Automation in Construction 91 (2018) 53–61

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Falls from heights: A computer vision-based approach for safety harness detection

Weili Fang[a,b], Lieyun Ding[a,b,*], Hanbin Luo[a,b], Peter E.D. Love[c]

[a] Dept. of Construction Management, School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei, China
[b] Hubei Engineering Research Center for Virtual, Safe and Automated Construction, (ViSAC), HUST, China
[c] Dept. of Civil Engineering, Curtin University, Perth, Western Australia, Australia

ARTICLE INFO

Keywords:
Convolution neural network
Falls from height
Harness
Unsafe behavior

ABSTRACT

Falls from heights (FFH) are major contributors of injuries and deaths in construction. Yet, despite workers being made aware of the dangers associated with not wearing a safety harness, many forget or purposefully do not wear them when working at heights. To address this problem, this paper develops an automated computer vision-based method that uses two convolutional neural network (CNN) models to determine if workers are wearing their harness when performing tasks while working at heights. The algorithms developed are: (1) a Faster-R-CNN to detect the presence of a worker; and (2) a deep CNN model to identify the harness. A database of photographs of people working at heights was created from activities undertaken on several construction projects in Wuhan, China. The database was then used to test and train the developed networks. The precision and recall rates for the Faster R-CNN were 99% and 95%, and the CNN models 80% and 98%, respectively. The results demonstrate that the developed method can accurately detect workers not wearing their harness. Thus, the computer vision-based approach developed can be used by construction and safety managers as a mechanism to proactively identify unsafe behavior and therefore take immediate action to mitigate the likelihood of a FFH occurring.

1. Introduction

Falls from heights (FFH) are a major problem in construction [1–6]. Research has revealed that FFH account for approximately 48% of serious injuries and 30% of fatalities [7]. Numerous safety policies and procedures have been established to protect people working at heights in construction [8]. For example, scaffolds/platforms and the use fall prevention solutions such as travel restraints systems (e.g. lines and belts) are required when working above a certain height [9].

Yet, despite the considerable amount of research that has been undertaken and the implementation of policies, procedures and the development of protection measures, FFH remain a pervasive problem, particularly for scaffolders and roofers [10]. In China, for example, people working above a height of two metres are required by law to use fall arrest equipment [10]. There has, however, been a reluctance from scaffolders to use a harnesses, in spite of its use being a legal requirement and workers being cognizant of their exposure to a fall [10]. Reasons for such non-compliance have been found to be attributable to discomfort while wearing the harness and the restrictions it place on movement [10]. While such reasons may well have a degree of validity, such workers tend to have a poor awareness and risk perception. Thus, good communication, effective consultation, improved training and reasonable adjustments can often be enough to head off objections to wearing a harness.

But more fundamentally, behavioral and cultural change is required to address the reluctance to wear personal protective equipment (PPE), but this can take a considerable amount of time to implement. To expedite and enact behavioral change, it is suggested that real-time monitoring of harnesses that are worn by people working at heights can contribute to preventing falls. Construction and safety managers require practical methods to monitor and ensure workers are using their harnesses, particularly scaffolders. However, the safety inspection process can be toilsome and is often undertaken intermittently [11]. As a result, safety compliance is unable to be assured and therefore the likelihood a FFH remains a risk.

To address this problem, the research presented in this paper develops an automatic and non-invasive approach using a computer-vision-based method to monitor the use of harnesses. Computer vision-based methods have been widely used in construction. For example, to track workers on-site [12,13], progress monitoring [14], productivity

* Corresponding author at: Dept. of Construction Management, School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei, China.
E-mail address: dly@hust.edu.cn (L. Ding).

https://doi.org/10.1016/j.autcon.2018.02.018
Received 15 September 2017; Received in revised form 21 January 2018; Accepted 7 February 2018
Available online 08 March 2018
0926-5805/

analysis [15], safety and health monitoring [5], automated documentation [5], and postural ergonomic assessment [16].

In comparison with sensing techniques (e.g., Radio Frequency Identification (RFID), Geographical Positioning Systems (GPS) and Ultra-Wide Band (UWB)), which tend to be limited to providing location data for a specific entity being monitored, computer vision can provide a rich set of information (e.g., locations and behaviors of project entities and site conditionss) by analyzing images or videos [5]. While technologies, such as RFID, for instance, have been widely used in construction and applied to an array of PPE types [17], there has been an absence of research that has monitored the use of harnesses. The paper commences by reviewing existing methods that have been used to prevent FFH in construction and then introduces a novel approach based on convolutional neural network (CNN) that can be used to monitor the use of safety harnesses worn by people working at heights. The technical challenges of the developed CNN approach are presented and the implications for future research are identified.

2. Falls from heights

The unique, dynamic, and complex working environment of construction sites and non-standardized design and work procedures can increase workers' exposure to hazards [5]. The prevention of FFH has received a significant amount of attention from construction safety and health management researchers and professionals [18]. Undertaking regular safety inspections and risk assessments to identify hazards has been repeatedly identified in the literature as a core activity to preventing the occurrence of falls [7,19]. A comprehensive review of the FHH literature is therefore eschewed, as this can be found in Nadhim et al. [7]. But for the purposes of brevity key studies that are aligned with the research presented in this paper are drawn upon.

Strategies to prevent and mitigate the severity of injuries can be categorized as being passive or proactive [7]. Strategies of a passive nature are based on analyzing fall accident data to develop future prevention plans. For example, identifying those factors that have contributed to fatal occupational falls from accident reports and acquiring data from regular safety inspections [20]. FFH preventive measures that have been derived from an analysis of accident records and autopsy records include [20]: (1) fixed barriers; (2) travel restraint systems (e.g., belts), fall arrest systems (e.g., harness); and (3) fall containment systems (e.g., nets). Factors that have been found to contribute to roofers FFH include cognitive slips and lapses, weather, and schedule demands [21]. The emergent risk factors contributing to FFH are often prioritized and then used to develop mitigation strategies [22,23]. For example, an automated Building Information Modeling-based safety checking platform that is integrated with safety risks has been developed, which supports fall prevention planning prior to the commencement of construction [24,25].

Proactive strategies are precautionary measures that place emphasis on safety training and education. For example, the implementation of specific fall protection training programs [19]; and the design of short courses, seminars and talks that focus on working the risks of working at heights with the aim to improve people's safety behavior. While enforcement of regulations may increase the use of PPE [1], this is a reactive approach to addressing the issue of safety and does not necessarily change people behaviors [5]. Hence, it is more important to influence the mind-sets, attitudes and culture (i.e. values and beliefs) of workers than solving specific violations [26]. It has been suggested that effective measures to enhance the use of PPE are needed, especially in the context of FFH, as scaffolders, are often reluctant use their harness [10]. The purpose of harness monitoring is to ensure that it is being used correctly by workers and to ensure an organization's safety and health plans and standards are being met.

3. Computer vision-based approaches

Computer vision is an interdisciplinary field of endeavour that deals with how computers can acquire a high-level of understanding from digital images or videos. From an engineering perspective, it seeks to automate tasks that the human visual system is unable to do. Vision-based applications have been developed to capture and process video [27–29]. This had been aided by the development of new algorithms (e.g. Faster R-CNN) that can be used to detect and track resources (e.g., people, plant and equipment), as well as identify the unsafe behavior of workers [30–36].

A fundamental tenet of computer vision-based is action recognition, which is used to exploit handcrafted features (e.g., shapes) from images or videos. To extract features of workers' actions, descriptors such as Histogram of Oriented Gradients (HOG) [37], Histogram of Optical Flow (HOF) [38], and Bag-of-features (Bof) [39] have all been employed to compute on the image or videos. Hand-crafted feature-based methods usually employ a three-stage procedure, which consists of: (1) extraction; (2) representation; and (3) classification.

Image representation that is used to recognize human actions can extract features such as shapes and temporal motions from images. Action recognition features, however, need to contain rich information so that a wide range of actions can be identified and analyzed. Techniques that can be used to analyze such features include classifier tools (e.g. Support Vector Machine (SVM)), temporal state-space models (e.g., Hidden Markov models (HMM), conditional random fields (CRF)), and detection-based methods (e.g., bag-of-words coding). However, the use of these approaches may lead to overfitting and therefore weaken the ability to derive generalizations from a dataset.

Another approach that is often used to collect motion data from stereo videos and to reconstruct a three-dimensional (3D) skeleton model are depth sensors (Kinect™) [5,40–45]. Kinect™ and multiple video cameras have been used to monitor the behavior of workers by estimating the positioning of individual joints in 3D [41–45]. This method provides a useful way to obtain accurate motion data. But more specifically, it provides the ability to record, model, and analyze the human motion that has occurred from the performing an unsafe act. However, monitoring the positioning of workers using 3D can require lengthy computational periods and the line of motion may also be hampered by sensitivities in lighting [46,47].

4. Convolutional neural networks in construction

Deep learning methods that incorporate CNNs have been demonstrated to be effective for computer vision and pattern recognition [48,49]. LeCun et al. [50] developed the LeNet-5 (a CNN model), which recognizes handwritten numbers, based on a dataset created by the Mixed National Institute of Standards and Technology. CNN models can effectively and automatically recognize features from static images by stacking multiple convolutional and pooling layers.

Krizhevsky et al. [49] was the first to achieve substantially high levels of image classification accuracy at the ImageNet Large Scale Visual Recognition Challenge (LSVRC) by training a deep CNN. Since the inception of Krizhevsky's [41] deep CNN, almost all the effective algorithms used for image classification, object recognition, and visual tracking that have been developed are based on this fundamental work. Krizhevsky et al. [49] used deep CNNs to classify 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into a thousand different classes. Hong et al. [51] proposed an online visual tracking algorithm by learning a discriminative saliency map using a CNN, which provided superior results compared to other state-of-the-art tracking algorithms (e.g., discrete fourier transform (DSK), local sparse and K-selection (LSK), and circulant tructure of tracking-by-detection with kernels (CSK)).

The success of region-based CNNs and region proposal methods has prompted advancements in object detection and their use in

construction for the purpose of visual detection [52,53]. For example, Ding et al. [54] proposed a hybrid learning model that integrated CNNs and long short-term memory (LSTM) to detect worker unsafe behavior. Fang et al. [55] developed a computer vision method to detect non-hardhat-use workers on jobsites. Cha et al. [56] developed a deep architecture of CNN for detecting concrete cracks without extracting the defect's feature. Similarly, Feng et al. [57] constructed a deep active learning system to detect defects in structure and classify in an image. Roberts et al. [58] adapted to CNN to detect and classify cranes for monitoring safety hazards using Unmanned Aerial Vehicles. Noteworthy, the use of a CNN requires a significantly large dataset to train its capacity to learn [59].

A popular state-of-the-art detection network is the *Faster R-CNN*. It consists of a fully convolutional region proposal network (RPN) for proposing candidate regions and a downstream classifier [60]. Essentially, the Faster R-CNN is able to accurately identify objects more than any other deep learning method that has been proposed [52]. A major challenge confronting the detection of a harness is its color; it is often similar to that of a worker's clothing.

5. Research approach

In tackling this problem, a design science research approach is adopted to design and develop a CNN that can automatically detect workers who are not wearing their safety harness. Design science focuses on describing, explaining and predicting the current natural or social world, by not only understanding problems, but also designing solutions to improve human performance [61,62]. In doing so, design science can be used to develop the corresponding knowledge and applications to design and implement a product that has value to an organization [63,64]. The research process used to design and develop the CNN for detecting harness compliance by workers is presented in Fig. 1.

5.1. Design development of human-harness network

5.1.1. Design of human network

The Faster R-CNN detection network is selected for this research due to its ability to accurately identify objects with a minimal time lag. The Faster R-CNN employs the Zeiler and Fergus network [65], which comprises of five convolutional layers. The Faster R-CNN model effectively initiates the acquisition of the feature maps through its extraction from within the CNN. The CNN then combines the RPN and Fast R-CNN. As a result, it directly connects the proposal extracted by the RPN to the regions of interest (ROI) pooling layer. This enables the end-to-end target detection of the CNN, thereby accelerating its speed to identify the target.

The core module of the Faster R-CNN is the RPN. The RPN employs $n \times n$ spatial windows that slides onto the feature map of the last convolution (Conv) layer of the original image. Each sliding window is mapped into a D-dimensional vector. Then, it is used as the input for two fully connected (f_c) layers, namely, the box classification (*cls*) and box regression (*reg*) layers. The former provides the probability of objects/non-objects, and the latter provides the coordinates of the predicted object bounding box (Bbox). When $n \times n$ sliding windows reach the end of the convolution feature matrix, the *cls* layer outputs $2k$ scores that represent the probability of the anchor that belongs to the foreground or the background, and the *reg* layer outputs $4k$ coordinates that represent the transformation parameters of the real target frame.

5.1.2. Design of safety harness network

In developing the process to test for the detection of a harness, a crop of worker patches obtained from Faster R-CNN are re-entered by constructing a depth neural network. This is undertaken to develop the

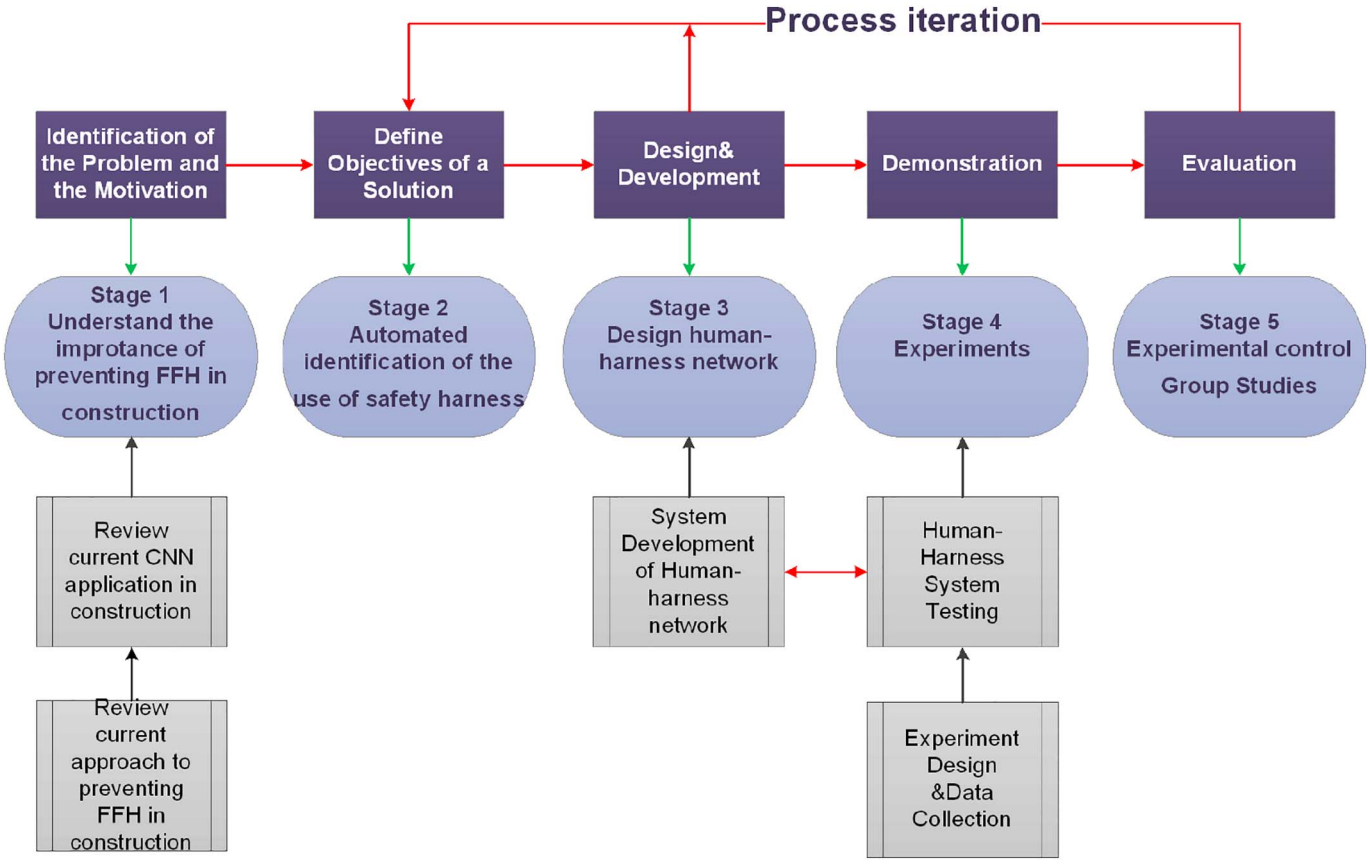


Fig. 1. Design science approach: Research process [61].

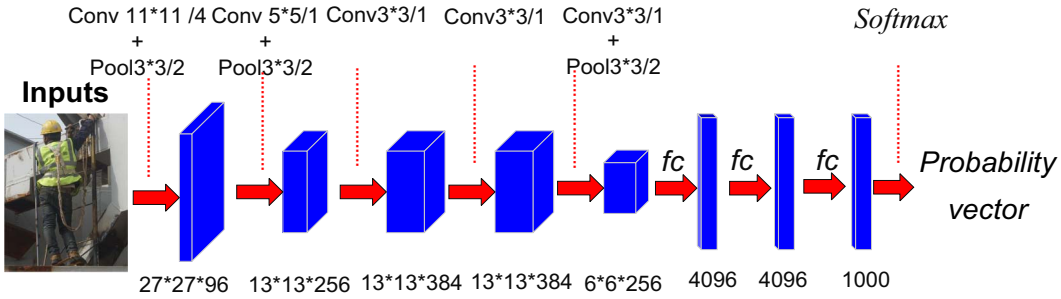


Fig. 2. Architecture of safety harness detection network.

CNNs ability to learn and classify those workers wearing a harness from an image by using forward propagation and gradient processes. The process to crop an image for training the CNN is as follows:

The output of the Faster R-CNN is O_F:

$[[p, x_1, y_1, x_2, y_2]_1\ [p, x_1, y_1, x_2, y_2]_2 \ldots [p, x_1, y_1, x_2, y_2]_n]$.

For i in Range (length (O_F)):

$D[i] = I\ [x_1^{(i)}: x_2^{(i)}, y_1^{(i)}: y_2^{(i)}:]$

Where, p is the confidence of the classification results; (x_1, y_1) is the upper-left coordinate of the rectangle; (x_2, y_2) is lower-right coordinate of the rectangle; n is number of detected human; I is the matrix of original image, the dimensional of matrix is three (length, width, RGB); D is an assembly of output detected human images matrix.

As denoted in Fig. 2 the harness detection network consists of five convolutional layers, three fully connected layers, and one *Softmax* classifier layer (Fig. 2). The *Softmax* function [56] used in the classification process is expressed as a probabilistic function:

$$P(y^{(i)} = n \mid x^{(i)}; W) = \begin{bmatrix} y^{(1)} = 1 \mid x^{(1)}; W \\ y^{(2)} = n \mid x^{(2)}; W \\ \vdots \\ y^{(i)} = n \mid x^{(i)}; W \end{bmatrix} = \frac{1}{\sum_{j=1}^{n} e^{W_j^T x^{(i)}}} \begin{bmatrix} e^{W_1^T x^{(i)}} \\ e^{W_2^T x^{(i)}} \\ \vdots \\ e^{W_n^T x^{(i)}} \end{bmatrix} \quad (1)$$

where, P stands for the i^{th} training example out of m number of training examples, the j^{th} class out of n number of classes, and weights W; $W_j^T x^{(i)}$ stands for the inputs of the *Softmax* layers.

The structure of proposed harness-network can be seen in Fig. 3. A detailed description of the deep CNN used as the basis of the research presented in this paper can be found in Krizhevsky et al. [49]. The architecture of deep CNN model is selected due to its ability to accurately classify images. For example, it achieved top-1 and top-5 error rates of 37.5% and 17.0%, which is better than previously developed state-of-the-art methods [42].

It can be seen in Fig. 2 that the processing and output dimensions of each layer for the network vary. The network accepts the original pixel of the input image and produces an output in the form of a probability vector, as noted in Fig. 3.

The implementation of the convolution and pooling layers of the network play a critical role in the process of feature extraction. The convolutional layer is a feature extraction mechanism used to form the eigenvector by setting a filter or a convolution kernel. For each layer, a convolution operation and activation function on the output of the previous layer in the forward propagation phase is employed, which is formalized as:

$$X_{ij}^k = f(W^k * x)_{ij} + b_k \quad (2)$$

where, f is activation function, b_k is the bias for this feature map, W^k is the value of the kernel connected to the kth feature map.

The input of the pooling layer is generally derived from the output of previous convolution layers. Its main function is to maintain a translation invariance (such as rotation, translation, and expansion) and reduce the number of parameters to prevent overfitting.

6. Experiments

The developed CNN framework was tested using an experiment to detect people who were not wearing their harness (Fig. 4). All algorithms were performed on a server with a 2.40 GHz Intel(R) Xeon(R) E5-2680 CPU, NVIDIA(R) TITAN X GPU, and 64 GB RAM. For the purpose of this research the Python programming language was used. The Caffe deep learning framework was drawn upon, which enabled labelled data to be fed into a Python interface so the calculation and updating of weights could be performed.

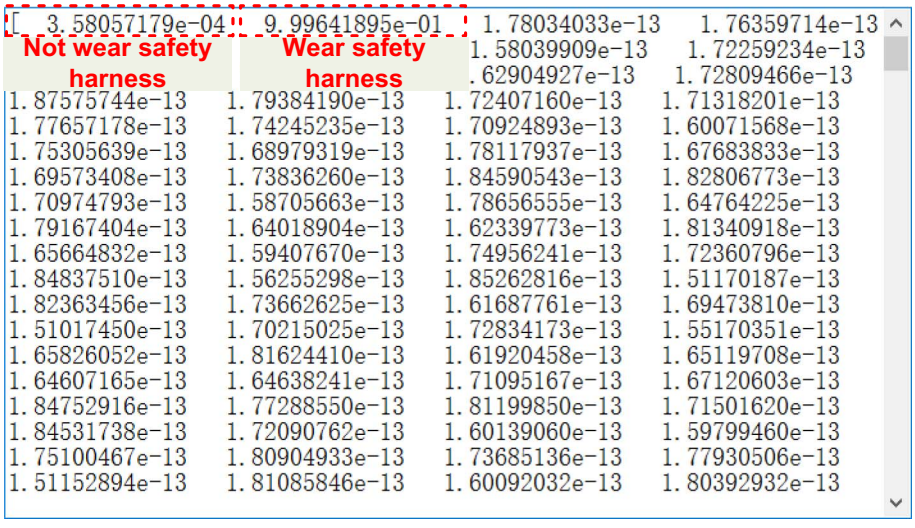
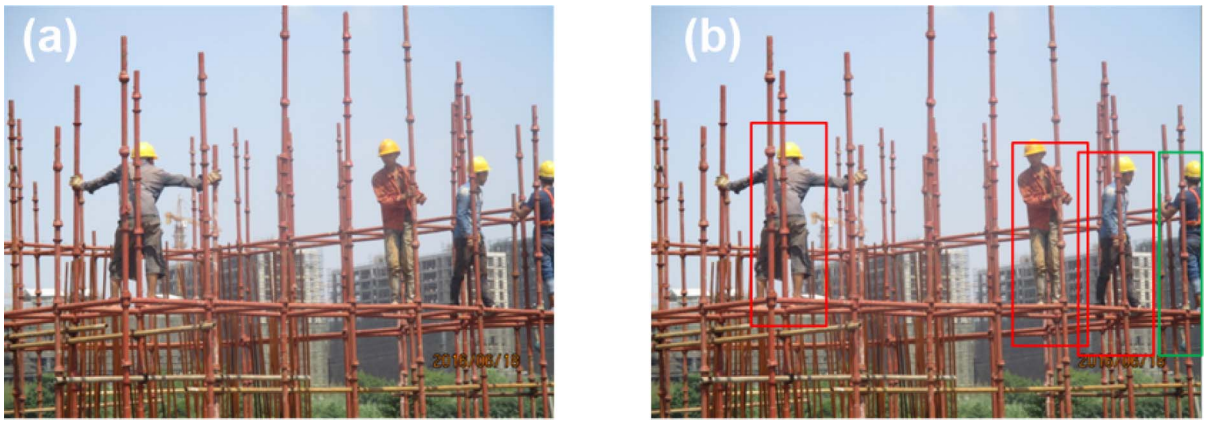


Fig. 3. Example of an output a probability vector.



(a) Workers *not* wearing harness (b) Workers are identified as they did not wear harness

Fig. 4. Detection of workers not wearing harness.

To assess the performance of the detection algorithm the following two measures are used: (1) precision (i.e. the fraction of relevant instances among those retrieved); and (2) recall (i.e. the fraction of relevant instances that have been retrieved over the total amount). These measures are widely used to determine the ability of a system to classify objects [66]. The value of precision and recall rates can be calculated using Eq. (3):

$$\begin{cases} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \end{cases} \quad (3)$$

where, *TP* is the number of true positives, *FN* is the number of false negatives, *FP* is the number of false positives, and TN is the number of true negatives.

6.1. Experiment design and data collection

Prior to testing the algorithm, a comprehensive dataset of images of workers, equipment and materials from construction sites is needed for the purpose of training. However, there is limited access to such datasets, which has hindered the implementation of intelligent monitoring systems in construction [5]. Due to the unavailability of such datasets, they were created to overcome this limitation.

Using a monocular camera, a dataset of 770 images of people working at heights was collected from a number of construction sites in Wuhan, China. Video recordings were also collected of people working at various heights that had been involved with welding steel beams and reinforcement. The experimental dataset was randomly divided into two parts: (1) training and (2) testing. To avoid bias, different views, scale, occlusions, and illumination needed to be considered when creating the collection of images that formed the datasets (Fig. 5). A subset of 693 randomly selected positive images (i.e. workers wearing a harness) and > 5000 negative images (i.e. workers *not* wear a harness) were used to extract and generalize image features during the algorithms training stage. A subset of 77 images (i.e. workers wearing a harness) and other 53 images (workers not wearing a harness) that included different scales, occlusions, illumination, and other characteristics were randomly selected as test data.

6.2. Human-harness system testing

6.2.1. Processing of worker detection

As noted above, a Faster R-CNN is used to accurately detect workers in real time with an example being presented in Fig. 6. The original image was inputted into the CNN to extract and generalize image features, which were subsequently shared by the RPN and Fast R-CNN (FRCN) as their respective input. The RPN module was used to extract the region proposal from the convolution network feature map, which enabled its target score and regressed bounds to be acquired.

To deal with the different scales and aspect ratios of objects, anchors were introduced in the RPN. An anchor was placed in the center of each spatial window at each sliding location of the convolutional maps. Three different scales (128^2, 256^2, 512^2) and aspect ratios (1:1, 1:2, 2:1) were set, and $k = 9$ anchors were placed at each location. Each proposal is parameterized to correspond to an anchor. If the size of the feature map in the last convolution layer is $H \times W$, then the number of possible proposals in a feature map would be $H \times W \times k$.

The ROI classification and regression network module pool and

Fig. 5. Examples of different working environments.

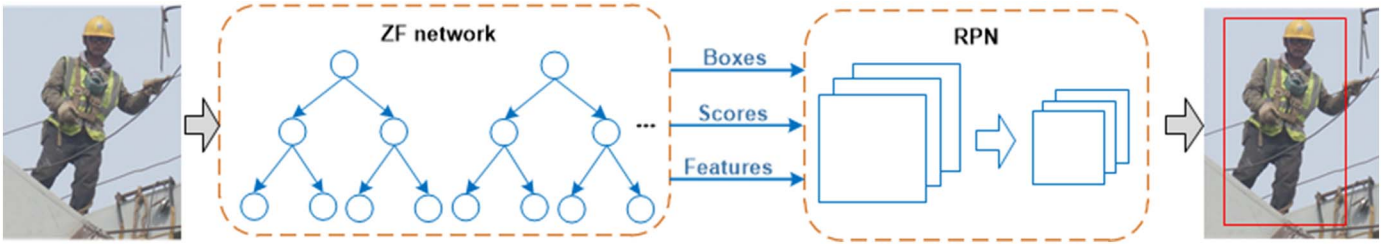


Fig. 6. The pipeline of detecting a worker from an image.

process feature map of the convolution network so that the target recognition and decision making can be achieved. This process is undertaken to share the same feature map between the RPN and FRCN.

The research of Ren et al. [45] was drawn upon to train the Faster R-CNN. The parameters used to train the Faster R-CNN are: (a) base-learning rate is 0.001; (b) step size is 10,000; and (c) Gamma is 0.1. For the purpose of tuning, the base learning rate and step size were adjusted and after 10,000 iterations, drop a Gramma (0.1) was observed.

6.2.2. Process to detect a harness

Fig. 6 (right side) identifies a red rectangular box surrounding the body of a person. This indicates that the program is searching for a harness. On detection of the worker, the coordinates of the rectangular box can be obtained. Then, rectangular box's pixels can be fed into the detection model as inputs. The harness in the images can be manually identified as being a positive training sample. When the network accepts the original pixel of the input image, and an output is generated based on the results obtained from the picture.

Taking the first convolution layer of this network as an example, its input is an image with a size of 227 × 227 × 3, using 96 filters with a size of 11 × 11. A total of 96 features maps with a size of 55 × 55 can be obtained using the convolution Eq. (1). The visual work is then performed to print the output of the first convolution layer, as shown in the middle of Fig. 7. Each convolution kernel presents characteristics of an image, such as the directions of the edges, and provides a feature map that records the different aspects of the image.

In the case of the first pooling layer, its input is 96 feature graphs with a size of 55 × 55. After drop sampling by a pooling factor of 3 × 3, 96 eigenvectors with a size of 27 × 27 are obtained (Fig. 7c).

6.3. Human-safety network evaluation

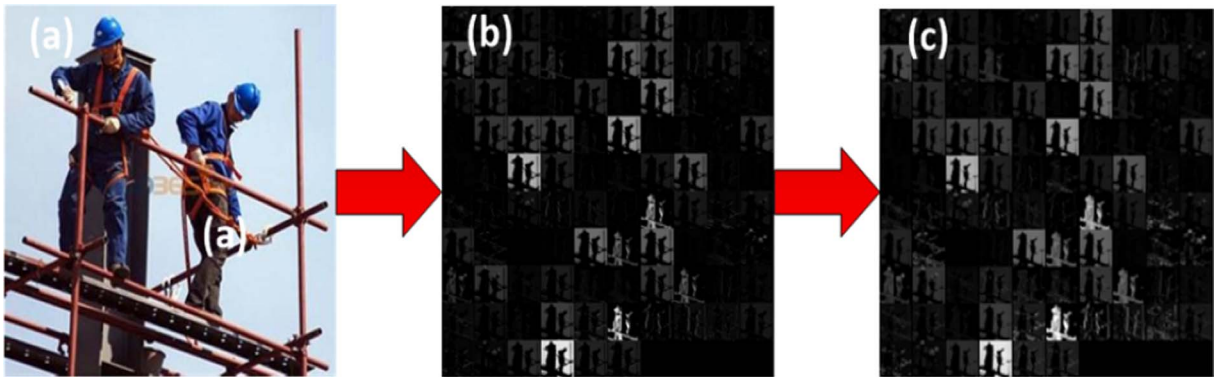
The results of the automatic detection of images that were recorded are presented in Table 1. At a high confidence threshold, the detection results are ambiguous and possess a high precision but low recall, and therefore are rejected. However, at a low confidence threshold, the results are more ambiguous with a high recall and low precision and thus are accepted. To obtain a high precision-recall rate, an acceptance threshold value of 0.8 for them is derived [52]. A value < 0.8, indicates a person has not or been falsely detected. As observed in Fig. 5, each image in the dataset is taken at a unique scale and in a specific pose, illumination, and occlusion condition. The *red rectangular box* illustrates workers without a harness. However, the *green rectangular box* identifies workers wearing their harness. A sample of the harness detection results are presented in Figs. 8 and 9.

Table 1
Detection results with a testing dataset in which workers are randomly taken (detection with $p > 0.8$).

Metric	Worker detection	Harness detection
Correctly detected (TP)	249	198
Mis-detected (FP)	2	51
Not detected (FN)	12	2
Precision	99%	80%
Recall	95%	98%

Note: TP is defined as the number of correctly detected workers/harness. FP is the number of incorrectly detected workers/harness, and FN is the number of undetected workers/harness.

Table 1 indicates that the developed CNN models are able to successfully detect workers and their harness within images. Notably, however, workers were able to be detected more easily than their harness due to occlusions. There was also similarity between the colors of the workers' clothes and their harness. However, the use of the proposed CNNs can improve the ability to train data under varying conditions and therefore increase the likelihood of detecting a harness. Several examples of TPs, FPs, and FN for the detection of the harness are shown in Figs. 10 and 11.



(a) Original map (b) Output map of the first convolution layer (c) Output map of the first pool layer

Fig. 7. Feature maps visualization.

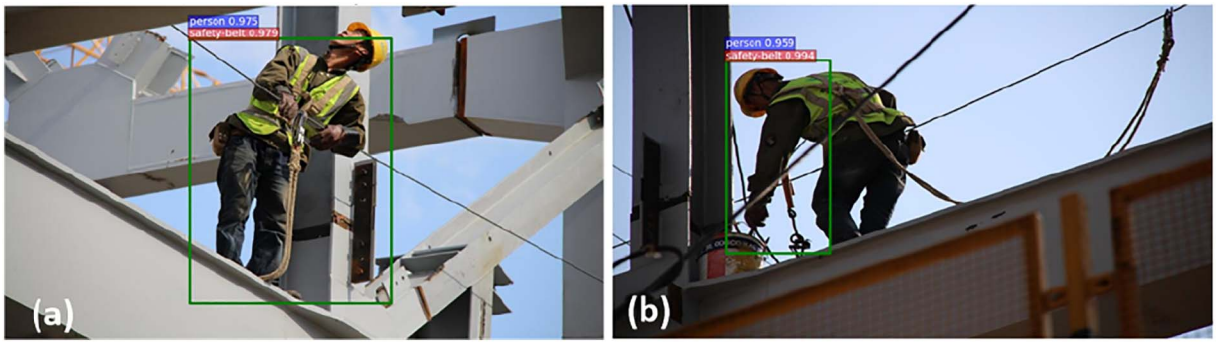


Fig. 8. Detection results of workers wearing harness.

Fig. 9. Detection results of workers not wearing a harness.

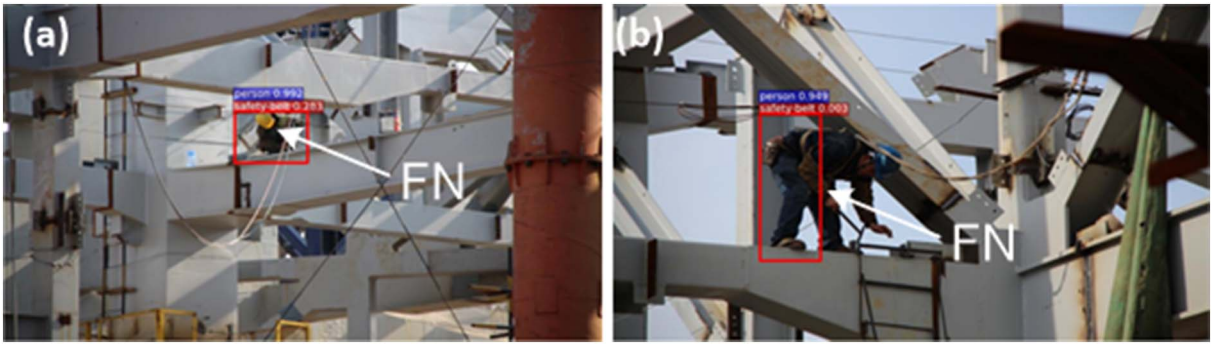


Fig. 10. Examples of false detections.

7. Discussion

To improve the efficiency and effectiveness of the safety inspection process and reduce the number of FFH incidents, a computer-vision approach to detect if workers are wearing their harness was developed. The approach provides site management with a mechanism to proactively identify unsafe behavior and take immediate action to mitigate the likelihood of FFH. It can also act as a safety intervention, as it can be used by site management as a means to highlight potential hazards to workers and the possible consequences that may materialize from their actions. If workers are made aware that they are being monitored, there is a greater likelihood that they will adhere to safety regulations. Contractors have a duty of care to protect their workers against health and safety hazards at work. It is, therefore, in the interest of all parties to have a harness monitoring system in place. In countries where there is a strong trade union presence within the construction industry such as Australia, and the United Kingdom, for example, the use of automatic harness monitoring systems would require contractors to collaborate with them to ensure the system was not being used for any other intended purpose and to penalize workers.

As noted above, there has been an absence of research that has focused on determining if people are wearing their harness while working at heights. In addressing this void, the research presented in this paper has demonstrated that the use of deep CNN can be used to accurately detect if a harness is being worn. Challenges associated with the color of the harness, differing viewpoint and illumination were encountered, but by using a combination of CNNs these issues were able to be overcome and ensure high degree of detection accuracy. With this in mind, enabling a robust method that can be used to automatically detect unsafe behavior.

8. Limitations

In spite of the novelty of the research that has been presented, it

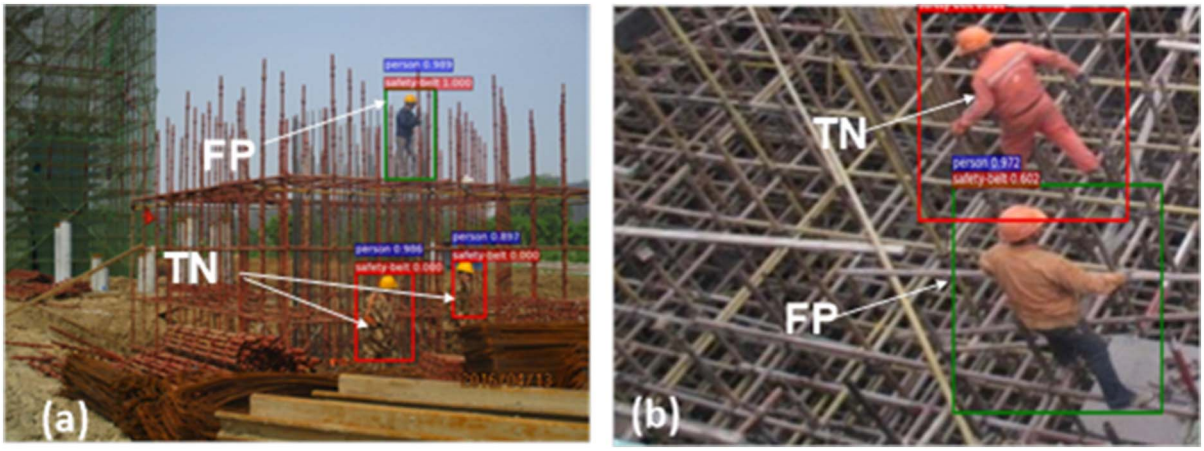


Fig. 11. Examples of missed harness detections.

needs to be acknowledged that several limitations exist. The study was limited to a select number of activities working at heights. Future research, is therefore, required to in increase the scope of activities that are examined and develop an automatic monitoring systems that can recognize unsafe behaviors. In some instances, the models were not able to detect workers and their harnesses. This was due to a number of issues, the sample size and the harness's color had a direct effect on the CNNs recognition ability. Learning requires numerous samples to achieve good testing results, otherwise, overfitting may arise. In the experiment, the sample of pictures was limited to 693 for training the networks, which resulted in the harness not being unrecognized. This may, in part, be attributable to the limited number of images possessing varying scales, which may have hindered the networks ability to learn and recognize the harness. In addition, cluttered construction sites can occlude the ability recognize the harness [67].

9. Conclusion

A new approach is proposed to detect individual workers who do not wear their harnesses while working at heights on construction sites. Two algorithms were developed: (1) a Faster-R-CNN to detect the presence of workers; and (2) CNN models to identify the harness attached to workers. The results that were achieved demonstrated that by combining the CNNs that a high degree of accuracy could be achieved in detecting workers that were not wearing their harness.

The precision and recall rates for the Faster R-CNN were 99% and 95%, respectively. Similarly, in case of the CNNs the precision and recall rates were marginally lower being 80% and 98%, respectively. It was revealed that the Faster R-CNN provided a better detection accuracy to detect people and the CNN models for the harness. The Faster R-CNN framework enabled the deep models to focus on annotating a worker. In addition, by mapping the feature map and sharing with the RPN and FRCN computation time was reduced. This enabled the extracted feature of the RPN to be directly connected to ROI pooling layer ensuring the CNN to quickly detect the worker.

Despite not being able to recognize the harness with 100% accuracy, the developed deep CNN approach can provide site management with several benefits to their everyday practice. Firstly, safety behavior can be monitored without disturbing people while they working. And secondly, a wide range of working areas can be simulteanously monitored, which can reduce the costs and time associated with inspections. Having a real-time monitoring system in place to monitor people working at heights provides a mechanism to reduce falls and improve safety.

Acknowledgments

This research is supported in part by a major project of The National Social Science Key Fund of China (Grant No.13&ZD175), supported by National Natural Science Foundation of China (Grant No.71732001, No.51678265, No.71301059), supported by “the Fundamental Research Funds for the Central Universities” (Grant NO. 2017KFYXJJ134, Grant No. 2015ZDTD023). The authors would like also acknowledge the constructive and insightful comments provided by the Associate Editor and four anonymous reviewers, which have helped improve the quality of this manuscript.

References

[1] F.P. Rivara, D.C. Thompson, Prevention of falls in the construction industry: evidence for program effectiveness, Am. J. Prev. Med. 18 (4) (2000) 23–26, http://dx.doi.org/10.1016/S0749-3797(00)00137-9.
[2] X. Huang, J. Hinze, Analysis of construction worker fall accidents, J. Constr. Eng. Manag. 129 (3) (2003) 262–271, http://dx.doi.org/10.1061/(asce)0733-9364(2003)129:3(262).
[3] S.M. Whitaker, R.J. Graves, M. James, P. McCann, Safety with access scaffolds: development of a prototype decision aid based on accident analysis, J. Saf. Res. 34 (3) (2003) 249–261, http://dx.doi.org/10.1016/s0022-4375(03)00025-2.
[4] P. Becker, M. Fullen, M. Akladios, G. Hobbs, Prevention of construction falls by organizational intervention, Inj. Prev. 7 (Suppl. 1) (2001) i64–i67, http://dx.doi.org/10.1136/ip.7.suppl_1.i64.
[5] J. Seo, S. Han, S. Lee, H. Kim, Computer vision techniques for construction safety and health monitoring, Adv. Eng. Inform. 29 (2) (2015) 239–251, http://dx.doi.org/10.1016/j.aei.2015.02.001.
[6] M.G. Yang, An empirical investigation of the average deployment force of personal fall arrest energy absorbers, J. Constr. Eng. Manag. 141 (1) (2015), http://dx.doi.org/10.1061/(asce)co.1943-7862.0000910.
[7] E.A. Nadhim, C. Hon, B. Xia, I. Stewart, D. Fang, Falls from height in the construction industry: a critical review of the scientific literature, Int. J. Environ. Res. Public Health 13 (7) (2016) 638, http://dx.doi.org/10.3390/ijerph13070638.
[8] M.-W. Park, N. Elsafty, Z. Zhu, Hardhat-wearing detection for enhancing on-site safety of construction workers, J. Constr. Eng. Manag. 141 (9) (2015) 04015024, , http://dx.doi.org/10.1061/(asce)co.1943-7862.0000974.
[9] K. Hu, H. Rahmandad, T. Smith-Jackson, W. Winchester, Factors influencing the risk of falls in the construction industry: a review of the evidence, Constr. Manag. Econ. 29 (4) (2011) 397–416, http://dx.doi.org/10.1080/01446193.2011.558104.
[10] M. Zhang, D. Fang, A cognitive analysis of why Chinese scaffolders do not use safety harnesses in construction, Constr. Manag. Econ. 31 (3) (2013) 207–222, http://dx.doi.org/10.1080/01446193.2013.764000.
[11] K. Shrestha, P.P. Shrestha, D. Bajracharya, E.A. Yfantis, Hard-hat detection for construction safety visualization, J. Constr. Eng. 2015 (2015), http://dx.doi.org/10.1155/2015/721380.
[12] M. Golparvar-Fard, A. Heydarian, J.C. Niebles, Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, Adv. Eng. Inform. 27 (4) (2013) 652–663, http://dx.doi.org/10.1016/j.aei.2013.09.001.
[13] M.-W. Park, I. Brilakis, Continuous localization of construction workers via integration of detection and tracking, Autom. Constr. 72 (2016) 129–142, http://dx.doi.org/10.1016/j.autcon.2016.08.039.
[14] K.K. Han, M. Golparvar-Fard, Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs, Autom. Constr. 53 (2015) 44–57, http://dx.doi.org/10.1016/j.autcon.2015.02.007.
[15] J. Gong, C.H. Caldas, Computer vision-based video interpretation model for automated productivity analysis of construction operations, J. Comput. Civ. Eng. 24 (3)

(2009) 252–263, http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000027.
[16] J. Seo, K. Yin, S. Lee, Automated postural ergonomic assessment using a computer vision-based posture classification, Construction Research Congress, San Juan, Puerto Rico, USA, 2016, pp. 809–818, , http://dx.doi.org/10.1061/9780784479827.082.
[17] S. Barro-Torres, T.M. Fernández-Caramés, H.J. Pérez-Iglesias, C.J. Escudero, Real-time personal protective equipment monitoring system, Comput. Commun. 36 (1) (2012) 42–50, http://dx.doi.org/10.1016/j.comcom.2012.01.005.
[18] E. Cheung, A.P. Chan, Rapid demountable platform (RDP)—a device for preventing fall from height accidents, Accid. Anal. Prev. 48 (2012) 235–245, http://dx.doi.org/10.1016/j.aap.2011.05.037.
[19] H. Jebelli, C.R. Ahn, T.L. Stentz, Fall risk analysis of construction workers using inertial measurement units: validating the usefulness of the postural stability metrics in construction, Saf. Sci. 84 (2016) 161–170, http://dx.doi.org/10.1016/j.ssci.2015.12.012.
[20] C.-F. Chi, T.-C. Chang, H.-I. Ting, Accident patterns and prevention measures for fatal occupational falls in the construction industry, Appl. Ergon. 36 (4) (2005) 391–400, http://dx.doi.org/10.1016/j.apergo.2004.09.011.
[21] H. Cakan, E. Kazan, M. Usmen, Investigation of factors contributing to fatal and nonfatal roofer fall accidents, Int. J. Constr. Educ. Res. 10 (4) (2014) 300–317, http://dx.doi.org/10.1080/15578771.2013.868843.
[22] O. Aneziris, I.A. Papazoglou, H. Baksteen, M. Mud, B. Ale, L.J. Bellamy, A.R. Hale, A. Bloemhoff, J. Post, J. Oh, Quantified risk assessment for fall from height, Saf. Sci. 46 (2) (2008) 198–220, http://dx.doi.org/10.1016/j.ssci.2007.06.034.
[23] P. Kines, Construction workers' falls through roofs: fatal versus serious injuries, J. Saf. Res. 33 (2) (2002) 195–208, http://dx.doi.org/10.1016/s0022-4375(02)00019-1.
[24] S. Zhang, J. Teizer, J.K. Lee, C.M. Eastman, M. Venugopal, Building information modeling (BIM) and safety: automatic safety checking of construction models and schedules, Autom. Constr. 29 (4) (2013) 183–195, http://dx.doi.org/10.1016/j.autcon.2012.05.006.
[25] S. Zhang, K. Sulankivi, M. Kiviniemi, I. Romo, C.M. Eastman, J. Teizer, BIM-based fall hazard identification and prevention in construction safety planning, Saf. Sci. 72 (8) (2015) 31–45, http://dx.doi.org/10.1016/j.ssci.2014.08.001.
[26] K.J. Nielsen, A comparison of inspection practices within the construction industry between the Danish and Swedish Work Environment Authorities, Constr. Manag. Econ. 35 (3) (2017) 154–169, http://dx.doi.org/10.1080/01446193.2016.1231407.
[27] T. Guan, L. Duan, J. Yu, Y. Chen, X. Zhang, Real-time camera pose estimation for wide-area augmented reality applications, IEEE Comput. Graph. Appl. 31 (3) (2011) 56–68, http://dx.doi.org/10.1109/mcg.2010.23.
[28] H. Pan, T. Guan, Y. Luo, L. Duan, Y. Tian, L. Yi, Y. Zhao, J. Yu, Dense 3D reconstruction combining depth and RGB information, Neurocomputing 175 (PA) (2016) 644–651, http://dx.doi.org/10.1016/j.neucom.2015.10.104.
[29] Z. Wang, J. Yu, Y. He, T. Guan, Affection arousal based highlight extraction for soccer video, Multimedia Tools Appl. 73 (1) (2014) 519–546, http://dx.doi.org/10.1007/s11042-013-1619-1.
[30] Y. Yu, H. Guo, Q. Ding, H. Li, M. Skitmore, An experimental study of real-time identification of construction workers' unsafe behaviors, Autom. Constr. (2017), http://dx.doi.org/10.1016/j.autcon.2017.05.002.
[31] H. Kim, K. Kim, H. Kim, Vision-based object-centric safety assessment using fuzzy inference: monitoring struck-by accidents with moving objects, J. Comput. Civ. Eng. 30 (4) (2016), http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000562.
[32] Z. Zhu, X. Ren, Z. Chen, Visual tracking of construction jobsite workforce and equipment with particle filtering, J. Comput. Civ. Eng. 30 (6) (2016), http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000573.
[33] L. Ma, R. Sacks, R. Zeibak-Shini, Information modeling of earthquake-damaged reinforced concrete structures, Adv. Eng. Inform. 29 (3) (2015) 396–407, http://dx.doi.org/10.1016/j.aei.2015.01.007.
[34] R. Zeibak-Shini, R. Sacks, L. Ma, S. Filin, Towards generation of as-damaged BIM models using laser-scanning and as-built BIM: first estimate of as-damaged locations of reinforced concrete frame members in masonry infill structures, Adv. Eng. Inform. 30 (3) (2016) 312–326, http://dx.doi.org/10.1016/j.aei.2016.04.001.
[35] L. Ma, R. Sacks, U. Kattel, T. Bloch, 3D object classification using geometric features and pairwise relationships, Comput. Aided Civ. Inf. Eng. (2017), http://dx.doi.org/10.1111/mice.12336.
[36] R. Sacks, A. Kedar, A. Borrmann, L. Ma, I. Brilakis, P. Hüthwohl, S. Daum, U. Kattel, R. Yosef, T. Liebich, B.E. Barutcu, S. Muhic, SeeBridge as next generation bridge inspection: overview, information delivery manual and model view definition, Autom. Constr. 90 (2018) 134–145, http://dx.doi.org/10.1016/j.autcon.2018.02.033.
[37] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, computer vision and pattern recognition, IEEE Computer Society Conference, San Diego, CA, USA, June 20–26 2005, pp. 886–893, , http://dx.doi.org/10.1109/cvpr.2005.177.
[38] H. Wang, C. Schmid, Action recognition with improved trajectories, IEEE International Conference on Computer Vision, 2014, pp. 3551–3558, , http://dx.doi.org/10.1109/iccv.2013.441.
[39] T. Guan, Y. Fan, L. Duan, J. Yu, On-device mobile visual location recognition by using panoramic images and compressed sensing based visual descriptors, PLoS One 9 (6) (2014) e98806, , http://dx.doi.org/10.1371/journal.pone.0098806.
[40] S.J. Ray, J. Teizer, Real-time construction worker posture analysis for ergonomics training, Adv. Eng. Inform. 26 (2) (2012) 439–455, http://dx.doi.org/10.1016/j.aei.2012.02.011.
[41] S. Han, S. Lee, A vision-based motion capture and recognition framework for behavior-based safety management, Autom. Constr. 35 (2013) 131–141, http://dx.doi.org/10.1016/j.autcon.2013.05.001.
[42] M. Liu, S. Han, S. Lee, Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis, Constr. Innov. 16 (3) (2016) 348–367, http://dx.doi.org/10.1108/ci-10-2015-0054.
[43] S. Han, S. Lee, F. Peña-Mora, Comparative study of motion features for similarity-based modeling and classification of unsafe actions in construction, J. Comput. Civ. Eng. 28 (5) (2014) A4014005, , http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000339.
[44] S.U. Han, M. Achar, S.H. Lee, F. Peña-Mora, Empirical assessment of a RGB-D sensor on motion capture and action recognition for construction worker monitoring, Vis. Eng. 1 (1) (2013) 6, http://dx.doi.org/10.1186/2213-7459-1-6.
[45] S. Han, S. Lee, F. Peña-Mora, Vision-based detection of unsafe actions of a construction worker: case study of ladder climbing, J. Comput. Civ. Eng. 27 (6) (2013) 635–644, http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000279.
[46] D. Wang, F. Dai, X. Ning, Risk assessment of work-related musculoskeletal disorders in construction: state-of-the-art review, J. Constr. Eng. Manag. 141 (6) (2015), http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000979.
[47] J. Gong, C.H. Caldas, C. Gordon, Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models, Adv. Eng. Inform. 25 (4) (2011) 771–782, http://dx.doi.org/10.1016/j.aei.2011.06.002.
[48] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117, http://dx.doi.org/10.1016/j.neunet.2014.09.003.
[49] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105, , http://dx.doi.org/10.1145/3065386.
[50] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, http://dx.doi.org/10.1109/9780470544976.ch9.
[51] S. Hong, T. You, S. Kwak, B. Han, Online tracking by learning discriminative saliency map with convolutional neural network, Comput. Sci. (2015) 597–606 (arXiv:1502.06796v1).
[52] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. (2015) 91–99, http://dx.doi.org/10.1109/tpami.2016.2577031.
[53] Q. Fang, H. Li, X. Luo, L. Ding, T.M. Rose, W. An, Y. Yu, A deep learning-based method for detecting non-certified work on construction sites, Adv. Eng. Inf. 35 (2018) 56–68, http://dx.doi.org/10.1016/j.aei.2018.01.001.
[54] L. Ding, W. Fang, H. Luo, P.E.D. Love, B. Zhong, X. Ouyang, A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory, Autom. Constr. 86 (2018) 118–124, http://dx.doi.org/10.1016/j.autcon.2017.11.002.
[55] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T.M. Rose, W. An, Detecting non-hardhat-use by a deep learning method from far-field surveillance videos, Autom. Constr. 85 (2018) 1–9, http://dx.doi.org/10.1016/j.autcon.2017.09.018.
[56] Y.J. Cha, W. Choi, O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, Comput. Aided Civ. Inf. Eng. 32 (5) (2017) 361–378, http://dx.doi.org/10.1111/mice.12263.
[57] C. Feng, M.Y. Liu, C.C. Kao, T.Y. Lee, Deep active learning for civil infrastructure defect detection and classification, ASCE International Workshop on Computing in Civil Engineering, ASCE, Seattle, Washington, USA, 25–27 June 2017, 2017, pp. 298–306, , http://dx.doi.org/10.1061/9780784480823.036.
[58] D. Roberts, T. Bretl, M. Golparvar-Fard, Detecting and Classifying cranes using camera-equipped UAVs for monitoring crane-related safety hazards, ASCE International Workshop on Computing in Civil Engineering 2017, ASCE, Seattle, Washington, USA, 25–27 June 2017, 2017, pp. 442–449, , http://dx.doi.org/10.1061/9780784480847.055.
[59] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105 http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.
[60] L. Zhang, L. Lin, X. Liang, K. He, Is faster R-CNN doing well for pedestrian detection? European Conference on Computer Vision, Springer, 2016, pp. 443–457, , http://dx.doi.org/10.1007/978-3-319-46475-6_28.
[61] J.E.V. Aken, Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules, J. Manag. Stud. 41 (2) (2004) 219–246, http://dx.doi.org/10.1111/j.1467-6486.2004.00430.x.
[62] M. Chu, J. Matthews, P.E.D. Love, Integrating mobile building information modelling and augmented reality systems: an experimental study, Autom. Constr. 85 (2018) 305–316, http://dx.doi.org/10.1016/j.autcon.2017.10.032.
[63] J.E.V. A., Management research as a design science: articulating the research products of mode 2 knowledge production in management, Br. J. Manag. 16 (1) (2005) 19–36, http://dx.doi.org/10.1111/j.1467-8551.2005.00437.x.
[64] G.L. Geerts, A design science research methodology and its application to accounting information systems research, Int. J. Account. Inf. Syst. 12 (2) (2011) 142–151, http://dx.doi.org/10.1016/j.accinf.2011.02.004.
[65] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, European conference on computer vision, Springer, 2014, pp. 818–833, , http://dx.doi.org/10.1007/978-3-319-10590-1_53.
[66] D.M.W. Powers, Evaluation: from precision, recall and F-factor to ROC, Informedness, Markedness & Correlation, J. Mach. Learn. Technol. 2 (2011) 2229–3981 http://hdl.handle.net/2328/27165.
[67] J. Yang, M.-W. Park, P.A. Vela, M. Golparvar-Fard, Construction performance monitoring via still images, time-lapse photos, and video streams: now, tomorrow, and the future, Adv. Eng. Inform. 29 (2) (2015) 211–224, http://dx.doi.org/10.1016/j.aei.2015.01.011.