Automation in Construction 85 (2018) 1–9

Contents lists available at ScienceDirect

Automation in Construction

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

Detecting non-hardhat-use by a deep learning method from far-field surveillance videos

Qi Fang[a,b], Heng Li[b,*], Xiaochun Luo[b], Lieyun Ding[a], Hanbin Luo[a], Timothy M. Rose[c], Wangpeng An[d]

[a] School of Civil Engineering & Mechanics, Huazhong University of Science & Technology, Wuhan, China
[b] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong
[c] School of Civil Engineering and Built Environment, Queensland University of Technology, Australia
[d] Department of Computing, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Keywords:
Construction safety
Non-hardhat-use
Far-field surveillance video
Deep learning
Faster R-CNN

ABSTRACT

Hardhats are an important safety measure used to protect construction workers from accidents. However, accidents caused in ignorance of wearing hardhats still occur. In order to strengthen the supervision of construction workers to avoid accidents, automatic non-hardhat-use (NHU) detection technology can play an important role. Existing automatic methods of detecting hardhat avoidance are commonly limited to the detection of objects in near-field surveillance videos. This paper proposes the use of a high precision, high speed and widely applicable Faster R-CNN method to detect construction workers' NHU. To evaluate the performance of Faster R-CNN, more than 100,000 construction worker image frames were randomly selected from the far-field surveillance videos of 25 different construction sites over a period of more than a year. The research analyzed various visual conditions of the construction sites and classified image frames according to their visual conditions. The image frames were input into Faster R-CNN according to different visual categories. The experimental results demonstrate that the high precision, high recall and fast speed of the method can effectively detect construction workers' NHU in different construction site conditions, and can facilitate improved safety inspection and supervision.

1. Introduction

Construction is a high-risk activity requiring construction workers to operate in awkward postures, excessive lifting and high-intensity operations [1], which are key factors leading to occupational injuries. According to the United States' Bureau of Labor Statistics, the number of fatalities in the US has gradually increased from 849 to 985 between 2012 and 2015 [2]. Similarly, according to the UK Health and Safety Executive (HSE), 38 construction workers suffered fatal injuries in Great Britain between April 2014 and March 2015, while this figure rose to 45 [3] during the same period the following year. Such statistics reinforce the need for greater safety measures to reduce the occurrence of construction accidents.

The consequences of head injuries are the most serious of all construction accidents. Although accidents involving legs, feet and toes most lead to some injury, those involving head and neck are often fatal [4]. Many head and neck injuries are caused by falling from height or being stuck by vehicles and other moving plant and equipment. From 2003 to 2010, 2210 construction workers in the United States died as a result of traumatic brain injuries [5,6], accounting for 24% of the total number of deaths from construction accidents.

Wearing a hardhat is an effective protective measure for minimizing the risk of traumatic brain injury. In construction accidents, hardhats protect workers by resisting penetration by objects, absorbing shock from direct blows to the head by objects and reducing electrical shock hazards. The United States Occupational Safety & Health Administration (OSHA) has two standards requiring workers to wear hardhats when there is a potential for head injury from "impacts, falling or flying objects, or electrical shock" [7]. Despite the vital role of hardhats in protecting life, a survey conducted by the US Bureau of Labor Statistics (BLS) suggests that 84% of workers who had suffered impact injuries to the head were not wearing head protection equipment [8].

An automated monitoring method helps to improve the supervision of construction workers to ensure hardhats are appropriately worn. It is argued that traditional safety management methods, such as risk analysis and safety training, are not sufficient to protect construction workers, as it is difficult to accurately predict the prevention of all kinds

[*] Corresponding author.
E-mail address: bshengli@polyu.edu.hk (H. Li).

http://dx.doi.org/10.1016/j.autcon.2017.09.018
Received 17 May 2017; Received in revised form 21 August 2017; Accepted 28 September 2017
Available online 06 October 2017
0926-5805/

of accidents at the design stage [9]. An automatic monitoring method is conducive to the realization of real-time site monitoring, which cannot only save labor costs but also enhance site security.

However, previous research into non-hardhat-use (NHU) detection methods is still beyond practical applications. Although existing methods perform well in near-field pedestrian recognition, they are less effective in far-field surveillance video detection. This is because the image resolution of construction workers is high enough to extract the facial features that are clearly visible in near-field image frames [10]. Additionally, most scenes captured by near-field cameras have a minimal change of background, which is inconsistent with the common ever-changing background of a construction site.

Most construction site surveillance cameras are installed on the site boundary at a high altitude. The location of cameras determines the far-field nature of their shoot. A far-field surveillance video is distinguished from other videos by the small pixel size of workers (as small as 30-pixels tall), broad background and the various postures of individuals [11,12]. Therefore, the continuous movement of equipment, resources, people and the environment is all captured in a far-field video, which is a major challenge to the detection of NHU on a construction site.

This paper proposes an object recognition method for NHU detection in far-field surveillance videos on construction sites. With the aim of verifying the adaptability of the method to the construction environment, this research analyzes the various visual conditions of construction sites and classifies image frames according to their visual conditions. The image frames are then input into the Faster R-CNN model according to their different visual categories. The precision and recall rates of the results are verified to judge how much the Faster R-CNN method is suitable for different construction conditions. The experimental results demonstrate that Faster R-CNN is highly robust for various backgrounds and worker posture changes in NHU detection. The precision and recall rates are all above 90%, which is sufficient to improve construction site safety supervision.

2. Literature review

2.1. Necessity for safety monitoring NHU

A survey conducted by the United States Bureau of Labor Statistics found the proportion of up to 90% traumatic brain injuries as a result of NHU [13,14]. Previous research demonstrates that wearing a hardhat can significantly reduce the probability of skull fracture, neck sprain and concussion [14,15]. As such, it is a statutory requirement to wear hardhats in construction activities all around the world [16,17]. Unfortunately, not all construction workers are aware of the importance of wearing a hardhat. In practice, many workers tend to take off their hardhats [18–20] because of discomfort due to weight and to cool off in high temperatures. If not rectified, this reinforced negative behavior can continue in the future until an accident occurs. Clearly, it is necessary to enhance monitoring of the hardhat use of construction site workers.

The ratio of NHU in a construction work group is related to two conditions [21,22]: 1) the workers' safety awareness and attitude towards construction risk and 2) safety supervision. Effective external supervision can not only prevent unsafe behavior in the first instance, but can also gradually improve the level of safety awareness and attitude of workers [23]. Therefore, an automated supervision method to detect construction workers NHU is required to minimize the risk of accidents and improve their safety.

2.2. Related research into the detection of NHU

At present, research into NHU detection can be divided into sensor-based detection and computer vision-based detection methods.

Sensor-based detection primarily relies on positioning technology to locate workers and hardhats. Kelm, et al. [19] designed a mobile Radio Frequency Identification (RFID) portal for checking Personal Protective Equipment (PPE) compliance of personnel. The RFID readers were located at the construction site entrance, and therefore only those who enter the construction site are checked, while workers in other areas are not. Additionally, the tagging of PPE with a worker's identification card only indicates that the distance between the worker and PPE is close, but unable to determine whether the PPE is being worn, held or has been placed on the ground. Barro-Torres, et al. [24] introduce a novel Cyber Physical System (CPS) to monitor how PPE is worn by workers in real time. Rather than being located at the construction site entrance, their sensors were integrated into the clothing of workers for constant monitoring. However, again, this system is unable to determine if a worker is wearing their hardhat or is just near it. Dong et al. [25] use a location system with virtual construction technology to track whether a worker should be wearing a hardhat and transmit a warning. A pressure sensor is placed in the hardhat to collect and store pressure information to indicate whether the hardhat was being worn, and then transmitted via Bluetooth for monitoring and response. However, if a worker exceeds an acceptable range from the monitoring center for long durations, information on the worker can be lost, making it difficult to identify whether they have been wearing their hardhat when out of range. Further, the Bluetooth devices are required to be regularly charged after a period of use. The need to regularly charge the Bluetooth transmitter can limit its use and practicability on site and can be detrimental to the long-term and widespread use of this technology. In general, the use of existing sensor-based detection and tracking techniques is limited by the need for each construction worker to wear a physical tag or sensor. This can be seen as intrusive to workers and generally requires a large up-front investment in additional equipment, including the physical tag or sensor. Many workers are unwilling to wear such tracking equipment because of health and privacy concerns.

In comparison to localization techniques, image recognition is receiving increased attention on construction sites for its enhanced monitoring abilities. RGB-D cameras, such as Kinect and VICON are one kind of popular tools to collect workers' unsafe behaviors [26–28]. Whereas, RGB-D sensors are limited in the range of around 1 to 4 m [29]. Also susceptibility to interference from sunlight and ferromagnetic radiation, making them unsuitable for NHU detection on construction sites [30]. In this regard, the use of regular cameras, particularly a single camera has a competitive advantage for practical application. However, several problems still exist in automatic NHU detection. For example, Du et al. [31] present a NHU detection method based on facial features, motion and color information. Facial feature and color information recognition methods have two important assumptions: 1) all workers turn their face towards the camera while working and 2) all hardhats are of the same color. These two assumptions can be inconsistent on an actual construction site. Further, Shrestha, et al. [32] use edge detection algorithms to recognize the edge of objects inside the upper head region where a hardhat may be recognized. This method also relies on the recognition of facial features, where workers who turn their face away from the cameras cannot be recognized. Rubaiyat, et al. [33] propose another automatic NHU detection method for construction safety by mixing a Histogram of Oriented Gradient (HOG) with Circle Hough Transform (CHT) to obtain the features of workers and hardhats. Again, this method relies on the detection of facial features and has similar limitations to previous algorithms. Park and Zhu et al. [8,34]develop a new NHU detection algorithm based on HOG that does not rely on the detection of facial features. They can contrast objects captured in images by developing a HOG feature template of a human object. Thus, it will be recognized as a worker if the detected object is similar to the previously proposed template. Compared to previous methods that rely on facial recognition, this method depends on the application of the HOG feature template. However, a limitation of this method is that they will not be recognized by the algorithm if workers, while working, act in different ways from the HOG feature template. Accordingly, this paper presents a new automatic detection algorithm based on Faster R-CNN to address the limitations of previous methods. Faster R-CNN has a faster processing time and higher precision than previous methods, which can support improved safety management on construction sites.

2.3. Deep learning based object detection

Automatic NHU detection is a specific application of object detection in construction. The goal of visual object detection is to determine if a given image contains one or more objects belonging to the class of interest [35]. Earlier object detection originating from the HOG (Histograms of oriented gradients) [36] was proposed by Dalal et al. in 2005 for the purpose of pedestrian detection. Since then, improved methods based on HOG have been developed, such as the DPM (deformable part-based model) [37] and larger HOG filters [38]. However, these approaches have been traditionally challenged [39] because of the difficulties in training and the use of specially designed learning procedures.

In 2012, Krizhevsky et al. trained a large and deep CNN (convolutional neural network) [40] for image classification. This was the beginning of deep learning. Deep learning [41] allowed the development of multiple processing layers to extract the main features of raw data. Since that time, advances in many domains of science have occurred, especially in object detection.

It has been shown that the deep CNN method performs well in object detection, and a great deal of research has been invested in its improvement. The R-CNN framework is proposed by Girshick et al. [42], who introduce a selective search to obtain 2000 bounding boxes from the image, and then use the CNN method to extract the features of each bounding box. Further, He et al. add an SPP (spatial pyramid pooling) layer between the last convolution layer and its full connectivity layer [43] to avoid graphics distortion, caused by warp or crop, to a proposed region. Girshick [44] propose Fast R-CNN where features of bounding boxes are obtained from the feature map of whole image, and then an ROI (region of interest, which is only one layer of SPP) pooling layer is inserted so that only one feature extraction is needed. Fast R-CNN greatly improves the processing speed of R-CNN, but still cannot be used in real-time applications. In response, Ren et al. propose their improved Faster R-CNN [45], introducing Region Proposal Networks (RPN) instead of selective search to obtain the bounding boxes. The RPN is distinguished from selective search by using convolutional neural networks to generate proposal regions directly through a sliding window. The speed of Faster R-CNN reaches the threshold of real time. Compared with Faster R-CNN, Redmon et al. propose a faster but lower precision algorithm - YOLO [46]. YOLO combines object location with the object recognition process, resulting in a faster algorithm, but which may also miss many small objects at the same time. Liu et al. then present SSD [47], which greatly improves small object detection. SSD uses anchors for classification and BBox regression on various feature maps. Since each layer of a feature map has different receptive fields, multi-scale sampling can be realized.

At present, Faster R-CNN, YOLO and SSD are the most widely used methods of all of these algorithms, which perform well in face detection [48], pedestrian detection [49,50], vehicle detection [51,52], driver hand localization [53], cancer cell detection [54] and in other domains. In comparison, although Faster R-CNN is not the fastest, its mean precision is the highest [55]. The 5fpts speed of Faster R-CNN can fulfill the requirements of a construction site. Compared with speed, high recognition precision and recall rate are more important for NHU detection. Therefore, in this paper, Faster R-CNN is proposed for the detection of construction NHU worker.

3. Methodology

Construction sites are highly complex work environments. The variety of weather events, changes in illumination, visual range, site occlusions and individuals' posture can have a significant impact on worker detection in far-field surveillance videos. There remain challenges in the detection and recognition of workers in live streaming videos. Thus, existing vision-based detection methods are limited in their practical application in real scenarios. In response to these limitations, the overall objective of this paper is to develop a new method for monitoring site workers and evaluating whether the proposed method could be used to detect NHU in various construction site environments. The methodology is introduced in this section.

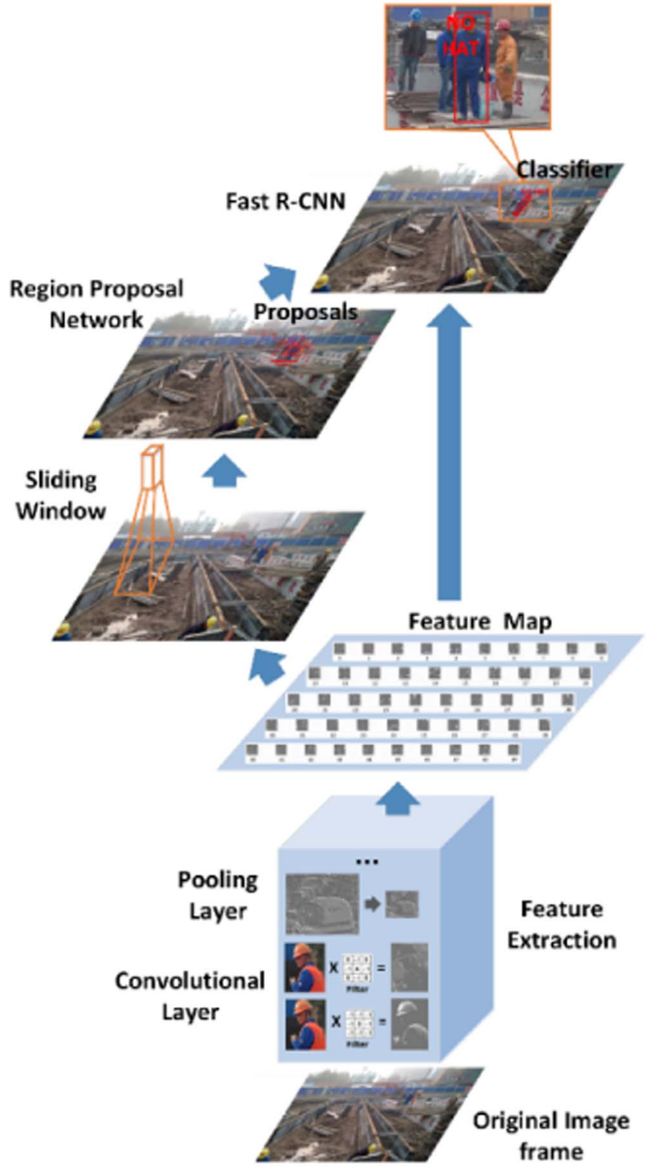


Fig. 1. Framework of the proposed method.

3.1. Faster R-CNN

Faster R-CNN is an object detection method proposed by Ren, et al. [45] in 2015. Faster R-CNN introduces a Region Proposal Network (RPN) that can generate high-quality region proposals and was used to detect and classify objects based on the region proposals and sharing full-image convolutional features with RPN. Faster R-CNN has a frame rate of 5fps (including all steps) on a GPU, and thus is a practical object detection system in terms of both speed and precision.

The Faster R-CNN method involves three steps as illustrated in Fig. 1. Firstly, the features of a photograph are extracted. The network then applies the CNN method to process the whole image with several convolutional layers and max pooling layers to produce a convolutional feature map. The second module is a deep, fully convolutional, network that uses the features to propose regions. Since the whole picture contains many unnecessary objects and the individuals always appear very small in a whole image, it is difficult to judge whether the human is wearing a hardhat simply from the feature maps. Therefore, we need to distinguish foreground regions, which may include humans from other background regions and then abandon the background regions. Only foreground regions are used for recognizing NHU. The third module is the Fast R-CNN detector that uses the proposed regions and corresponding extracted features to classify whether the proposal region is a worker wearing a hardhat or not.

Compared to other methods used to detect NHU in previous studies, Faster R-CNN has three advantages. Firstly, Faster R-CNN is robust in dealing with complex construction site environments. For example, the previous methods only identify specific postures, especially standing. However, Faster R-CNN can automatically learn features without manually establishing various human posture models. Therefore, this method offers robust human detection despite varying site worker postures, weather, illumination, visual range and occlusions. Secondly, the high precision of Faster R-CNN can fulfill the needs of practical engineering applications. Compared to the 10.2% precision of HOG tested in the PASCAL VOC 2006 [56], the precision of Faster R-CNN tested in the Pascal VOC 2012 person dataset is 89.6% [55]. Thirdly, coupled with the short processing time of Faster R-CNN, real-time monitoring of NHU can be achieved.

Therefore, previous methods are feasible for monitoring in near-field images, but have many limitations when applied in far-field surveillance videos. In contrast, Faster R-CNN can handle various situations with greater ease due to its higher precision and shorter calculation time, fulfilling the practical safety monitoring requirements on a variety of construction jobsites.

3.2. Development of the construction worker image dataset

An appropriate dataset was developed to train the Faster R-CNN to detect NHU behavior since there is no off-the-shelf dataset available. In order to develop a diverse and rich dataset of construction worker images, different construction sites were visited and thousands of construction site images were collected. The worker-of-interest (WOI) in each image was annotated to generate the ground truth for training.

3.2.1. Data collection

Since many video cameras are placed on construction sites, we were able to directly clip and save the key frames of the videos. The collection of key frames followed two requirements: 1) the videos covered various construction sites conditions; and 2) sufficient number of samples is necessary. In order to meet these requirements, the cameras covered the entire construction site.

3.2.2. Image annotation

After collecting the image frames of construction workers, the next step was to annotate them using the graphical image annotation tool *LabelImg* [57]. The annotations included the identification of NHU workers. The annotations were saved as XML files in PASCAL VOC format as can be used by Python.

3.3. Metrics for performance evaluation

The performance of the method was measured based on its correctness, speed and robustness as follows.

3.3.1. Correctness

The correctness metrics were selected based on the main purpose of the method being to detect individual NHU workers. The final classifier was divided into two categories: NHU workers and the rest. The first metric is precision, which is popular in evaluating pattern recognition. To clarify the meaning of precision, we have first to define the meaning of TP (true positive), FP (false positive), and FN (false negative). Specifically, TP is the number of NHU workers and where the test results are correct. FP is the number of NHU objects detected, but results are incorrect. For example, if the worker is wearing a hardhat but the model recognizes the worker as NHU, or even other objects are mistaken as a NHU worker, we count these situations as FP. FN is the number of NHU workers, but where the test results are incorrect. Table 1 presents the definitions of TP, FP and FN and several site examples are presented in Fig. 2.

Table 1
Definitions TP, FP and FN for NHU detection.

Category	Actual NHU	Predicted NHU
TP	Yes	Yes
FP	No	Yes
FN	Yes	No

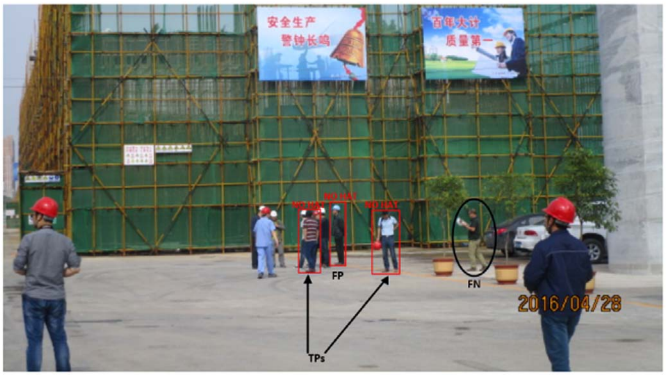


Fig. 2. Examples of TP, FP and FN in NHU detection.

Precision is defined as a ratio of TP to TP + FP and measures the reliability of the detection. TP + FP is the number of workers detected as NHU based on the method. Recall is the ratio of TP to TP + FN. TP + FN means the actual number of NHU workers. Miss rate is the opposite to recall and indicates how many NHU workers are missed by the method.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Miss rate} = 1 - \text{Recall} = \frac{\text{FN}}{\text{TP} + \text{FN}} \tag{3}$$

3.3.2. Speed

The speed of Faster R-CNN refers to the time consumed by complete NHU worker detection for one image. The calculation speed is much faster than previous methods, since this method is calculated on the GPU. As the objective here is to apply Faster R-CNN to real construction jobsite surveillance videos, is necessary to ascertain whether Faster R-CNN can fulfill the real-time requirement.

3.3.3. Robustness

Robustness represents the degree of tolerance of an object detection method when applied to testing various images. Construction sites are usually in open outdoor environments and contain a large amount of workers, equipment and material. Therefore, changes in weather, illumination, individual postures, visual range and occlusions frequently occur on construction sites. These factors inevitably have a significant impact on the visual verisimilitude on such work sites. A good algorithm should be robust to such changes and not degrade significantly under varying conditions. Correctness and speed in different situations are indicators reflecting the robustness of the model.

4. Experiments and results

4.1. Experimental data collection and setup

To establish the construction worker image dataset, we collected more than 100,000 image frames of surveillance videos from 25 different construction projects. In order to create a comprehensive dataset (of assorted situations), the videos were collected for more than one year. A total of 81,000 images from this dataset were randomly selected to comprise the training dataset. These were annotated and used to develop a Faster R-CNN-based NHU model. The rest of the images constituted the testing dataset.

The performance of the method was evaluated using the correctness, speed and robustness metrics. To measure these aspects, we tested correctness and speed in different situations. Therefore, all the images in the testing dataset were classified into several categories based on weather, illumination, individuals' posture, visual range and occlusions. The information in each category is listed in Table 2. The next step was to evaluate correctness and speed in all of these situations.

4.2. Results

A wide range of collected images were tested in the Faster R-CNN model. The correctness performance under different situations follows.

4.2.1. Choice of confidence threshold

The Faster R-CNN model provides a confidence value for every detected object. The confidence value here is defined as the probability of an object being a NHU worker. For example, a 0.9 confidence value means that the probability of an object being a NHU worker is 90%. Positive samples are identified when the confidence values are above the confidence threshold. Therefore, the value of the confidence threshold has an influence on the classification of positive and negative samples. Fig. 3 illustrates the precision-recall (P/R) curve based on different confidence thresholds. As is shown, a high confidence threshold is inclined to reject ambiguous samples and results in high precision but low recall, while a low confidence threshold situation accepts more ambiguous samples but with high recall and low precision. To achieve a ‘win-win’ result of both high precision and recall, we chose 0.7 as the confidence threshold for the study.

Table 2
Information of image samples in different situations.

Categories	No.	Value	Total number of NHU workers	Number of images
Weather	1	Sunny	2582	1000
	2	Cloudy	2249	1000
	3	Rainy	1684	1000
	4	Haze	2350	1000
Illumination	1	8:00–10:00 am	2120	1000
	2	10:00–12:00 am	2437	1000
	3	2:00–4:00 pm	2646	1000
	4	4:00–6:00 pm	2035	1000
Individual posture	1	Standing	1660	1000
	2	Bending	1335	1000
	3	Squatting	1090	1000
	4	Sitting	1004	1000
Visual range	1	Small	3654	1000
	2	Middle	2167	1000
	3	Large	1136	1000
Occlusions	1	Whole body visible	1200	1000
	2	Upper body visible	1101	1000
	3	Head visible	1046	1000
	4	Only part of head visible	1120	1000

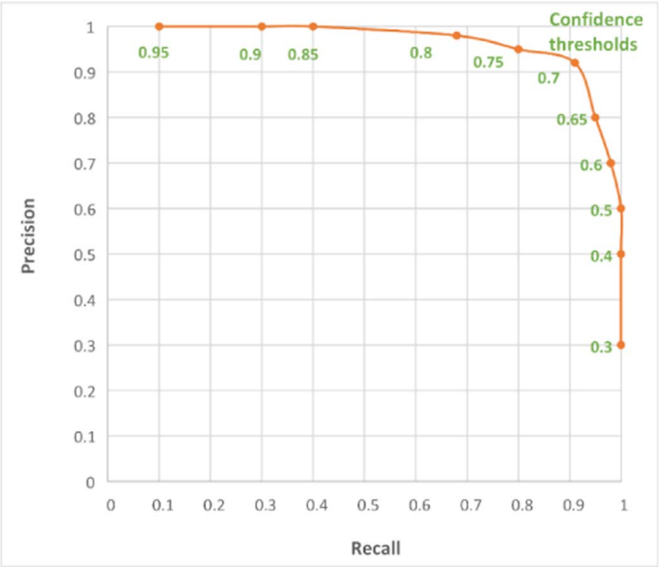


Fig. 3. Precision-recall (P/R) curve.

4.2.2. Impact of visual range

As surveillance cameras are placed in different locations on construction jobsites and the trajectory of workers is stochastic, workers were captured in different resolutions in the surveillance videos. If workers are close to the camera, they are captured in a larger pixel size and have richer image features. Conversely, workers far away from the cameras are captured in a smaller pixel size and have fuzzy image features. The image frames were manually divided into three categories to evaluate the robustness of the trained model on the different pixel sizes of workers. As shown in Fig. 4, workers were captured in a small pixel size in the large visual range category, with a large pixel size in small visual range.

The test dataset includes 1000 images for each category and the test results are shown in Table 3. As the distance between camera and workers increases, the pixel size of a person in the image gradually becomes smaller. Although the precision and recall rate of image detection gradually decreased, the overall precision and recall rate remained greater than 90%. Consequently, the trained model proved to be robust in detecting workers in different pixel sizes and with satisfactory results overall.

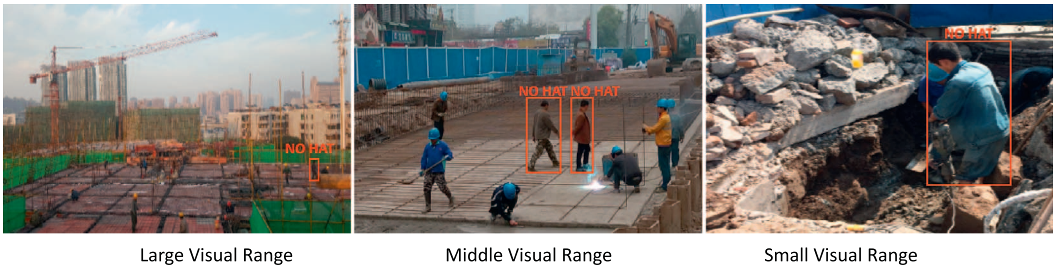
4.2.3. Impact of weather

Construction sites are predominantly exposed to the outdoor environment and thus can be significantly affected by natural conditions. As such, changes in the weather have an impact on the quality of the surveillance video. Heavy rain and severe haze are excluded as they often lead to work being suspended. As represented in Fig. 5, four common weather types are included: sunny, cloudy, misty rain and hazy.

The test results (Table 4) of the robustness of the model in different weather conditions indicate that they have little effect on detection performance, with the precision and recall rate remaining robust in misty rain and hazy conditions although the results were better on sunny and cloudy days.

4.2.4. Impact of illumination

The working time on the sample construction sites was 8:00 am to

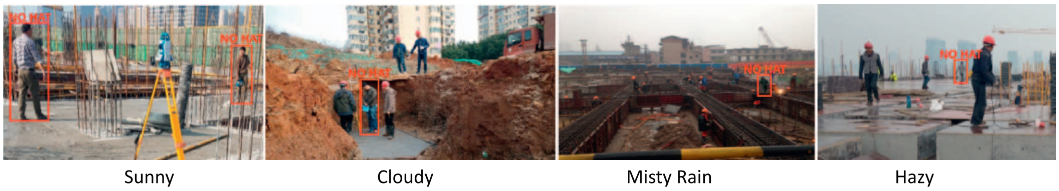


Large Visual Range Middle Visual Range Small Visual Range

Fig. 4. Image frame examples with different visual ranges.

Table 3
Precision, recall and miss rate ratios under different visual range.

Categories	No.	Value	TP	FP	FN	Precision (%)	Recall (%)	Miss rate (%)	Speed (s)
Visual range	1	Large	3374	226	280	93.7	92.3	7.7	0.212
	2	Middle	2065	91	102	95.8	95.3	4.7	0.207
	3	Small	1089	18	47	98.4	95.9	4.1	0.204

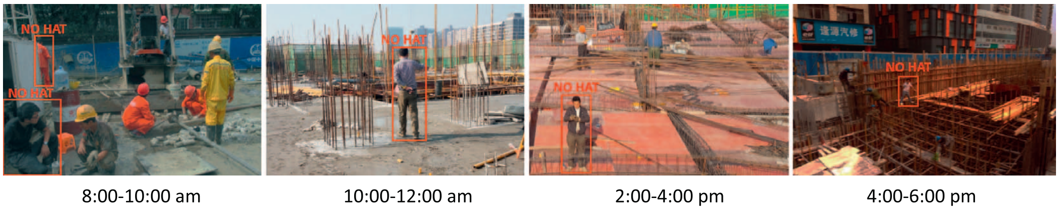


Sunny Cloudy Misty Rain Hazy

Fig. 5. Image frame examples under different weather conditions.

Table 4
Precision, recall and miss rate ratios in different weather conditions.

Categories	No.	Value	TP	FP	FN	Precision (%)	Recall (%)	Miss rate (%)	Speed (s)
Weather	1	Sunny	2459	83	123	96.7	95.2	4.8	0.204
	2	Cloudy	2155	98	94	95.7	95.8	4.2	0.202
	3	Misty rain	1586	107	98	93.7	94.2	5.8	0.209
	4	Hazy	2186	123	164	94.7	93.0	7.0	0.210



8:00-10:00 am 10:00-12:00 am 2:00-4:00 pm 4:00-6:00 pm

Fig. 6. Image frame examples under different illumination levels.

Table 5
Precision, recall and miss rate ratios under different illumination levels.

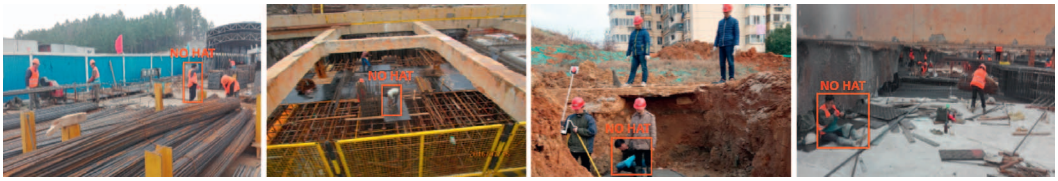
Categories	No.	Value	TP	FP	FN	Precision (%)	Recall (%)	Miss rate (%)	Speed (s)
Illumination	1	8:00–10:00 am	2005	92	115	95.6	94.6	5.4	0.209
	2	10:00–12:00 am	2334	82	103	96.6	95.8	4.2	0.207
	3	2:00–4:00 pm	2528	78	118	97.0	95.5	4.5	0.208
	4	4:00–6:00 pm	1907	62	128	96.9	93.7	6.3	0.210

12:00 am and 2:00 pm to 6:00 pm. Fig. 6 illustrates that the illumination level was mild in the morning (8:00 am to 10:00 am), strongest during the day with a peak at midday (10:00 am to 12:00 am and 2:00 pm to 4:00 pm) and progressively weaker from 4:00 pm to 6:00 pm. We divided the image frames into these four categories to test the impact of illumination on the precision, recall and miss rate.

The test results show that the illumination did not affect the video detection. As shown in Table 5, the precision and recall rate declined only slightly as illumination reduced. The miss rate was acceptable in each group.

4.2.5. Impact of individual posture

Different types of work, mechanical tools used and activity locations determine a variety of postures of workers in construction. Pedestrian

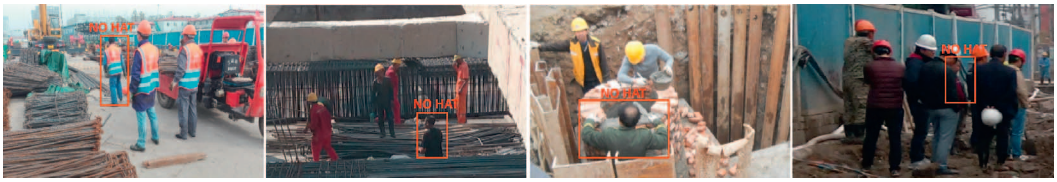


Standing Bending Squatting Sitting

Fig. 7. Image frame examples including different individual postures.

Table 6
Precision, recall and miss rate ratios under different individual postures.

Categories	No.	Value	TP	FP	FN	Precision (%)	Recall (%)	Miss rate (%)	Speed (s)
Individual's posture	1	Standing	1608	54	52	96.8	96.9	3.1	0.209
	2	Bending	1255	58	80	95.6	94.0	6.0	0.208
	3	Squatting	1019	68	71	93.7	93.5	6.5	0.205
	4	Sitting	988	56	16	94.6	98.4	1.6	0.207



Whole body visible Upper body visible Head visible Only part of head visible

Fig. 8. Image frame examples including different occlusions.

Table 7
Precision, recall and miss rate ratios under different occlusions.

Categories	No.	Value	TP	FP	FN	Precision (%)	Recall (%)	Miss rate (%)	Speed (s)
Occlusions	1	Whole body visible	1143	54	57	95.5	95.3	4.8	0.205
	2	Upper body visible	1077	43	24	96.2	97.8	2.2	0.206
	3	Head visible	996	41	50	96.0	95.2	4.8	0.204
	4	Only part of head visible	686	75	434	90.1	61.3	38.8	0.209

Table 8
Comparison of the applicability of various methods.

Methodology	Independent of reading range/figure size	Independent of facial features	Independent of worker posture	Independent of partial occlusion	Independent of workers' coordination and participation
RFID [19,24]	×	√	√	√	×
Bluetooth + pressure sensor [25]	×	√	√	√	×
Haar-like features [31]	×	×	√	√	√
Edge detection algorithms [32]	×	×	√	√	√
HOG + CHT [33]	√	×	×	×	√
HOG + SVM [8,34]	√	√	×	×	√
Faster R-CNN	√	√	√	√	√

detection generally is only concerned with a standing posture, whereas four common construction worker postures comprise standing, bending, squatting and sitting (Fig. 7).

The dataset in this test contained 1000 image frames for each category. There were multiple workers in each image frame, so NHU workers in each image frame had a variety of postures. For the convenience of metrics, we chose image frames in which NHU workers all adopted the same posture. The test result (Table 6) shows a high overall precision, representing the excellent performance of the model in testing various worker postures. The recall rates of bending and squatting are slightly lower than the others. However, the overall recall rate is above 91%.

4.2.6. *Impact of occlusions*

Construction sites are generally occupied by a large amount of workers, equipment and materials. In far-field surveillance videos, workers are often obstructed by equipment, materials and other workers. Therefore, many workers in the videos are incomplete (example shown in Fig. 8). To test the impact of occlusions, we classify the occlusion degree into four categories: 'whole body visible', 'upper body visible', 'head visible' and 'only part of the head visible' as represented in Table 7.

Similar to the posture test, every image frame contained one kind of occlusion category of NHU workers. The test results show that the precision and recall rate of 'whole body visible', 'upper body visible' and 'head visible' are all beyond 95%. The precision is still robust for 'only part of head visible', but the recall rate is only 64.5% since it is difficult to detect a NHU worker without their full head visible.

5. Discussion

The use of hardhats can protect workers from head injuries caused when struck by objects, punctures, and extrusions. Research conducted by the United States Bureau of Labor Statistics identified that 84% of workers who had suffered head injuries were not wearing head protection [8]. This suggests that strengthening the supervision of workers by proactively detecting and alerting to NHU can reduce the threat of head injuries. Therefore, the proposed automatic detection method warns NHU workers for immediate response. The objective of this study was to develop a new method for the detection of NHU workers in various construction site environments. This is a comprehensive study for practical use on construction sites. The proposed method can achieve real-time monitoring with high precision and recall in different scenarios and thus, can efficiently provide an early warning to NHU workers. Additionally, the active monitoring of NHU can also contribute as a leading indicator of overall site safety performance. It is argued by tracking active lead indicators provides a more accurate assessment of construction site safety performance [58].

In this paper, the various methods for NHU detection and the development of object detection technologies are discussed. Previous studies have adopted specific methods to solve this problem with limitations in adaptation and practical feasibility to construction site conditions. We reviewed the limitations of each of these methods (represented in Table 8), and discussed the development of vision based methods in the history of computer vision. Existing sensor-based detection methods, including RFID based methods, are limited by the need for a physical tag or sensor to be worn by each construction worker. Additionally, sensor transmitters via Bluetooth require regular charging that can impact on their practicality on a construction site. In general, the large investment in numerous devices across a complete network, and the heavy dependence on workers' coordination and participation, can constrain the applicability of sensor-based methods in PPE monitoring. Further, considering the powerful performance of deep learning and the limitations of HOG in practical use (for example, those not facing the cameras cannot be recognized), deep learning offers a significant improvement in automatic NHU detection.

In selecting the most suitable method for construction site applications, we discussed the image characteristics of construction sites and analyzed the range of factors affecting NHU detection. Faster R-CNN was selected for its robust performance in PASCAL VOC Challenge. To address previous method limitations, we tested the performance of Faster R-CNN on a variety of construction site images. 19,000 images were collected as the test dataset and the precision and recall rates were verified by manual calibration. The test dataset covers a variety of visual conditions that may occur on construction sites including visual range, weather, illumination, individual posture and occlusions. The results demonstrate the robustness of Faster R-CNN in various visual conditions of the construction site. The recognition precision and recall rates were consistently more than 90% except for the low precision for 'only part of head visible' under 'the impact of occlusions' - an expected result of computer vision, because even the best algorithm cannot accurately detect a hardhat on an obstructed head. However, considering the real-time processing speed of Faster R-CNN and frequent posture changes of workers, those who are not detected for 'only part of head visible' may be detected in their next movement as their heads become more revealed.

A new method is needed to adapt to the far-field nature of surveillance videos on construction sites and this study offers a new and robust deep learning method in detecting NHU workers this way. Previous studies have been constrained by the requirement for a standing and camera-facing posture, which is not consistent with construction jobsite conditions. A large number of image frames are investigated from far-field surveillance videos on real construction sites, which are characterized by pixel-sized images of workers, broad backgrounds and various worker postures. Faster R-CNN has been proven feasible in NHU detection from far-field images by experiments.

6. Conclusion

Construction continues to be one of the most dangerous job sectors globally. Despite hardhats offering significant protection in resisting penetration by objects and absorbing shock from direct blows to the head, they do not always prevent on-site accidents resulting in head injury. A primary reason for this is site-worker ignorance. To effectively manage on-site safety, it is critical to improve the detection of NHU workers. Since previous NHU detection methods cannot be adapted to various changes in the visual field, they are still far from practical use.

This paper proposes a new method for NHU detection in far-field surveillance videos based on Faster R-CNN. Firstly, a large number of image frames were collected as a training dataset to develop a Faster R-CNN model for NHU detection. Secondly, the image frames in the testing dataset were divided into 19 categories according to the real visual characteristics of construction sites. The performance of the proposed method was then evaluated under various site conditions. The test results indicate that the method could successfully detect NHU construction workers with a precision and recall rate of 95.7% and 94.9% respectively under a variety of conditions. The high precision and recall rates indicate that the proposed method can be effectively used to detect NHU workers in far-field surveillance videos. The method offers a significant opportunity to contribute to real-time site monitoring and improve the safety management of workers on construction sites. Currently, our algorithm is able to detect NHU workers but not identity the workers involved. Therefore, it is recommended that future research focus on the identification and integration of worker information into real-time safety monitoring systems as this will then enable disciplinary action and targeted safety training to be carried out.

Acknowledgement

The authors would like to acknowledge Lei Zhang, Xianbiao Qi, Chengqian Li, Tianshui Chen and Yachun Huang for their help. We are also thankful for the financial support of 1) National Science Foundation of China (grant no.51678265). 2) The Research Grants Council of Hong Kong grant titled "Proactively Monitoring Construction Progress by Integrating 3D Laser-scanning and BIM" (PolyU 152093/14E). 3) National 12th Five-Year Plan Major Scientific and Technological Issues (NFYPMSTI) through Grant 2015BAK33B04.

References

[1] S. Schneider, P. Susi, Ergonomics and construction: a review of potential hazards in new construction, Am. Ind. Hyg. Assoc. J. 55 (7) (1994) 635–649, http://dx.doi.org/10.1080/15428119491018727.

[2] Bereau of Labor Statistics, Construction: NAICS 23,2017, https://www.bls.gov/iag/tgs/iag23.htm.

[3] Health and Safety Executive, Statistics on fatal injuries in the workplace in Great Britain 2016, 2016, http://www.hse.gov.uk/statistics/.

[4] B.Y. Jeong, Occupational deaths and injuries in the construction industry, Appl. Ergon. 29 (5) (1998) 355–360, http://dx.doi.org/10.1016/S0003-6870(97)00077-X.

[5] S. Konda, H.M. Tiesman, A.A. Reichard, Fatal traumatic brain injuries in the construction industry, 2003–2010, Am. J. Ind. Med. 59 (3) (2016) 212–220, http://dx.doi.org/10.1002/ajim.22557.

[6] A. Colantonio, D. McVittie, J. Lewko, J. Yin, Traumatic brain injuries in the construction industry, Brain Inj. 23 (11) (2009) 873–878, http://dx.doi.org/10.1080/02699050903036033.

[7] Occupational Safety & Health Administration, Determining the need for hard hat and eye protection on construction sites, 2004, https://www.osha.gov/pls/oshaweb/owasrch.search_form?p_doc_type=INTERPRETATIONS&p_toc_level=3&p_keyvalue=1926.100&p_status=CURRENT.

[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] B. Naticchia, M. Vaccarini, A. Carbonari, A monitoring system for real-time interference control on large construction sites, Autom. Constr. 29 (2013) 148–160, http://dx.doi.org/10.1016/j.autcon.2012.09.016.

[10] M. Paul, S.M. Haque, S. Chakraborty, Human detection in surveillance videos and its applications-a review, EURASIP J. Adv. Signal Process. 2013 (1) (2013) 176, http://dx.doi.org/10.1186/1687-6180-2013-176.

[11] Y. Tian, R.S. Feris, H. Liu, A. Hampapur, M.-T. Sun, Robust detection of abandoned and removed objects in complex surveillance videos, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 41 (5) (2011) 565–576, http://dx.doi.org/10.1109/TSMCC.2010.2065803.

[12] C.C. Chen, J.K. Aggarwal, Recognizing human action from a far field of view, 2009 Workshop on Motion and Video Computing (WMVC), 2009, pp. 1–7, , http://dx.doi.org/10.1109/WMVC.2009.5399231.

[13] H. Li, X. Li, X. Luo, J. Siebert, Investigation of the causality patterns of non-helmet use behavior of construction workers, Autom. Constr. (2017), http://dx.doi.org/10.1016/j.autcon.2017.02.006.

[14] A. Hume, N. Mills, A. Gilchrist, Industrial Head Injuries and the Performance of the Helmets, Proceedings of the International IRCOBI Conference on Biomechanics of Impact, Switzerland, (1995).

[15] B.L. Suderman, R.W. Hoover, R.P. Ching, I.S. Scher, The effect of hardhats on head and neck response to vertical impacts from large construction objects, Accid. Anal. Prev. 73 (2014) 116–124, http://dx.doi.org/10.1016/j.aap.2014.08.011.

[16] R.R. Cabahug, A survey on the implementation of safety standards of on-going construction projects in Cagayan de Oro City, Philippines, Mindanao J. Sci. Technol. 12 (1) (2014) 12–24.

[17] I.W. Fung, Y. Lee, V.W. Tam, H. Fung, A feasibility study of introducing chin straps of safety helmets as a statutory requirement in Hong Kong construction industry, Saf. Sci. 65 (2014) 70–78, http://dx.doi.org/10.1016/j.ssci.2013.12.014.

[18] 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).

[19] A. Kelm, L. Laußat, A. Meins-Becker, D. Platz, M.J. Khazaee, A.M. Costin, M. Helmus, J. Teizer, Mobile passive radio frequency identification (RFID) portal for automated and rapid control of personal protective equipment (PPE) on construction sites, Autom. Constr. 36 (2013) 38–52.

[20] F. Akbar-Khanzadeh, Factors contributing to discomfort or dissatisfaction as a result of wearing personal protective equipment, J. Hum. Ergol. 27 (1–2) (1998) 70–75, http://dx.doi.org/10.11183/jhe1972.27.70.

[21] N. Cavazza, A. Serpe, Effects of safety climate on safety norm violations: exploring the mediating role of attitudinal ambivalence toward personal protective equipment, J. Saf. Res. 40 (4) (2009) 277–283, http://dx.doi.org/10.1016/j.jsr.2009.06.002.

[22] D.A. Lombardi, S.K. Verma, M.J. Brennan, M.J. Perry, Factors influencing worker use of personal protective eyewear, Accid. Anal. Prev. 41 (4) (2009) 755–762, http://dx.doi.org/10.1016/j.aap.2009.03.017.

[23] R. Flin, K. Mearns, P. O'Connor, R. Bryden, Measuring safety climate: identifying the common features, Saf. Sci. 34 (1) (2000) 177–192, http://dx.doi.org/10.1016/S0925-7535(00)00012-6.

[24] 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.

[25] S. Dong, Q. He, H. Li, Q. Yin, Automated PPE Misuse Identification and Assessment for Safety Performance Enhancement, ICCREM 2015, (2015), pp. 204–214, http://dx.doi.org/10.1061/9780784479377.024.

[26] 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.

[27] 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) (2013) A4014005, , http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000339.

[28] 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.

[29] M. Liu, D. Hong, S. Han, S. Lee, Silhouette-Based On-Site Human Action Recognition in Single-View Video, Construction Research Congress, (2016), pp. 951–959, http://dx.doi.org/10.1061/9780784479827.096.

[30] R. Starbuck, J. Seo, S. Han, S. Lee, A stereo vision-based approach to marker-less motion capture for on-site kinematic modeling of construction worker tasks, Comput. Civ. Build. Eng. 2014 (2014) 1094–1101, http://dx.doi.org/10.1061/9780784413616.136.

[31] S. Du, M. Shehata, W. Badawy, Hard hat detection in video sequences based on face features, motion and color information, 2011 3rd International Conference on Computer Research and Development, vol. 4, IEEE, 2011, pp. 25–29, , http://dx.doi.org/10.1109/ICCRD.2011.5763846.

[32] 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.

[33] A.H. Rubaiyat, T.T. Toma, M. Kalantari-Khandani, S.A. Rahman, L. Chen, Y. Ye, C.S. Pan, Automatic Detection of Helmet Uses for Construction Safety, 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), IEEE, (2016), pp. 135–142, http://dx.doi.org/10.1109/WIW.2016.045.

[34] Z. Zhu, M.-W. Park, N. Elsafty, Automated monitoring of hardhats wearing for onsite safety enhancement, 11th Construction Specialty Conference, 2015, http://dx.doi.org/10.14288/1.0076342.

[35] S.S. Bucak, R. Jin, A.K. Jain, Multiple kernel learning for visual object recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 1354–1369, http://dx.doi.org/10.1109/TPAMI.2013.212.

[36] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, IEEE, 2005, pp. 886–893, , http://dx.doi.org/10.1109/CVPR.2005.177.

[37] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1627–1645.

[38] L. Zhu, Y. Chen, A. Yuille, W. Freeman, Latent hierarchical structural learning for object detection, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 1062–1069, , http://dx.doi.org/10.1109/CVPR.2010.5540096.

[39] C. Szegedy, A. Toshev, D. Erhan, Deep neural networks for object detection, Adv. Neural Inf. Proces. Syst. (2013) 2553–2561.

[40] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Proces. Syst. (2012) 1097–1105, http://dx.doi.org/10.1145/3065386.

[41] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, http://dx.doi.org/10.1038/nature14539.

[42] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587, , http://dx.doi.org/10.1109/CVPR.2014.81.

[43] K. He, X. Zhang, S. Ren, J. Sun, Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, European Conference on Computer Vision, Springer, 2014, pp. 346–361, http://dx.doi.org/10.1007/978-3-319-10578-9_23.

[44] R. Girshick, Fast r-cnn, Proceedings of the IEEE International Conference on Computer Vision, (2015), pp. 1440–1448, http://dx.doi.org/10.1109/ICCV.2015.169.

[45] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, Adv. Neural Inf. Proces. Syst. (2015) 91–99, http://dx.doi.org/10.1109/TPAMI.2016.2577031.

[46] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788, , http://dx.doi.org/10.1109/CVPR.2016.91.

[47] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: Single Shot Multibox Detector, European Conference on Computer Vision, Springer, 2016, pp. 21–37, http://dx.doi.org/10.1007/978-3-319-46448-0_2.

[48] H. Jiang, E. Learned-Miller, Face Detection with the Faster R-CNN, (2016), http://dx.doi.org/10.1109/FG.2017.82 (arXiv preprint arXiv:1606.03473).

[49] 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.

[50] Q. Peng, W. Luo, G. Hong, M. Feng, Y. Xia, L. Yu, X. Hao, X. Wang, M. Li, Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO, 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) vol. 2, IEEE, 2016, pp. 562–565, http://dx.doi.org/10.1109/IHMSC.2016.130.

[51] H. Kim, Y. Lee, B. Yim, E. Park, H. Kim, On-road Object Detection Using Deep Neural Network, Consumer Electronics-Asia (ICCE-Asia), IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), IEEE, 2016, pp. 1–4, http://dx.doi.org/10.1109/ICCE-Asia.2016.7804765.

[52] Y. Zhou, H. Nejati, T.-T. Do, N.-M. Cheung, L. Cheah, Image-based Vehicle Analysis using Deep Neural Network: A Systematic Study, 2016 IEEE International Conference on Digital Signal Processing (DSP), (2016), http://dx.doi.org/10.1109/ICDSP.2016.7868561.

[53] A. Rangesh, E. Ohn-Bar, M.M. Trivedi, Driver hand localization and grasp analysis: A vision-based real-time approach, Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016, pp. 2545–2550, http://dx.doi.org/10.1109/ITSC.2016.7795965.

[54] J. Zhang, H. Hu, S. Chen, Y. Huang, Q. Guan, Cancer Cells Detection in Phase-Contrast Microscopy Images Based on Faster R-CNN, Computational Intelligence and Design (ISCID), 2016 9th International Symposium on Computational Intelligence and Design (ISCID), vol. 1, IEEE, 2016, pp. 363–367, http://dx.doi.org/10.1109/ISCID.2016.1090.

[55] PASCAL VOC, Detection results: VOC2012, 2017, http://host.robots.ox.ac.uk:8080/leaderboard/displaylb.php?cls = mean&challengeid = 11&compid = 4&submid = 9222.

[56] P. Ott, M. Everingham, Implicit Color Segmentation Features for Pedestrian and Object Detection, Computer Vision, 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 723–730, http://dx.doi.org/10.1109/ICCV.2009.5459238.

[57] GitHub, LabelImg: a graphical image annotation tool, 2005, https://github.com/tzutalin/labelImg.

[58] J. Hinze, S. Thurman, A. Wehle, Leading indicators of construction safety performance, Saf. Sci. 51 (1) (2013) 23–28, http://dx.doi.org/10.1016/j.ssci.2012.05.016.