios and different examinees. Therefore, we tested the examinees outside the dataset as a comparison group to analyze whether the model applied to other types of scenarios and workers.

Similar to the previous dataset establishment method, the comparison group examinees were asked to perform the five unsafe actions according to their usual manner of behavior. The shooting scene of the comparison group was entirely different from the scene in the original dataset. The results are shown in Table 3 and Figure 8.

Table 3. Test results of the model's overall performance.

Indicators	Accuracy	Precision	Recall	F1-Measure
Original	93.46%	99.71%	93.72%	96.63%
Comparison	83.38%	87.79%	63.79%	73.89%

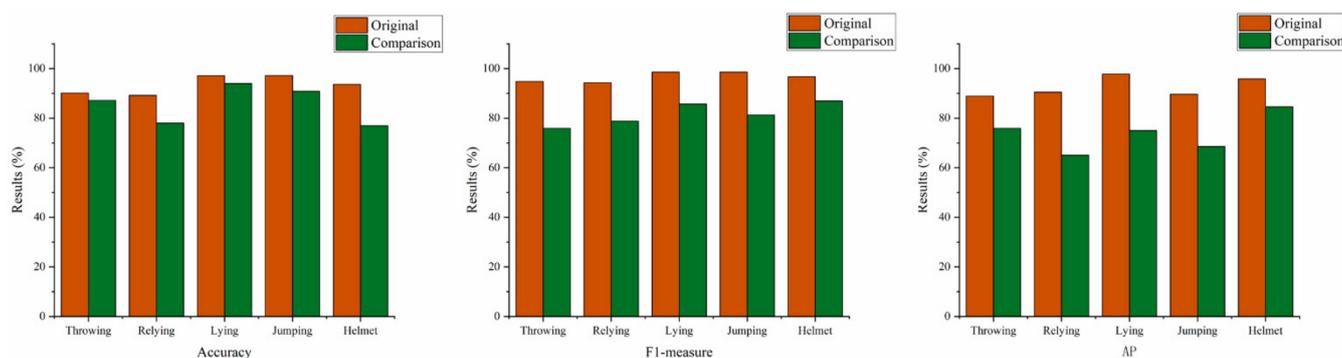


Figure 8. Test results of the model's performance on each unsafe action (between the original and comparison group).

We compared the testing results of the comparison group with the original dataset. The performance of the comparison group was worse than the original samples. Nevertheless, they were still more than 60 percent accurate. The reason could be attributed to the individual differences in the actions performed. There are significant numbers of construction sites, as well as numbers of workers. An unsafe action detection dataset built for a specific engineering construction site is more suitable for workers at a specific construction site for a long time period. To detect workers' unsafe actions at other engineering projects, rebuilding the unsafe action model dedicated to those projects would be more conducive to observing workers. The interference environment of images also affects the recognition accuracy, such as light, rain and fog.

4.2.2. FPs Analyzation

In the FP analysis, it was observed that FPs were usually due to the similar characteristics of two unsafe actions. For example, Figure 9 shows an image sequence of a throwing action. The testing result of the throwing action in Figure 9 was erroneously detected as relying on the railing. At the early stage of the throwing action, the detection result was relying (Figure 9a). As the image sequence changed, the detection was a result of both throwing and relying on the railing (Figure 9b). At the end of the throwing action, a throwing result was detected when the projectile was about to leave the hand (Figure 9c).

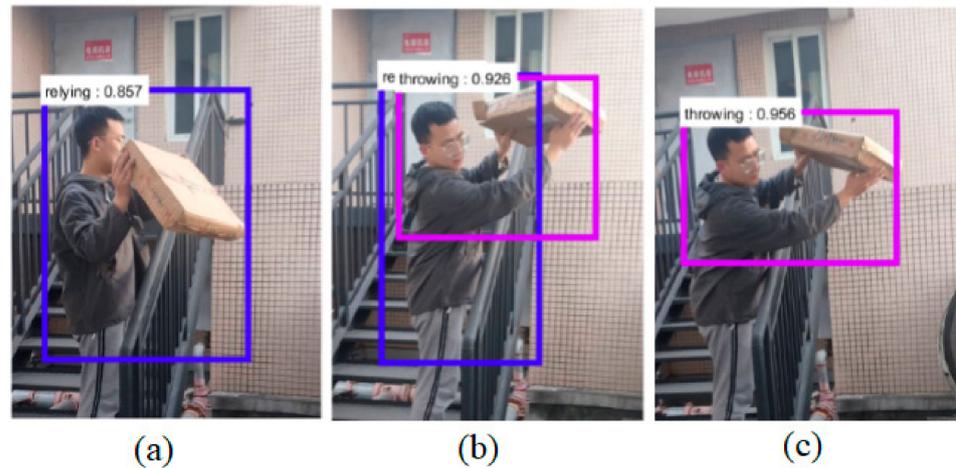


Figure 9. Testing results analysis for a FP sample. (a) At the early stage of the throwing action, the detection result was relying; (b) As the image sequence changed, the detection was a result of both throwing and relying on the railing; (c) At the end of the throwing action, a throwing result was detected when the projectile was about to leave the hand.

There are two explanations for the phenomenon that appeared above.

First, most of the image samples of the relying action in this study were relying on the railing. When performing sample labeling, the railing object was usually included in the sample labeling box of the relying action. The CNN learns all types of features of the images. These features include but are not limited to color, shape, and texture. Therefore, when the ZF-net was learning action features, it might mistake the railing for the feature of the relying action.

Secondly, the recognition method of the human skeleton may help to explain this problem. In previous studies, the posture of the human body could be distinguished by the distribution of human bone parameters [61], such as elbow angle and torso angle. The similarity of the human bone shape parameters in these two actions (throwing and relying) was very high, as shown in Figure 9. When a worker throws an object beside a railing, it is accompanied by a relying posture. Hence, it was not surprising that such a result was reached. In recent related research, Openpose has been used as an advanced algorithm that could achieve accurate human posture. The general process of CNN in learning image features has usually been divided into three layers [57], which could simply reveal the process of accepting and understanding the news for the V1. In the first layer, the learned features usually indicate whether there are edges in a specific direction and position in the image. In the second layer, the pattern is detected by discovering the specific arrangement of the edges, regardless of small changes in edge positions. In the third layer, the patterns are assembled into larger combinations of parts corresponding to familiar objects, and subsequent layers detect the object as a combination of these parts. This is a bottom-up strategy. In contrast to this learning strategy, part of the affinity field (PAF) is used in Openpose, which can improve the ability of computer vision in human pose estimation [62]. This provides a new research method in workers' unsafe actions management.

4.2.3. Helmet Detection Analysis

For the helmet detection analysis, the action detection of whether the worker wears a helmet is different from other unsafe action detection methods. Detecting only the helmet could not determine whether the worker is wearing a helmet. Figure 10 is an intuitive explanation of this conclusion. When a helmet is not worn, it could also be detected that the object is a helmet (Figure 10a). However, based on “people” detection, helmet detection could effectively avoid this problem. Therefore, on the one hand, for the helmet-wearing test, if there is a result that the person and the helmet are detected simultaneously, as shown in Figure 10b, it was defined as safe. On the other hand, if only one of them appears, it could not judge whether it was an unsafe action, Figure 10a.

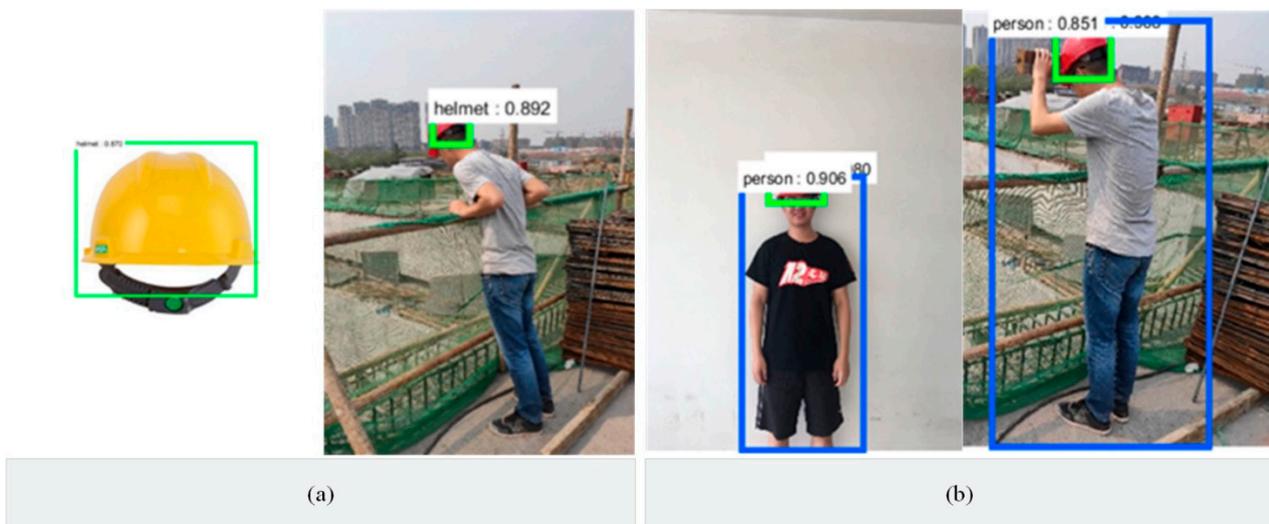


Figure 10. Helmet wearing test (a) Helmet; (b) Helmet & person.

5. Discussion

This research proposed an automatic identification method based on an improved Faster R-CNN algorithm to detect workers' unsafe actions considering a heights working environment. Based on the method, this research designs and carries out a series of experiments involving five types of unsafe actions to examine their efficiency and accuracy. The results illustrate and verify the method's feasibility for improving safety inspection and supervision, as well as its limitations. According to the experiment and results, it could be found that it is an excellent way to detect workers' unsafe behaviors through the proposed method.

Compared with previous studies that have utilized computer vision for construction management, this research has the following advantages. This work combines human knowledge with computer semantics that consider a high workplace. Thus, leading to better intervention in construction safety management when workers work at heights. For the observation and recording part of BBS, manually observing was a waste of human resources and could not capture the worker's action information comprehensively. Computer vision has an advantage in this aspect. Computer vision has been utilized to observe workers, including the efficiency of their activities [49], helmet-wearing [37], and even construction activities at night-time [63]. These works are significant to project management and achieved a good effect. However, accidents are most likely to occur when workers work at heights. There is a lack of research on observing workers' behavior in high working environments. The proposed method is for workers' observation in a high working place. There was a dataset of unsafe actions that commonly happen in high places on construction sites. It may provide a more reliable method for observing workers' behavior in a high places.

Three factors affect the robustness of the model:

- (1) Whether a deep learning algorithm could fully extract image features.
- (2) Whether the dataset could adequately represent the detection object. Deep learning is learning the features in a dataset. Therefore, the characteristics contained in the dataset determine the final effect of the model.
- (3) The quality of the image used for recognition also affects the robustness. It includes the quality of data acquisition equipment, image acquisition angle and object obscured, and the influence of environmental factors, such as light, rain and fog on image quality.

This research also has application limitations. It lacks a large-scale dataset. There comprised approximately 2200 images utilized for the CNN model training, which was deemed to be small. The quality and quantity of images in the dataset will affect the effect of the model. Future studies will consider further dataset improvement to enhance the robustness of the model.

6. Conclusions

Most construction sites are equipped with cameras to observe their safety status. However, manual observation is laborious and may not accurately capture workers' unsafe behavior. Computer vision technology's time-sparing and intelligent advantages could help construction safety management when workers work at heights. This paper proposed a deep learning model that could be applied to detect unsafe actions when working in a high place automatically. The workers' unsafe actions worth observing and detecting were defined first to achieve this work. The dataset including five workers' unsafe actions for deep learning was built. Finally, an automatic recognition model was built for training, validation, and testing the unsafe actions model. The model's accuracy in detecting throwing, relying, lying, jumping actions, and helmet wearing were 90.14%, 89.19%, 97.18%, 97.22%, and 93.67%, respectively.

This work combines human knowledge with computer semantics, thus leading to a better intervention in construction safety management when workers work at heights. Its contribution is to enable computers to identify the unsafe behavior of workers in a high working environment. Its application is that it could provide workers' unsafe action information intelligently to safety managers and assist with intervention. Besides, an unsafe action detection dataset built for a specific engineering construction site is more suitable for workers who would engage with the specific construction site for a long time. It could become a new means for the BBS observation procedure. All of these would benefit workers, managers, and supervisors working in a hazardous construction environment.

Since the research mainly focuses on whether unsafe actions could be well-detected, all scenarios where workers made these actions were not fully considered in the dataset production process. According to the occurrence rules of the accident, an accident is often coupled with multiple factors. Although unsafe actions were an essential factor in accidents, they do not lead to accidents directly but in a specific scenario.

Author Contributions: Y.B. proposed the conceptualization and wrote this manuscript under the supervision of Q.H. L.H. provided the resources. J.H. and H.W. conducted the experiment planning and setup. G.C. provided valuable insight in preparing this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (52178357), the Sichuan Science and Technology Program (2021YFSY0307, 2021JDRC0076), and the Graduate Research and Innovation Fund Program of SWPU (2020cxyb009).

Data Availability Statement: All data generated or analyzed during the study are included in the submitted article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fang, W.; Ma, L.; Love, P.E.; Luo, H.; Ding, L.; Zhou, A. Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Autom. Constr.* **2020**, *119*, 103310. [CrossRef]
2. Lingard, H. Occupational health and safety in the construction industry. *Constr. Manag. Econ.* **2013**, *31*, 505–514. [CrossRef]
3. Available online: <https://www.osha.gov/data/work> (accessed on 16 February 2022).
4. Ministry of Housing and Urban-Rural Development of China (MHURDC), 2010–2019. Available online: <https://www.mohurd.gov.cn/> (accessed on 16 February 2022).
5. Li, H.; Lu, M.; Hsu, S.-C.; Gray, M.; Huang, T. Proactive behavior-based safety management for construction safety improvement. *Saf. Sci.* **2015**, *75*, 107–117. [CrossRef]
6. Li, S.; Wu, X.; Wang, X.; Hu, S. Relationship between Social Capital, Safety Competency, and Safety Behaviors of Construction Workers. *J. Constr. Eng. Manag.* **2020**, *146*, 4020059. [CrossRef]
7. Guo, S.; Ding, L.; Luo, H.; Jiang, X. A Big-Data-based platform of workers' behavior: Observations from the field. *Accid. Anal. Prev.* **2016**, *93*, 299–309. [CrossRef]
8. Heinrich, H. *Industrial Accident Prevention: A Safety Management Approach*; McGraw-Hill Book Company: New York, NY, USA, 1980; Volume 468.
9. Chen, D.; Tian, H. Behavior Based Safety for Accidents Prevention and Positive Study in China Construction Project. In Proceedings of the International Symposium on Safety Science and Engineering in China, Beijing, China, 7–9 November 2012; Volume 43, pp. 528–534. [CrossRef]
10. Zhang, M.; Fang, D. A continuous Behavior-Based Safety strategy for persistent safety improvement in construction industry. *Autom. Constr.* **2013**, *34*, 101–107. [CrossRef]
11. Ismail, F.; Hashim, A.; Zuriea, W.; Ismail, W.; Kamarudin, H.; Baharom, Z. Behaviour Based Approach for Quality and Safety Environment Improvement: Malaysian Experience in the Oil and Gas Industry. *Procedia-Soc. Behav. Sci.* **2012**, *35*, 586–594. [CrossRef]
12. Barber, B. Resistance by Scientists to Scientific Discovery. *Science* **1961**, *134*, 596–602. [CrossRef]
13. Kuhn, T. *The Structure of Scientific Revolutions*; University of Chicago Press: Chicago, IL, USA, 1996.
14. Ding, L.; Fang, W.; Luo, H.; Love, P.; Zhong, B.T.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Autom. Constr.* **2018**, *86*, 118–124. [CrossRef]
15. Bhagwat, K.; Kumar, V.S.; Nanthagopalan, P. Construction Safety Performance Measurement Using Leading Indicator-based Jobsite Safety Inspection Method—A Case Study of Building Construction Project. *Int. J. Occup. Saf. Ergon.* **2021**, 1–27. [CrossRef]
16. Khosrowpour, A.; Niebles, J.; Fard, M. Vision-based workplace assessment using depth images for activity analysis of interior construction operations. *Autom. Constr.* **2014**, *48*, 74–87. [CrossRef]
17. Cheng, T.; Teizer, J.; Migliaccio, G.; Gatti, U. Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic posture data. *Autom. Constr.* **2013**, *29*, 24–39. [CrossRef]
18. Guo, H.; Yu, Y.; Skitmore, M. Visualization technology-based construction safety management: A review. *Autom. Constr.* **2017**, *73*, 135–144. [CrossRef]
19. Li, H.; Li, X.; Luo, X.; Siebert, J. Investigation of the causality patterns of non-helmet use behavior of construction workers. *Autom. Constr.* **2017**, *80*, 95–103. [CrossRef]
20. Lingard, H.; Rowlinson, S. Behavior-based safety management in Hong Kong's construction industry. *J. Saf. Res.* **1997**, *28*, 243–256. [CrossRef]
21. Li, H.; Chan, G.; Wong, J.; Skitmore, M. Real-time locating systems applications in construction. *Autom. Constr.* **2016**, *63*, 37–47. [CrossRef]
22. Kim, H.; Kim, H.; Hong, Y.; Byun, Y. Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. *J. Comput. Civ. Eng.* **2018**, *32*, 04017082. [CrossRef]
23. Angah, O.; Chen, A. Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy. *Autom. Constr.* **2020**, *119*, 103308. [CrossRef]
24. Hu, Q.; Bai, Y.; He, L.; Cai, Q. Intelligent Framework for Worker-Machine Safety Assessment. *J. Constr. Eng. Manag.* **2020**, *146*, 04020045. [CrossRef]
25. Hu, Q.; Ren, Y.; He, L.; Bai, Y. A Novel Vision Based Warning System For Proactive Prevention Of Accidenets Induced By Falling Objects Based On Building And Environmental Engineering Requirements. *Fresenius Environ. Bull.* **2020**, *29*, 7867–7876.
26. Liu, X.; Hu, Y.; Wang, F.; Liang, Y.; Liu, H. Design and realization of a video monitoring system based on the intelligent behavior identify technique. In Proceedings of the 9th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics, Datong, China, 15–17 October 2016.
27. Vigneshkumar, C.; Salve, U.R. A scientometric analysis and review of fall from height research in construction. *Constr. Econ. Build.* **2020**, *20*, 17–35. [CrossRef]
28. Hoła, A.; Sawicki, M.; Szóstak, M. Methodology of Classifying the Causes of Occupational Accidents Involving Construction Scaffolding Using Pareto-Lorenz Analysis. *Appl. Sci.* **2018**, *8*, 48. [CrossRef]
29. Szóstak, M.; Hoła, B.; Bogusławski, P. Identification of accident scenarios involving scaffolding. *Autom. Constr.* **2021**, *126*, 103690. [CrossRef]

30. Martínez-Rojas, M.; Gacto, M.J.; Vitiello, A.; Acampora, G.; Soto-Hidalgo, J.M. An Internet of Things and Fuzzy Markup Language Based Approach to Prevent the Risk of Falling Object Accidents in the Execution Phase of Construction Projects. *Sensors* **2021**, *21*, 6461. [[CrossRef](#)] [[PubMed](#)]
31. Fang, W.; Ding, L.; Zhong, B.; Love, P.; Luo, H. Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach. *Adv. Eng. Inform.* **2018**, *37*, 139–149. [[CrossRef](#)]
32. Fan, H.; Su, H.; Guibas, L. A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 2463–2471. [[CrossRef](#)]
33. Zhu, J.; Wan, X.; Wu, C.; Xu, C. Object detection and localization in 3D environment by fusing raw fisheye image and attitude data. *J. Vis. Commun. Image Represent.* **2019**, *59*, 128–139. [[CrossRef](#)]
34. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005. [[CrossRef](#)]
35. Rubaiyat, A.; Toma, T.; Masoumeh, K.; Rahman, S.; Chen, L.; Ye, Y.; Pan, C. Automatic Detection Of Helmet Uses For Construction Safety. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence Workshops, Leipzig, Germany, 23–26 August 2017.
36. Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2012**, *60*, 84–90. [[CrossRef](#)]
37. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.; An, W. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Autom. Constr.* **2018**, *85*, 1–9. [[CrossRef](#)]
38. Qiao, L.; Qie, Y.; Zhu, Z.; Zhu, Y.; Zaman, U.; Anwer, N. An ontology-based modelling and reasoning framework for assembly sequence planning. *Int. J. Adv. Manuf. Technol.* **2017**, *94*, 4187–4197. [[CrossRef](#)]
39. Fang, W.; Zhong, B.; Zhao, N.; Love, P.; Luo, H.; Xue, J.; Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [[CrossRef](#)]
40. Fang, W.; Ding, L.; Luo, H.; Love, P. Falls from heights: A computer vision-based approach for safety harness detection. *Autom. Constr.* **2018**, *91*, 53–61. [[CrossRef](#)]
41. Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* **2019**, *338*, 321–348. [[CrossRef](#)]
42. Zhu, J.; Zeng, H.; Liao, S.; Lei, Z.; Cai, C.; Zheng, L. Deep Hybrid Similarity Learning for Person Re-Identification. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 3183–3193. [[CrossRef](#)]
43. Hubel, D.; Wiesel, T. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **1959**, *148*, 574–591. [[CrossRef](#)] [[PubMed](#)]
44. Hubel, D.; Wiesel, T. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106–154. [[CrossRef](#)]
45. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016.
46. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2014**, *115*, 211–252. [[CrossRef](#)]
47. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Zitnick, C.L. *Microsoft COCO: Common Objects in Context*; Springer International Publishing: Cham, Switzerland, 2014. [[CrossRef](#)]
48. Kolar, Z.; Chen, H.; Luo, X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. *Autom. Constr.* **2018**, *89*, 58–70. [[CrossRef](#)]
49. Luo, H.; Xiong, C.; Fang, W.; Love, P.; Zhang, B.; Ouyang, X. Convolutional neural networks: Computer vision-based workforce activity assessment in construction. *Autom. Constr.* **2018**, *94*, 282–289. [[CrossRef](#)]
50. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
51. Zeiler, M.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014. [[CrossRef](#)]
52. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Comput. Sci.* **2014**. [[CrossRef](#)]
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 30 June 2016; pp. 770–778. [[CrossRef](#)]
54. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2818–2826.
55. Morteza, H. Learning representations from dendrograms. *Mach. Learn.* **2020**, *109*, 1779–1802. [[CrossRef](#)]
56. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. *J. Mach. Learn. Res.* **2010**, *9*, 249–256.
57. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
58. Chu, M.; Matthews, J.; Love, P. Integrating mobile Building Information Modelling and Augmented Reality systems: An experimental study. *Autom. Constr.* **2018**, *85*, 305–316. [[CrossRef](#)]
59. GB 50870-2013; Unified Code for Technique for Constructional Safety. China Planning Press: Beijing, China, 2013.

60. Wang, H.; Li, X. *Construction Safety Technical Manual*; China Building Materials Press: Beijing, China, 2008.
61. Yu, Y.; Guo, H.; Ding, Q.; Li, H.; Skitmore, M. An experimental study of real-time identification of construction workers' unsafe behaviors. *Autom. Constr.* **2017**, *82*, 193–206. [[CrossRef](#)]
62. Cao, Z.; Hidalgo, M.; Simon, T.; Wei, S.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 172–186. [[CrossRef](#)]
63. Xiao, B.; Lin, Q.; Chen, Y. A vision-based method for automatic tracking of construction machines at nighttime based on deep learning illumination enhancement. *Autom. Constr.* **2021**, *127*, 103721. [[CrossRef](#)]