Applications of Deep Learning in Intelligent Construction

Yang Zhang and Ka-Veng Yuen

Abstract Smart construction site is the concrete embodiment of the concept of smart city in the construction industry. It provides all-round, three-dimensional, and real-time supervision of construction sites with intelligent systems of controllability, data, and visualization. In particular, it is common to install many cameras in smart construction sites. These cameras only play the role of visualization, and further analysis is necessary to extract information from the images/videos and to provide safety warning signals. With the development of deep learning in the field of image processing, automatic deep feature extraction of images is possible for construction safety monitoring. This chapter summarizes the development and application of deep learning in construction safety, such as bolt loosening damage, structural displacement, and worker behavior. Finally, the application scenarios of deep learning in smart construction sites are further discussed.

Keywords Smart construction site · Deep learning · Construction safety · Computer vision

1 Introduction

Construction safety is an important issue for the society. There are many hidden dangers in construction sites, such as various types of workers, large construction machinery and complex environment. In the process of dynamic construction, the load of building structure is time-varying, and various support structures interact among themselves, leading to the stability risk of the entire construction. Buildings

Y. Zhang · K.-V. Yuen (✉)
State Key Laboratory of Internet of Things for Smart City, Department of Civil and Environmental Engineering, University of Macau, Macau 999078, China
e-mail: kvyuen@um.edu.mo

Guangdong-Hong Kong-Macau Joint Laboratory for Smart Cities, University of Macau, Macau 999078, China

Y. Zhang
e-mail: yangzhang@um.edu.mo

A. Cury et al. (eds.), *Structural Health Monitoring Based on Data Science Techniques*, Structural Integrity 21, https://doi.org/10.1007/978-3-030-81716-9_11

and structures under construction are prone to collapse, overturning and other accidents due to the influence of natural conditions, construction level, and construction quality. In addition, workers suffer from injuries caused by construction machinery, such as blow or collision. Therefore, the construction safety monitoring is necessary.

In order to improve the construction level of engineering projects and reduce the incidence of accidents, the concept of smart construction site was proposed [1]. Smart construction site is the concrete embodiment of the concept of smart city in the construction industry. It provides all-round, three-dimensional, and real-time supervision of construction sites with intelligent systems of controllability, data, and visualization. At present, the development of smart construction site is still in the initial stage, i.e., the perception stage. It uses advanced sensing technologies to monitor workers, machineries, and structures, and then identifies and locates the potential hazards. Building information modeling (BIM) takes the three-dimensional graphics of buildings as the carrier to further integrate all types of building information. The parameterized model can be used to realize construction simulation, collision detection, and other applications. Hu et al. used network analysis to improve collision detection. A component network centered on conflicting objects was constructed to represent the dependencies of components. This method can effectively identify the irrelevant conflicts and reduce the number of irrelevant conflicts by 17% [2]. Mirzaei et al. developed a novel 4D-BIM dynamic conflict detection and quantification system for the identification of spatiotemporal conflicts [3]. 3D laser scanning technology uses the principle of laser ranging to scan objects and quickly obtain 3D models. There have been many attempts in building inspection, cultural relics protection and other methods. Yang et al. [4] used terrestrial laser scanning to detect the deformation of the arch structure. The surface approximation method was used to cover the blank of measurement area, and the uncertainty of surface of different order was studied. Valenca et al. [5] proposed an automatic crack assessment method based on image processing and terrestrial laser scanning (TLS) technology. The geometric information measured by TLS was used to correct the captured image. It improved the identification accuracy of structural cracks. The combination of 3D laser scanning technology and BIM can realize the detection of structural quality and deformation. Ham et al. [6] proposed a structural safety diagnosis method based on laser scanning and BIM. The laser scanning data and BIM model were compared and analyzed to determine the deformation degree of pipe support. Chen et al. [7] proposed a point-to-point comparison method for deviation detection between automatic scanning and BIM. When there is a deviation between BIM and point clouds, it will be highlighted to remind users for further investigation.

Sensor networks can detect the deformation of high formwork support, tower crane, and bridge by different sensor nodes. Kifouche et al. [8] developed and deployed a sensor network that could collect data from multiple types of sensors. The data was transferred to a server for visualization and real-time processing. Kuang et al. [9] used fiber-optic sensors to monitor the deflection and cracks of beams. The results showed that it was possible to detect fine crack and final failure crack by optical fiber. Casciati et al. [10] used GPS to detect the displacement of steel structures in real time. The accuracy was of the order of subcentimeters. With the continuous upgrading of camera equipment, camera measurement technology has also been substantially developed. Cameras can be used to detect the deformation and vibration of structures in a close distance. Feng et al. [11] demonstrated the potential of low-cost visual displacement sensors for structural health monitoring. Meanwhile, experimental results showed that vision sensors have high precision in the full-field displacement measurement. Harvey et al. [12] used visual sensors to measure interlayer drift in real time. The dynamic characteristics were extracted to detect structural damage.

Artificial intelligence has accelerated its development, presenting new features such as deep learning, cross-border integration, human–machine collaboration, open intelligence, and autonomous control [13]. It has great potential in the field of construction safety monitoring. Deep convolutional neural networks have achieved remarkable results in the field of image processing. Compared with traditional machine learning methods, deep learning does not need to manually extract features and has a strong ability of autonomous feature extraction. Therefore, deep learning has been widely applied in various research fields [14–16].

With the rapid development of smart construction site, many cameras have been installed on construction sites. These cameras provide a lot of real-time data for construction safety detection. However, it is still an important research topic to deeply understand the image and to provide more accurate and timely monitoring results for construction safety. As one of the non-contact detection techniques, camera measurement has attracted much attention in the field of structural health monitoring. Some researchers have tried to combine deep learning with machine vision for construction safety monitoring. According to the construction characteristics, the detection content can be roughly divided into structure, worker, and mechanical operation safety. Section 2 introduces several kinds of deep learning algorithms commonly used in the field of construction safety monitoring. Section 3 presents the structural safety monitoring using deep learning. Section 4 focuses on worker safety management based on deep learning. Section 5 describes safety management of construction machinery using deep learning in the construction process. Section 6 summarizes the current situation and development of deep learning in the field of construction safety monitoring.

2 Related Deep Learning Algorithms

Convolutional neural network is the cornerstone for deep learning to achieve breakthrough achievements in the field of computer vision in recent years. Convolutional neural network uses convolutional layers to replace fully connected (FC) layers for effective feature extraction. A convolutional neural network usually consists of multiple convolutional modules. The front convolution layer has a small receptive field, which can capture local and detailed information of an image. The receptive field of the latter convolution layer is gradually enlarged to capture more complex and abstract information in the image. At first, deep convolutional neural network

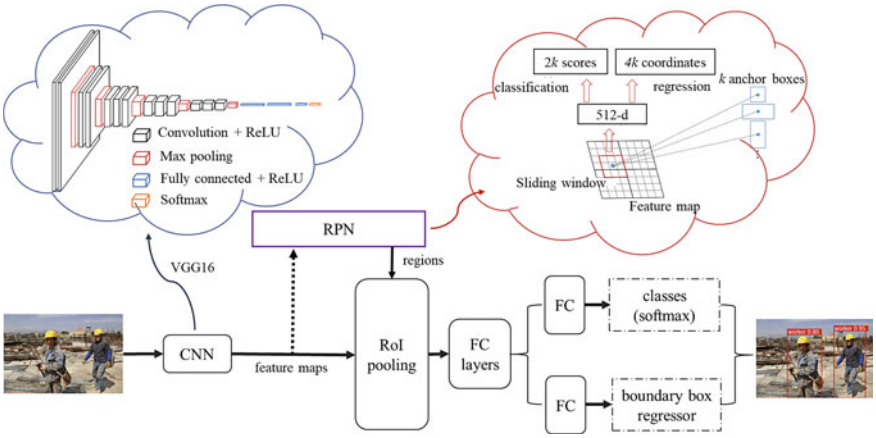


Fig. 1 Faster R-CNN

was used to perform image classification tasks and determine object categories in the image. After AlexNet was concerned by researchers, some image classification networks were produced, such as VGGNet, GoogLeNet, and ResNet. The recognition accuracy of these networks in the image classification task has almost reached the human level. However, image classification is a very crude task. Objects in an image need not only to be classified, but also to be positioned. Some algorithms combine candidate regions with image classification networks to achieve target location based on bounding boxes, such as region-based convolutional neural networks (R-CNN), Fast R-CNN, single-shot multibox detector (SSD), you only look once (YOLO), Faster R-CNN, and so on. The overall architecture of Faster R-CNN is shown in Fig. 1. Image classification network is used to extract features, and region proposal network (RPN) is used to generate candidate regions [17]. It further improves the accuracy and positioning accuracy of object recognition.

Although bounding boxes can locate objects in an image, the bounding box contains not only the target object, but also the background and other content. In order to achieve more precise recognition of objects in images, semantic segmentation and instance segmentation are proposed. These two types of algorithms belong to the pixel-level object recognition algorithm. Segmentation algorithms mostly adopt code-decoding structures [18–20], such as fully convolutional networks (FCN), SegNet, U-Net, and DeepLab. The overall architecture of FCN is shown in Fig. 2. Multiple convolutional layers are used as encoding structures for down-sampling, and multiple deconvolutional layers are used as decoding structures for up-sampling.

Human pose recognition is a special object detection task. Key point detection of human body is very important to describe human posture and predict human behavior. Action classification, abnormal behavior detection, and other tasks can be accomplished by estimating the key points of human body. Mask R-CNN, an instance segmentation algorithm, can not only identify targets at pixel level, but also estimate the key points of human body. In addition, network structures such as DensePose,

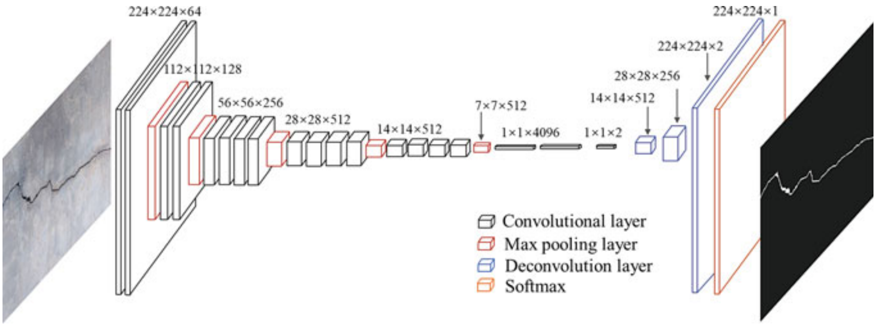


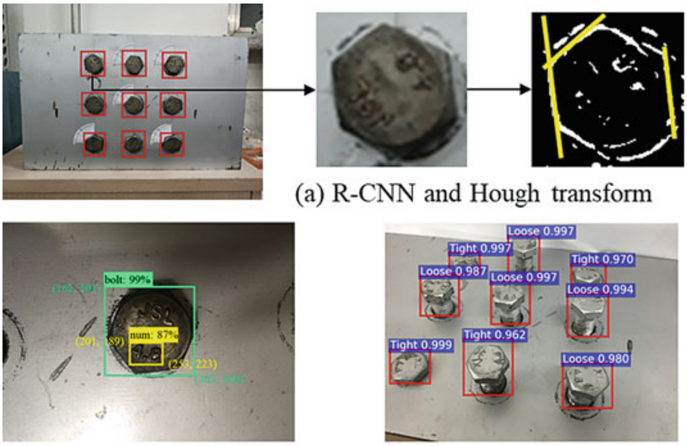
Fig. 2 FCN

OpenPose, AlphaPose, and DeepPose have also achieved remarkable effects in the field of human pose estimation.

3 Structural Safety Monitoring

3.1 Bolted Joints

In the process of dynamic construction, structural stability is poor, which can easily cause accidents. Therefore, it is necessary to monitor the status of structures under construction. Steel structures are often used in construction sites, such as steel trusses, steel columns, and support structures. Bolted joints have the advantages of simple structure, convenient installation, and strong reliability. These reasons have made bolts the preferred fastener. Under the influence of dynamic and static loads, bolts may be loosened and fallen off. In order to avoid interference with construction, non-contact monitoring methods can be preferred. The combination of machine vision and deep learning can quickly detect and count the state of bolts in steel structures. A camera or smartphone can quickly capture images of bolts. Faster R-CNN was used to identify and locate bolts in images. The image of each bolt is extracted on the basis of the localization information. Then, the edge lines of bolts in the image are extracted by using binarization and Hough transform. The detection result is shown in Fig. 3a. According to the change of edge information, the looseness angle could be recognized [21]. The head of the shoulder bolt is a regular hexagon, so this method could detect looseness angle within 60°. Grades of fastener appear on the head of each bolt. The "bolt" and "num" were identified and located simultaneously by SSD, as shown in Fig. 3b. According to the localization information, the looseness angle within 180° could be recognized [22]. When the bolt looseness angle is greater than 180°, the above method cannot effectively detect bolt looseness angle. A bolt to be loosen when the loosening is evident in a geometric change, which may even be visible to naked eyes. Therefore, Faster R-CNN can be used to directly detect large looseness



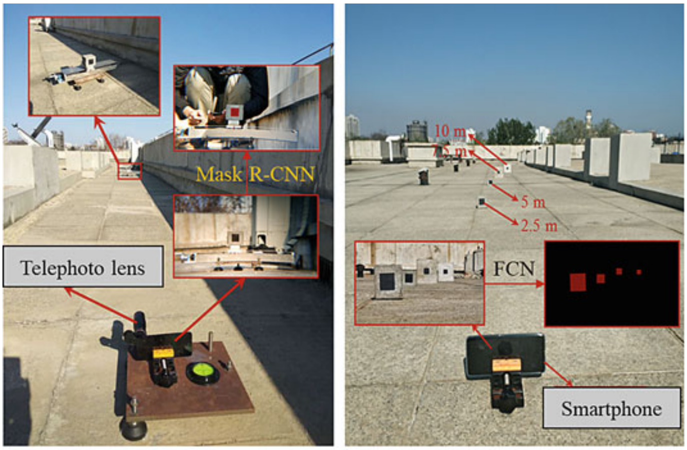
(a) R-CNN and Hough transform

(b) SSD

(c) Faster R-CNN

Fig. 3 Bolt damage detection

damage of bolts, as shown in Fig. 3c. This method has strong robustness even under the influence of structural vibration, illumination, and other environmental factors [23]. Deep learning is a data-driven recognition algorithm, and a large amount of data is the basic condition to ensure the recognition accuracy and generalization ability of model. 3D simulation software can quickly produce many bolt images. The detection model based on the simulation image still can accurately identify and locate bolts in a real image [24]. Although the simulation software cannot produce bolt images in a complex scene, this method provides a new idea for data augmentation.



(a) Single point

(b) Multi point

Fig. 4 Displacement monitoring based on semantic segmentation

3.2 Structural Displacement

During the construction process, the deformation and displacement of structures should be strictly controlled within the allowable range. With the development of high-rise and super high-rise buildings, high-support formwork and deep foundation pit appear more frequently in the construction process of building structures. These two types of structures, which have more potential hazards and high accident frequency, are the core parts of construction safety monitoring. The height of high-support formwork is more than five meters. Once it collapses, the impact and economic loss will be very serious. During the procedure of pouring concrete, the load borne by high-support formworks increase sharply, causing large deformation or even partial collapse and overall overturning. The foundation pit, located in the urban areas area, is not surrounded by sufficient space, so it is generally constructed by vertical excavation. Meanwhile, the depth of deep foundation pit is more than five meters, so it is very easy to collapse the side wall of the foundation pit. Therefore, it is necessary to monitor the displacement of high-support formwork and deep foundation pit. Deep learning can identify and locate target in an image at the pixel level. It provides the possibility for displacement monitoring. Portable devices such as smartphones are used to photograph artificial targets, which are identified and located by Mask R-CNN. Some parameters of target can be easily extracted from segmentation results. These parameters can be used to calculate the target displacement. Zhang et al. verified the feasibility of this method through static and dynamic experiments, respectively [25]. However, the proposed method can only monitor the short-range displacements. To achieve remote displacement monitoring, they installed a 22× optical zoom lens on a smartphone, as shown in Fig. 4a. The lens improves the ability of smartphones to photograph long-range targets. However, this method is limited by the lens, and it can only monitor the longitudinal one-point displacement. In order to monitor the displacement of deep foundation pit, Zhang et al. [26] proposed a longitudinal multipoint displacement monitoring method using FCN and smartphone. They used Huawei P30 Pro as the acquisition equipment, which has a high-performance camera with a 50× optical zoom. The detection result is shown in Fig. 4b. It can be used to detect the displacement of four longitudinal points within 10 m. This method does not require an external lens and is more portable.

Compared with the above method, the optical flow estimation-based method detects small motions without the use of paints or markers on the structure surface. However, the complex calculation process limits its real-time inference capacity. Deep learning can estimate optical flow with fewer parameters. With the help of GPU-accelerated computation, the optical flow method based on deep learning can be used for real-time monitoring [27]. The full-field optical flow and homography matrix were used to obtain the full-field structural displacement. Theoretically, the displacement of any point of structures can be obtained according to the full-field structural displacement diagram. Nevertheless, the optical flow estimation is affected by the background clutter. Dong et al. [28] proposed a full-field optical flow estimation algorithm based on deep learning. It reduces manual manipulation and provides more accurate measurements with less computation time. Subpixel subdivision technology can make up for the shortage of hardware and improve image resolution. Luan et al. propose a deep learning approach based on CNN to extract full-field high-resolution displacements at subpixel levels [29]. The results showed that the trained network can identify the pixels with sufficient texture contrast as well as their subpixel motions.

3.3 Structural Surface Quality

Construction quality evaluation is an important part of construction quality management. Once forms are removed, concrete surfaces may appear void, pockmarked surface, crack, and other damage. Deep convolutional neural network can identify and extract structural surface damage, it can improve the efficiency and accuracy of manual inspection. To locate surface damage in an image, sliding window algorithm cropped the image into several small images, which are fed into convolutional neural networks for classification [30]. In order to improve the speed of sliding window detector, various object recognition algorithms based on candidate regions are proposed, such as Faster R-CNN and YOLO. The identification result is shown in Fig. 5. Deng et al. [31] used YOLO to identify and locate cracks in images and compared the recognition effect with Faster R-CNN.

Object detection algorithms can locate surface damages through bounding box, which cannot be used to extract damage area, width, length, and other parameters. Semantic segmentation algorithm can identify and locate targets in an image at pixel level. For example, SegNet and FCN can be used for pixel-level recognition of cracks in images [32, 33]. Compared with object detection, the result of semantic segmentation is more precise. The identification result using semantic segmentation is shown in Fig. 6. Lee et al. [34] proposed crack width estimation method based on shape-sensitive kernels and semantic segmentation. Firstly, SegNet was used to identify cracks at pixel level. Then, the maximum width of the crack is measured based on

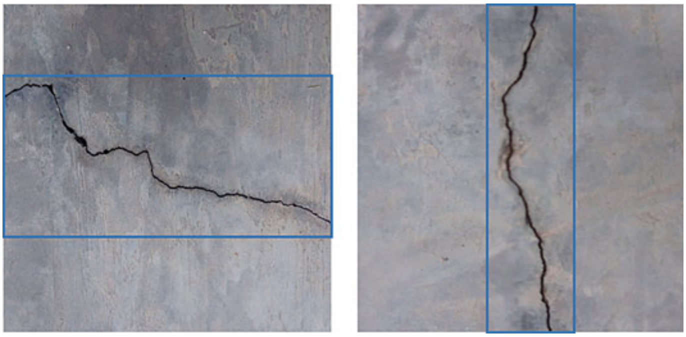


Fig. 5 Crack identification based on bounding box

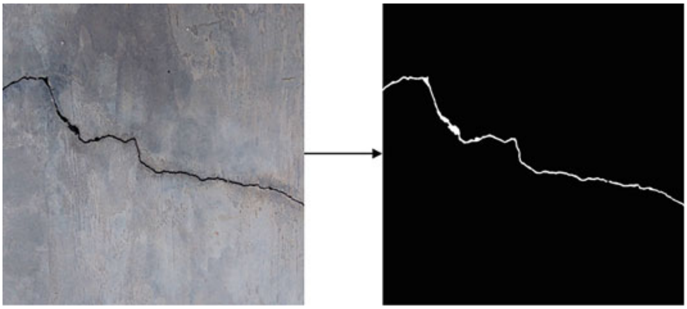


Fig. 6 Crack identification based on semantic segmentation

the identification results. A practical detection method should consider not only the detection accuracy, but also the detection speed. Cha et al. [35] proposed a real-time crack segmentation network, which has a great improvement in detection accuracy and speed. The model processes in real-time images at 1025 × 512 pixels, which is 46 times faster than in a recent work. According to the results of semantic segmentation, the parameters such as length, width, and area of damages can be obtained more conveniently. However, semantic segmentation requires pixel-level annotation data, and the annotation cost is very large. Meanwhile, semantic segmentation model is difficult to train and requires high computational performance.

To reduce the detection costs, object detection and traditional methods are combined to achieve crack refined detection. Object detection can identify and locate cracks accurately and quickly by bounding box. The positioning method based on the bounding box cannot quantitatively analyze the parameters of cracks. Some image processing methods such as binarization and filtering can be used to extract cracks in the bounding box at the pixel level. The identification result is shown in Fig. 7. Jiang et al. [36] proposed a real-time crack evaluation system based on deep learning and wall-climbing UAV. Firstly, SSD was used to identify and locate cracks in the

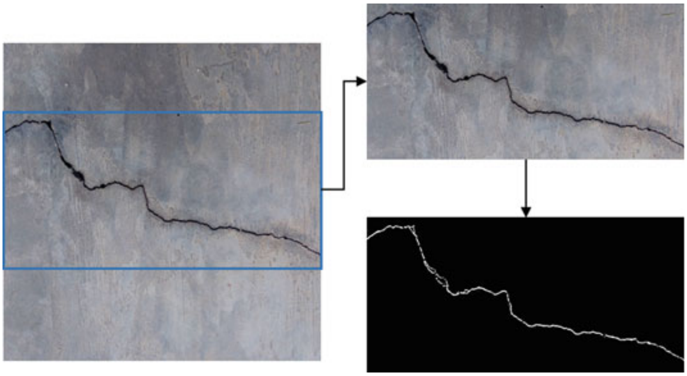


Fig. 7 Crack identification based on bounding box and image processing

images transmitted by UAV. Then, the images in bounding box were converted into binary image. The grayscale images were transformed into binary image through Otsu threshold segmentation, and morphological processing was applied to ensure continuity of cracks. These are image-based identification methods, which can only obtain two-dimensional spatial parameters of damage. Deep convolutional neural networks have also achieved remarkable results in the field of 3D point cloud segmentation. Beckman et al. [37] combined convolutional neural network with depth camera to detect the volume of concrete surface damages.

4 Worker Safety Management

A large number of workers on construction sites and the crosscutting of jobs are the main reasons for the higher casualty rate in construction industry than in any other industry. There are many potential hazards in the construction sites and workers should be vigilant of their surroundings. Therefore, it is desirable to track the number and status of workers in the construction site in real time. Object detection algorithms such as Faster R-CNN can identify and locate workers in different scenes and postures [38], as shown in Fig. 8. It can quickly count the number and location of workers in a large scene, providing a new possibility for construction worker detection.

Heavy equipment may lead to serious worker injuries, and equipment operators' view is easily obscured. Thus, it is necessary to detect workers near the equipment to provide safety warning for operators. In order to improve visibility, heavy equipment manufacturers install cameras on each side of the equipment (i.e., front, right, left, and rear) to provide a comprehensive view of the area around the equipment. However, this monitoring system cannot automatically extract information from images. Son et al. [39] proposed a real-time warning system using visual data and Faster R-CNN. This system used monocular cameras to estimate worker's position in three dimensions. In addition, some support structures without guardrails are also one of the important hazards. During the construction of an engineering

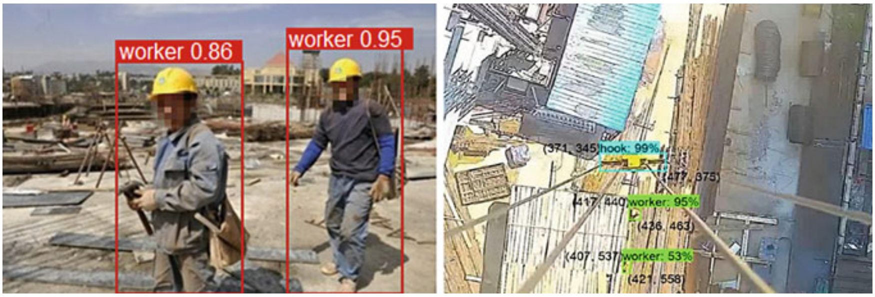


Fig. 8 Identification of worker

structure, workers tend to take shortcuts by crossing supports to perform daily activities and save time. However, crossing the support is very dangerous and forbidden. Fang et al. [40] used instance segmentation to detect workers crossing structural supports during the construction of a project. First, workers and structural supports are identified at the pixel level by the Mask R-CNN. Then, the positioning relationship between the structural support and the workers could be used to determine whether the worker is passing through the support. The proposed method could obtain the distance between workers and hazard sources, thus providing safety warning for managers and performing the correct behavior. At construction sites, workers are always in dynamic walking positions. Predicting workers' trajectories has great potential to improve workplace safety. Object identification-based tracking methods only rely on entity operation information and do not make full use of context information. Cai et al. [41] proposed a context-augmented long short-term memory approach for worker trajectory prediction. Compared with the traditional one-step prediction method, the proposed method could predict multistep trajectory to avoid error accumulation and effectively reduce final displacement error by 70%.

Workers should be vigilant of potential hazards. However, personal protection of workers is also very necessary. Safety helmets play an important role in protecting construction workers from accidents. Nevertheless, workers sometimes do not wear safety helmets for convenience. For the safety of construction workers, a high-precision and strong robustness helmet detection algorithm is urgently needed. Object detection algorithms such as Faster R-CNN, SSD, and YOLO can detect safety helmet in an image [42–44], as shown in Fig. 9. The recognition accuracy of detection algorithms based on deep learning depends on a large number of sample data. To evaluate the performance of Faster R-CNN, more than 100,000 image frames of construction workers were randomly selected from surveillance videos of 25 construction sites over a period of more than one year. The experimental results showed that this method has high accuracy, high recall rate, and fast speed, and could effectively detect non-helmet-use in different construction sites. It is conducive to improving the safety inspection and supervision level. Besides safety helmet and protective clothing, safety harness is the main protective equipment for workers working at height. Workers often forget or deliberately do not wear seat harness. This is a very

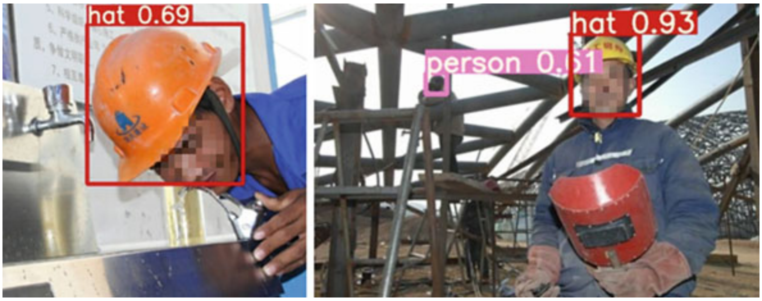


Fig. 9 Identification of safety helmet

dangerous behavior and is one of the main causes of falling from heights. Fang et al. [45] developed a visual-based automated approach that uses two convolutional neural networks to determine whether a worker is wearing a safety harness. First, Faster R-CNN was used to detect workers in an image, and the image in bounding box was cropped. Then, these cropped images were fed into a multilayer convolutional neural network to determine whether the workers wear seat harness. The accuracy rate and recall rate of the Faster R-CNN model were 99 and 95%, respectively, and the accuracy rate and recall rate of the CNN model were 80 and 98%, respectively.

Human body is very flexible, and posture can be used to identify worker behavior. Roberts et al. [46] proposed an activity analysis framework based on vision and deep learning that could estimate and track the posture of worker. Firstly, YOLO was used to identify and locate workers in an image, and the image in bounding box was cropped. Then, the cropped images were fed into a 2D skeleton detection network to estimate each joint coordinates of the worker. Posture estimation of worker is shown in Fig. 10. Finally, the pose of the same construction worker in different video frames was tracked. The proposed method was used to identify different worker activities (bricklaying and plastering), and the results showed that this method had the potential to evaluate individual worker activities. A 3D skeleton estimation network can directly determine the spatial position of human. The distance between workers can be used to estimate severity of crowding in working environment. High crowding may lead to dangerous working conditions and negative worker behavior. Yan et al. [47] proposed a vision-based crowd detection technology. First, Faster R-CNN was used to identify and locate workers in complex environment, and the image in bounding box was cropped. Then, the cropped images were fed into the 3D skeleton detection network to estimate each joint coordinates of the worker. Finally, according to the result of human skeleton estimation, the spatial position of each worker was obtained. Experimental results showed that this method could estimate the distance between two workers with an error of 0.45 m in three-dimensional space.

At present, worker status monitoring methods based on the fixed camera have some limitations, such as the perceptual mixing, occlusion, and illumination. Meanwhile, there is no camera inside the buildings under construction. Some mobile robots autonomously inspect indoor work sites and find potentially dangerous anomalies.

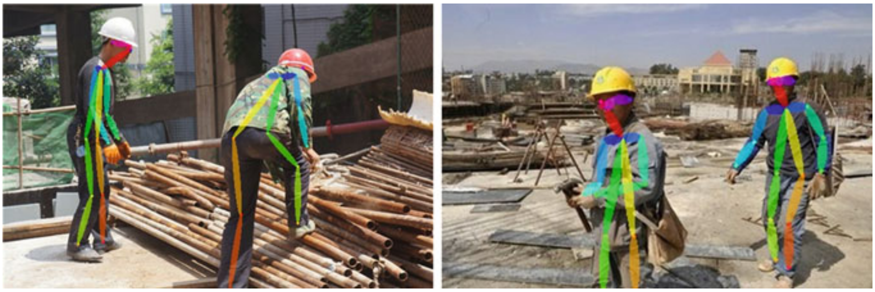


Fig. 10 Posture estimation of worker

Lee et al. [48] applied the perception module based on deep learning and simultaneous localization and mapping (SLAM) to target recognition and navigation of mobile robots. The proposed method could identify abnormal behaviors of some workers, such as not wearing safety helmets, standing on top of ladders, or falling. Identification of danger is only the first step of construction safety management. It is necessary to timely and accurately convey risk information to managers. Tang et al. [49] proposed a language-image framework that aims at understanding and detecting semantic roles of activities mentioned in safety rules. This framework includes semantic parsing of safety rules, construction object detectors using SSD and Faster R-CNN, and semantic role detectors. The experimental results showed that this framework can preliminarily describe the dangerous scene in an image in the form of language.

5 Construction Machinery Management

In order to improve construction efficiency, various machineries are used in the construction process, such as excavator, dump truck, bulldozer, scraper, crane, etc. However, safety accidents often occur with these construction machineries. Therefore, it is necessary to identify the position and status of construction machinery in real time. Object detection algorithms such as Faster R-CNN and R-FCN can identify and locate construction machinery in an image [50], as shown in Fig. 11. Kim et al. used R-FCN to identify five types of vehicles: dump truck, excavator, loader, cement mixer, and road roller [51]. The experimental results showed that the average accuracy of this method was 96.33%. Meanwhile, the synthesized image

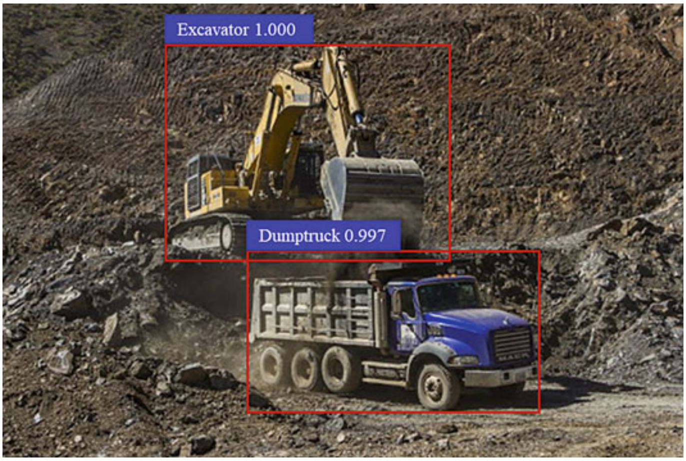


Fig. 11 Identification of construction machinery

data could be used for data augmentation, which can improve the vehicle detection performance [52]. The identification and location information of construction machinery can provide managers with the type, quantity, and distribution of construction machinery. Identifying and tracking construction machines also can avoid potential collisions and other accidents. Firstly, convolutional neural network was used to detect and track vehicles. Then, the tracking trajectory was fed into a hidden Markov model (HMM) that automatically discovers and assigns activity labels to objects. Roberts et al. [53] performed activity analysis of machineries based on the object detection results. The accuracy of activity analysis was found to be 86.8% for excavators and 88.5% for dump trucks. Of all construction machinery, tower cranes are the largest in size. The tower crane operator is far from the ground and cannot clearly observe the surrounding environment of lifting objects. High-definition cameras built into Unmanned Aerial Vehicles (UAVs) could help solve this problem. Roberts et al. [54] used SSD to identify and locate tower cranes in an image, which is photographed by UAVs.

All the aforementioned identified bounding box-based construction machinery recognition methods use horizontal detection algorithms for detection. However, construction machinery can be in any direction and position, and they are not necessarily horizontal or vertical. When construction machinery is densely parked together, rotated bounding box can be more accurately fitted to the machinery region in terms of the orientation. Guo et al. [55] proposed a construction machinery identification method based on orientation-aware feature (OAF) fusion convolutional neural network. The proposed OAF-SSD could be applied not only to the construction vehicle detection, but also to the detection problem of dense multiple objects in civil engineering, which can also identify the orientation of target objects (more useful for motion tracking and estimation).

In the similar fashion as human body, the posture and movements of construction machinery need to be estimated automatically. This has a significant impact on construction safety and the use of the machinery itself. Excavator is one of the most used construction equipment. Its operation is complicated, and its posture changes more. Hence, it is more difficult to evaluate the full body posture of excavators than other equipment. Liang et al. [56] proposed a vision-based marker-less pose estimation system for estimating the joints and components of excavators. This system adopted and improved the advanced convolutional network, namely the stack hourglass network for pose estimation. The results showed that this system could estimate excavator boom, stick, and bucket joint positions but had higher estimation error for the bucket location due to the occlusion issue. Luo et al. [57] built an integrated model based on stacked hourglass and stacked pyramid network to estimate the posture of excavators in an image. This model evaluated the posture of excavators by identifying six key points of excavators, as shown in Fig. 12. Occlusion can significantly interfere with the detection results of this method. Compared to the activity identification of construction workers, research on activity identification of construction machinery is very limited, mainly because of the lack of construction machinery actions datasets. Zhang et al. [58] produced a comprehensive video dataset of 2064 clips, which included five actions (digging, swinging, dumping, moving forward,

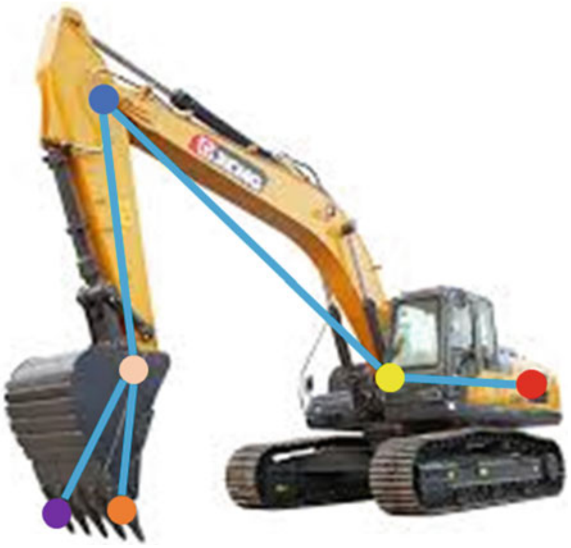


Fig. 12 Posture estimation of construction machinery

and moving backward) of excavators and dump trucks. CNN is used to extract image features, and long short-term memory (LSTM) is used to extract time characteristics from video frame sequences. These two types of features were used to identify these five actions.

6 Conclusion

Deep learning has been growing rapidly in the field of image processing. Some new network structures promote the development of construction safety monitoring using machine vision. On construction sites, structure, worker, and machinery are the three potential hazard sources, which are the most likely to cause accidents. Machine vision-based detection method is a non-contact detection technology, which can detect hazard sources without affecting the normal construction. At present, deep learning networks applied in the field of construction safety monitoring can be roughly divided into three categories: object detection, semantic segmentation, and pose estimation.

(1) Object detection using neural network is the simplest and most used detection method. Faster R-CNN, SSD, YOLO, R-FCN, and other object detection networks were used to identify and locate workers, machineries, and structural components. This type of detection method mainly relies on bounding box to identify and locate hazard sources. It can only complete detection tasks with low positioning accuracy.

(2) Compared with object detection, semantic segmentation has higher positioning accuracy and can achieve pixel-level object recognition. FCN, SegNet, U-Net, and other networks are used to locate targets, components, and workers with high precision. However, the disadvantages of this approach are also obvious. The training cost of semantic segmentation is high, and the detection speed is slow.

(3) Pose estimation networks recognize human posture through key point estimation. It can be used to detect different types of workers and assess worker activities. Meanwhile, it can also be used to locate the key points of construction machinery and identify the working status of machinery.

Many researchers have studied construction safety with deep learning and proposed many effective detection methods. However, from the perspective of smart construction site, the research on construction safety monitoring using deep learning is still in the initial stage. These detection techniques are far from being intelligent, and only identify and locate multiple targets on construction sites. In addition, the current detection technology also has some common problems. Deep learning is a data-driven algorithm. There is a lack of large image datasets for different construction scenes. Therefore, how to train a robust and high-precision detection model is still a difficult problem. Most importantly, visual detection method is easily affected by illumination, background, occlusion, and other factors. Now more and more complex construction site environment, some targets are often blocked, which is fatal to visual detection methods. A single type of sensor data often has great limitations, so the fusion of multitype sensor data may be a more feasible scheme for construction safety monitoring.

Acknowledgements This research has been supported by the Guangdong-Hong Kong-Macau Joint Laboratory Program under grant number: 2020B1212030009 and the Science and Technology Development Fund of the Macau SAR under research grant SKL-IOTSC-2021-2023. These generous supports are gratefully acknowledged.

References

1. Hammad A, Vahdatikhaki F, Zhang C, Mawlana M, Doriani A (2012) Towards the smart construction site: improving productivity and safety of construction projects using multi-agent systems, real-time simulation and automated machine control. In: Proceedings of the 2012 winter simulation conference, pp 1–12. IEEE, Germany
2. Hu Y, Castro-Lacouture D, Eastman CM (2019) Holistic clash detection improvement using a component dependent network in BIM projects. Autom Constr 105:102832
3. Mirzaei A, Nasirzadeh F, Parchami Jalal M, Zamani Y (2018) 4D-BIM dynamic time–space conflict detection and quantification system for building construction projects. J Constr Eng Manag 144(7):04018056
4. Yang H, Omidalizarandi M, Xu X, Neumann I (2017) Terrestrial laser scanning technology for deformation monitoring and surface modeling of arch structures. Compos Struct 169:173–179

5. Valença J, Puente I, Júlio E, González-Jorge H, Arias-Sánchez P (2017) Assessment of cracks on concrete bridges using image processing supported by laser scanning survey. Constr Build Mater 146:668–678
6. Ham N, Lee SH (2018) Empirical study on structural safety diagnosis of large-scale civil infrastructure using laser scanning and BIM. Sustainability 10(11):4024
7. Chen J, Cho YK (2018) Point-to-point comparison method for automated scan-vs-bim deviation detection. In: 17th international conference on computing in civil and building engineering, pp 1–8. Springer, Finland
8. Kifouche A, Baudoin G, Hamouche R, Kocik R (2017) Generic sensor network for building monitoring: design, issues, and methodology. In: 2017 IEEE conference on wireless sensors, pp 1–6. IEEE, Malaysia
9. Kuang KSC, Cantwell WJ, Thomas C (2003) Crack detection and vertical deflection monitoring in concrete beams using plastic optical fiber sensors. Meas Sci Technol 14(2):205–216
10. Casciati F, Fuggini C (2011) Monitoring a steel building using GPS sensors. Smart Struct Syst 7(5):349–363
11. Feng D, Feng MQ (2017) Experimental validation of cost-effective vision-based structural health monitoring. Mech Syst Signal Process 88:199–211
12. Harvey Jr, PS, Elisha G (2018) Vision-based vibration monitoring using existing cameras installed within a building. Struct Control Health Monit 25(11):e2235
13. Editorial Committee (Editor-in-chief: Xin Zhao) (2017). Report on informatization of building construction industry (2017): Application and development of smart construction sites. China Building Material Industry Press, (in Chinese)
14. Hashemi H, Abdelghany K (2018) End-to-end deep learning methodology for real-time traffic network management. Comput-Aided Civil Infrastruct Eng 33(10):849–863
15. Jia Y, Johnson M, Macherey W, Weiss RJ, Cao Y, Chiu CC, Wu Y (2019) Leveraging weakly supervised data to improve end-to-end speech-to-text translation. In: IEEE international conference on acoustics, pp 7180–7184. IEEE, UK
16. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Ronneberger O (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 24(9):1342–1350
17. Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
18. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440. IEEE, USA
19. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: international conference on medical image computing and computer-assisted intervention. Springer, Germany, pp 234–241
20. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848
21. Huynh TC, Park JH, Jung HJ, Kim JT (2019) Quasi-autonomous bolt-loosening detection method using vision-based deep learning and image processing. Autom Constr 105:102844
22. Zhao X, Zhang Y, Wang N (2019) Bolt loosening angle detection technology using deep learning. Struct Control Health Monit 26(1):e2292
23. Zhang Y, Sun X, Loh KJ, Su W, Xue Z, Zhao X (2020) Autonomous bolt loosening detection using deep learning. Struct Health Monit 19(1):105–122
24. Pham HC, Ta QB, Kim JT, Ho DD, Tran XL, Huynh TC (2020) Bolt-loosening monitoring framework using an image-based deep learning and graphical model. Sensors 20(12):3382
25. Zhang Y, Liu P, Zhao X (2020) Structural displacement monitoring based on mask regions with convolutional neural network. Construct Build Mater 120923
26. Zhang Y, Zhao X, Liu P (2019) Multi-point displacement monitoring based on full convolutional neural network and smartphone. IEEE Access 7:139628–139634

27. Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T (2017) Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2462–2470. IEEE, USA
28. Dong CZ, Celik O, Catbas FN, O'Brien EJ, Taylor S (2020) Structural displacement monitoring using deep learning-based full field optical flow methods. Struct Infrastruct Eng 16(1):51–71
29. Luan L, Wang ML, Yang Y, Sun H (2020) Extracting full-field subpixel structural displacements from videos via deep learning. arXiv preprint arXiv:2008.13715
30. Cha YJ, Choi W, Büyüköztürk O (2017) Deep learning-based crack damage detection using convolutional neural networks. Comput-Aided Civil Infrastr Eng 32(5):361–378
31. Deng J, Lu Y, Lee VCS (2020) Imaging-based crack detection on concrete surfaces using You Only Look Once network. Struct Health Monit 20(2):484–499
32. Zhang X, Rajan D, Story B (2019) Concrete crack detection using context-aware deep semantic segmentation network. Comput-Aided Civil Infrastruct Eng 34(11):951–971
33. Dung CV (2019) Autonomous concrete crack detection using deep fully convolutional neural network. Autom Constr 99:52–58
34. Lee JS, Hwang SH, Choi IY, Choi Y (2020) Estimation of crack width based on shape-sensitive kernels and semantic segmentation. Struct Control Health Monit 27(4):e2504
35. Choi W, Cha YJ (2019) SDDNet: Real-time crack segmentation. IEEE Trans Industr Electron 67(9):8016–8025
36. Jiang S, Zhang J (2020) Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system. Comput-Aided Civil Infrastruct Eng 35(6):549–564
37. Beckman GH, Polyzois D, Cha YJ (2019) Deep learning-based automatic volumetric damage quantification using depth camera. Autom Constr 99:114–124
38. Son H, Choi H, Seong H, Kim C (2019) Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. Autom Constr 99:27–38
39. Son H, Seong H, Choi H, Kim C (2019) Real-time vision-based warning system for prevention of collisions between workers and heavy equipment. J Comput Civ Eng 33(5):04019029
40. Fang W, Zhong B, Zhao N, Love PE, Luo H, Xue J, Xu S (2019) A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. Adv Eng Inform 39:170–177
41. Cai J, Zhang Y, Yang L, Cai H, Li S (2020) A context-augmented deep learning approach for worker trajectory prediction on unstructured and dynamic construction sites. Adv Eng Inform 46:101173
42. Wu J, Cai N, Chen W, Wang H., Wang G (2019) Automatic detection of hardhats worn by construction personnel: a deep learning approach and benchmark dataset. Autom Construct 106:102894
43. Zhao Y, Chen Q, Cao W, Yang J, Xiong J, Gui G (2019) Deep learning for risk detection and trajectory tracking at construction sites. IEEE Access 7:30905–30912
44. Fang Q, Li H, Luo X, Ding L, Luo H, Rose TM, An W (2018) Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Autom Constr 85:1–9
45. Fang W, Ding L, Luo H, Love PE (2018) Falls from heights: A computer vision-based approach for safety harness detection. Autom Constr 91:53–61
46. Roberts D, Torres Calderon W, Tang S, Golparvar-Fard M (2020) Vision-based construction worker activity analysis informed by body posture. J Comput Civ Eng 34(4):04020017
47. Yan X, Zhang H, Li H (2019) Estimating worker-centric 3D spatial crowdedness for construction safety management using a single 2D camera. J Comput Civ Eng 33(5):04019030
48. Lee MFR, Chien TW (2020) Intelligent robot for worker safety surveillance: deep learning perception and visual navigation. In: 2020 international conference on advanced robotics and intelligent systems, pp 1–6. IEEE, UK
49. Tang S, Golparvar-Fard M (2017) Joint reasoning of visual and text data for safety hazard recognition. In: Computing in civil engineering 2017, pp 450–457. ASCE, USA
50. Fang W, Ding L, Zhong B, Love PE, Luo H (2018) Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach. Adv Eng Inform 37:139–149

51. Kim H, Kim H, Hong YW, Byun H (2018) Detecting construction equipment using a region-based fully convolutional network and transfer learning. J Comput Civ Eng 32(2):04017082
52. Kim H, Bang S, Jeong H, Ham Y, Kim H (2018) Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation. Autom Constr 92:188–198
53. Roberts D, Golparvar-Fard M (2019) End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level. Autom Constr 105:102811
54. Roberts D, Bretl T, Golparvar-Fard M (2017) Detecting and classifying cranes using camera-equipped UAVs for monitoring crane-related safety hazards. In: Computing in civil engineering 2017, pp 442–449. ASCE, USA
55. Guo Y, Xu Y, Li S (2020) Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network. Autom Construct 112:103124
56. Liang CJ, Lundeen KM, McGee W, Menassa CC, Lee S, Kamat VR (2018) Stacked hourglass networks for markerless pose estimation of articulated construction robots. In: 35th international symposium on automation and robotics in construction, pp 843–849. Curran Associates, Germany
57. Luo H, Wang M, Wong PKY, Cheng JC (2020) Full body pose estimation of construction equipment using computer vision and deep learning techniques. Autom Construct 110:103016
58. Zhang J, Zi L, Hou Y, Wang M, Jiang W, Deng D (2020) A deep learning-based approach to enable action recognition for construction equipment. Adv Civil Eng 1–14