Transforming Construction with Reality Capture Technologies:
The Digital Reality of Tomorrow

August 23-25, 2022, Fredericton, New Brunswick, Canada

OBJECT-DETECTION APPLICATIONS FOR ON-SITE CONSTRUCTION PROCESSES: A REVIEW

Driouache, Y[1,*], and Motamedi, A[1,**]
[1] Department of Construction Engineering, École de technologie supérieure, Canada
[*] younes.driouache.1@ens.etsmtl.ca
[**] ali.motamedi@etsmtl.ca

Abstract: The advances in image capture technology have made it easier and more affordable to collect images and videos. Thus, most construction site are now equipped with surveillance cameras that continuously capture visual data which contain valuable information that can be used to assess the four main on-site construction processes: safety monitoring, progress monitoring, productivity monitoring and quality control. Computer vision algorithms and techniques are the most suitable tools to extract meanings (such as object location) from these data. The computer vision technique that has advanced the most and is the most used in the industry is object detection. This review aims to present the use of object detection to assess the four on-site construction processes and identify the challenges and opportunities that need to be addressed to correctly deploy and use this solution. Among the initial results there is a lack of research on object detection applied to quality control and the biggest challenges that object detection techniques face are the data acquisition, the construction environment and the limited hardware.

Keywords: Object detection; Computer vision; on-site processes; Construction

1 INTRODUCTION

During the construction phase of a project many workers perform various activities concurrently using various tools and equipment. To ensure that activities are performed safely and on schedule, effective monitoring needs to be performed. However, when a construction site is large, such monitoring is time consuming and requires high skills, which leads to high cost. It is claimed that in construction sites, around 400,000 images are taken for a typical commercial building project of approximately 750,000 sq. ft (Han et Golparvar-Fard 2017). With the advent of surveillance cameras (CCTV) used on construction sites and the development of mobile technologies (such as smartphones and tablets), this number will only rise. Moreover, the fact that surveillance cameras are continuously capturing images from the construction site, useful information regarding construction progress, compliance to safety rules or activity time of construction equipment/workers can be captured. With the development of computer vision techniques and their expansion across diverse industries, researchers in the construction industry are applying these techniques to automate the extraction of the previously mentioned information and help construction management team members in their decision-making tasks.

The goals of this report are to review the use of object-detection techniques in on-site construction processes such as safety, progress, productivity monitoring and quality control and identify the challenges and opportunities that can affect the deployment of such techniques in construction sites. The paper is structured as follows: 1) Discussion around the two types of object-detection techniques, 2) Methodology of the review, 3) Analysis of the collected data, and 4) Discussion about the application of object detection and challenges and opportunities.

2 BACKGROUND

Object detection is a technique that performs both classification and localization. Object detection is done first, to identify the types/classes of objects present in an image, then localization is done by drawing a bounding box around each of the classified objects. The methods used to perform object detection can be divided into two categories: traditional and deep learning methods. Traditional methods are based on handcrafted features such as color and shape, but objects can also be detected based on their motion within the scene. Deep learning methods can be categorized into one-stage and two-stage detectors.

2.1 Traditional methods

2.1.1 Motion-Based Methods

The main technique that is used for motion-based object detection is background subtraction. The purpose of background subtraction is to detect a moving object in a scene. To achieve this, the image is divided into a foreground that contains the moving object and a background that contains no moving objects. If the pixel wise (i.e., the calculation is performed pixel-by-pixel) difference between the background image and current frame exceeds a certain threshold, the pixel is considered as being a part of the foreground. This method can be affected by dynamic backgrounds and variation of illumination (Piccardi 2004).

2.1.2 Methods Based on Feature Extraction

Object detection based on feature extraction aims to recognize and localize a type of object (e.g., human, vehicle, etc.) by extracting specific feature representations of the object in the image. The two main image features that are extracted are colors and shapes. For shapes, the histogram of oriented gradients (HOG) descriptor has been used to describe the appearance and shape of an object where a histogram of gradient orientation or a histogram of edge orientation is computed. The concatenation of these histograms forms the representation (Dalal et Triggs 2005). The other shape features that have been used to identify the presence of the object in an image are the Haar-like features (Papageorgiou, Oren et Poggio 1998).
Color features have been used to recognize objects that are easily identifiable by their color (e.g., hardhat, safety vest). Swain and Ballard (1991) demonstrated that the color histogram is robust, easy to compute and that it is invariant to rotation and translation. Yet it is affected by variations in illumination.

2.2 Deep learning methods

2.2.1 Two-Stage Detectors

The **R-CNN** (Girshick *et al.* 2014) algorithm proposes regions in the image using a high-capacity convolutional neural network (CNN) that includes all possible object candidates. Then, the feature vectors of each region are extracted using a CNN to identify the object present in the proposed regions. **Fast R-CNN** (Girshick 2015) trains in the same network a detector and a bounding box regressor which reduce computation. Thus, **Fast R-CNN** can detect objects faster than **R-CNN**. **Faster R-CNN** (Ren *et al.* 2016) has been proposed to resolve the main issue of Fast R-CNN which is the region proposal computation. (Ren *et al.* 2016) developed a Region Proposal Network (RPN) which shares the extracted features with the detector network to enable cost-free region proposals. It provides improved accuracy and speed.

2.2.2 One-Stage Detectors

YOLO (You Only Look Once) (Redmon *et al.* 2016) as a one-stage detector can detect full-image objects in real-time. It performs the regression in grid regions and extracts features using CNN. In 2017, **YOLOv2 (YOLO9000)** was released (Redmon et Farhadi 2017). Its main advancements are 1) batch normalization (BN) that accelerates convergence and helps to regularize the model, 2) high-resolution classifier, and 3) use of anchor boxes. In 2018, **YOLOv3** was released (Redmon et Farhadi 2018) with these improvements: 1) multi-label classifier, 2) three different scale feature maps to predict bounding boxes, and 3) deeper backbone named Darknet53. In 2020, Alexey Bochkovskiy et al. introduced **YOLOv4** which uses data augmentations and post-process methods to improve performance (Bochkovskiy, Wang et Liao 2020).
As another one-stage detector, **Single-Shot Detector** (SSD) (Liu *et al.* 2016) takes advantage of the regression of YOLO and the anchor approach of the Faster R-CNN.

3 REVIEW METHODOLOGY

To identify relevant academic publications, two major databases were (i.e., Scopus and Web of Science) used in this research. These databases offer an extensive collection of academic resources and provide adequate tools to find relevant references. Moreover, one can export a large amount of structured data for analysis. The search keywords included the main on-site construction processes, namely, safety, progress, productivity monitoring and quality control. The main search was performed through Title, Abstract and Keywords and included the following keywords: (*"computer vision"* OR *"object detection"* OR *"object recognition"*) AND (*"construction"*). Thus, it was possible to identify a wide number of articles relevant to the topic. To refine the search on each of the construction processes, a secondary search was performed using these keywords: ("computer vision" OR "object recognition" OR "object detection" OR "Faster R-CNN" OR "YOLO*" OR "Mask R-CNN" OR "Fast R-CNN" OR "R-CNN" OR "CNN") AND (type of on-site process (safety, progress, productivity, quality)).

The initial search identified 482 articles. From this, the relevant articles were selected based on the following criteria: (1) the main source of data comes from images and/or videos; (2) the framework developed in the article uses an object-detection algorithm; (3) the method does not use additional 3D data, such as point cloud; and (4) research focuses on on-site construction processes. By applying the mentioned criteria, the final number of relevant articles to be used for the analysis was 56.

4 DATA OVERVIEW

Table 1 list the journals in which identified articles were published. The journal of Automation in Construction holds the most articles. The number of publications regarding object detection in construction has significantly increased over time. There has been a burst of publication since 2016-2017, as shown in Figure 1. It is speculated that the fast development of Deep Learning methods, the increase in computational power, the data availability and the sharing of frameworks and packages from Deep Learning researchers (e.g., Tensorflow (2015), Pytorch (2016)) have initiated this burst.

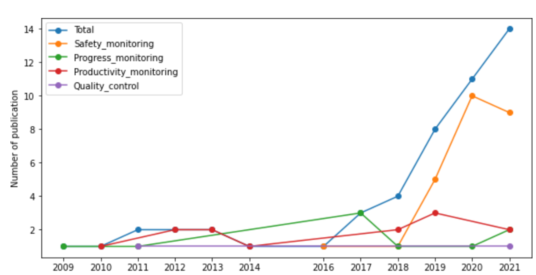

Figure 1: Publications per year across construction processes

Table 1: List of the most represented journals and conferences

Journal title	**Number of articles**
Automation in Construction	19
Journal of Computing in Civil Engineering	9
Advanced Engineering Informatics	4
Journal of Construction Engineering and Management	4
IEEE Access	2
Remote Sensing	2
Engineering, Construction and Architectural Management	2
Conference title	
ISARC - International Symposium on Automation and Robotics in Construction and Mining	3
EG-ICE, European Group for Intelligent Computing in Engineering	1

The relevant papers were grouped based on the following four construction processes: Safety, Progress, Productivity monitoring and Quality control. Figure 2a shows the distribution of papers in each category. Safety and productivity monitoring are the two categories where object detection has been the most used. Quality control could greatly benefit from object detection; however, very few papers focus on it.

The average publication year of the papers that used traditional object detection is around 2015, whereas those that used deep learning is around 2020. Figure 3 visualizes the use of various traditional and deep learning object-detection methods through the years. The results demonstrate that after 2017 very few traditional methods were used, which coincides with the arrival of more robust deep learning techniques. Moreover, after the release of YOLOv3 in 2018, the YOLO model family was the most used due to its robustness and its capacity to perform fast and accurate object detection. From 2018 to 2022, the YOLO model family was the most used with 15 papers followed by Faster R-CNN, Mask R-CNN and SSD.

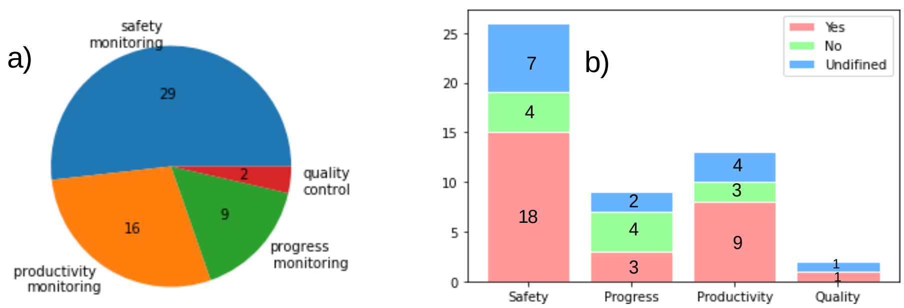

Figure 2: a) Number of papers per on-site construction process, b) Proportion of real-time object detection per type of process

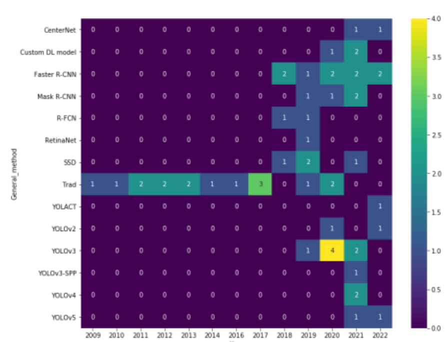

Figure 3: Object-detection techniques used from 2009 to 2022

Some methods provided by the different papers are not necessarily in real-time. Indeed, it is best to have real-time insights when decisions need to be made quickly. Real-time object detection is needed for safety, productivity monitoring and quality control. For progress monitoring, the reviewed papers showed that real-time detection is not a necessity because most of the techniques were not in real-time or did not mention that they were, as shown in Figure 2b.

The target objects that are detected in each article can be grouped into 9 categories. Each type of monitoring has its own set of target object. In safety monitoring, the most targeted objects are person/workers and PPE to ensure that workers are following the safety rules, but in some cases construction vehicles need to be detected in the case of collision. Similarly, progress and productivity monitoring target person/workers. However, they also target material, construction equipment, and building elements to provide quality information on how productive the workforce and construction vehicles such as excavator are, but also which type of material or building elements have been added. Quality control is mainly focused on materials. Figure 4 shows the distribution of detected objects for each type of monitoring.

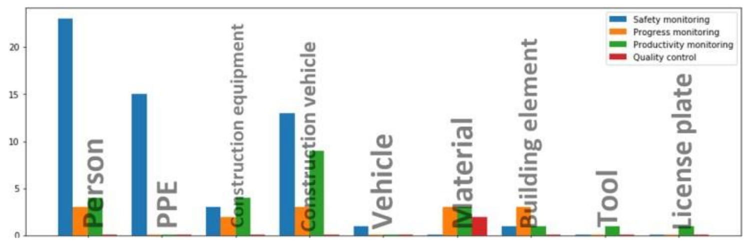

Figure 4: Distribution of types of detected objects for each type of monitoring

5. APPLICATION OF OBJECT DETECTION IN THE CONSTRUCTION INDUSTRY

5.1 Safety Monitoring

For safety monitoring, object-detection techniques have mainly been used to detect safety hazard scenarios and violation of safety rules. Safety hazard scenarios can be divided into two categories: static and dynamic. Both of these categories concern two types of objects: workers and PPE. For example, (Mneymneh, Abbas et Khoury 2019) proposed a framework to detect workers without helmets that used background subtraction and a classifier to identify workers and a color-based classification to detect hardhats. Other vision techniques based on Deep Learning have been used to locate workers within complex and dynamic environments (e.g., Fang *et al.* 2018a; Shen *et al.* 2021; Li *et al.* 2021a; Kim, Kim et Shchur 2021).

In static hazard scenarios, the focus is on the interaction between the worker and their surroundings. To understand interactions between workers and their environment and detect dangerous behavior, (Tang, Roberts et Golparvar-Fard 2020) developed a framework that first detected the workers and construction objects (such as a ladder, scaffolding) using Faster R-CNN and then used their HOI (Human-Object Interaction) recognition model to predict the interaction between worker and tools/equipment. To prevent falls from height, object-detection methods have been used to monitor compliance to fall protection system and equipment. For example, (Fang *et al.* 2019) developed a system to detect workers who traverse structural supports by identifying the relationship between the worker and structural support to determine unsafe behaviors. Similarly, (Fang *et al.* 2018b) developed a framework based on a Faster R-CNN to detect workers and a CNN to identify whether they were wearing their harness.

For dynamic hazard scenarios, the focus is mainly on construction vehicles and moving/falling objects. Many of the research projects aim to reduce “struck-by” accidents. (Zhang *et al.* 2020) proposed a framework that detects workers and construction vehicles (such as excavator) using a Faster R-CNN model. Once these objects had been detected, the risk of collision could be evaluated. Similarly, (Kim, Lee et Kamat 2020) developed a framework that first detects workers and construction vehicles using YOLOv3 model and then predicts their trajectory. Thus, the methods mentioned above help to detect the unsafe proximity of workers to a dynamic or static hazard.

5.2 Progress Monitoring

Accurate progress monitoring gives useful insights regarding the as-built states of the construction project to prevent time and cost overruns. As the process involves the analysis of visual data, efficient automated systems can be used to help construction teams reduce reworks and errors (Ekanayake *et al.* 2021). Progress monitoring is performed for interior and exterior construction.

To provide insights as to the progress of interior construction, (Roh, Aziz et Peña-Mora 2011) developed a framework that compares as-built photos with an as-planned 3D BIM model in a 3D walk-through model by first detecting the material in the as-built photo and then seeing if all the material from the 3D BIM model are present in the photo. Similarly, (Deng *et al.* 2020) proposed a method that detects tiles by extracting LBP (Local Binary Pattern) features. The edge coordinate of the detected tiles is then converted from pixel to real-world coordinates to transform it into a BIM model. A different approach was used to detect interior materials, such as studs, electrical outlets, insulation and three states for drywall sheets (i.e., painted, plastered, and installed), which is based on colors and shapes (Hamledari, McCabe et Davari 2017).

Moreover, exterior construction has also greatly benefited from object-detection algorithms to assess progress without necessarily comparing visual data with a BIM model. (Hevesi *et al.* 2021) proposed a two-stage method that first detects construction vehicles, construction equipment, resources and materials and then evaluates the relationship between these objects to estimate the progress. An alternative method to assess progress is to provide adequate information about workforce and equipment using videos. (Zhu, Ren et Chen 2017) developed a framework including an object detector based on HOG features and latent Support Vector Machine (Felzenszwalb *et al.* 2010) to identify and track workforce and equipment. Comparing as-built images with an as-planned BIM model is also utilized for exterior construction monitoring. For example, (Ibrahim *et al.* 2009) proposed a system that automatically generates work packages (groups of small, related tasks in a project) by analyzing construction site videos to detect changes. Wang *et al.* (2021) developed a framework that includes object detection and multi-object tracking to locate and acquire temporal information of precast walls. The collected information is then transferred to the BIM model and the corresponding walls are matched using the temporal information to assess progress.

5.3 Productivity Monitoring

The term *Productivity* in the construction industry is defined as being the maximization of an output while optimizing the inputs, in which the inputs are workers, equipment and material (Naoum 2016). Thus, vision-based techniques, such as object detection, can help automate productivity monitoring by locating workers, equipment and describing their interactions and activities. Productivity monitoring can be divided into human and equipment/vehicle productivity.

In this example of productivity monitoring, still images from construction sites that show the execution of an activity were used by (Luo *et al.* 2018) to first locate workers and tools using a Faster R-CNN model and then to recognize construction activity performed by the workers. Similarly, a zero-shot human-object interaction detection algorithm was used with a knowledge graph to extract a visual relationship and update construction activity knowledge graphs (Pan *et al.* 2022). In another study, only detected workers and their positional relationships were used to determine the type of activity (Li *et al.* 2022). Other techniques assess productivity by first detecting workers and work objects to determine how much time the workers spend on a specific task using their positional relationship (Li *et al.* 2021b; Li *et al.* 2022).

Assessing productivity of construction equipment and construction vehicles is particularly important for earthmoving processes. In a study from (Roberts et Golparvar-Fard 2019), the RetinaNet model was used to detect excavators and dump trucks and a FCNT (fully-convolutional network-based tracker) was used to track them. Then their trajectories were computed and fed to a hidden Markov model to evaluate the type of activities performed, their duration and the transition time between each activity. By using photogrammetry techniques and an object-detection algorithm, (Bügler *et al.* 2014) evaluated the volume of excavated soil and generated statistics about equipment activities, such as loading and idle times to accurately measure the productivity of construction equipment and vehicles. Moreover, after detecting dump trucks and excavators using an R-FCN model, (Kim *et al.* 2018) used their bounding boxes to establish a context reasoning system that could be used as an input for a process simulation software to generate reports on productivity and cost analysis.

5.4 Quality control

The construction industry frequently experiences cost and schedule overruns that are mainly due to delays in material delivery or reworks. Establishing quality control procedures can help to reduce the amount of rework. However, quality control procedures mainly require a large amount of manual labor. Thus, computer vision techniques, such as object detection, can be helpful to automate some of the related procedures due to their ability to extract spatial and dimensional information. So far, object-detection techniques have not really been fully explored for on-site quality control. However, some exploratory works have been performed. (Lin et Fang 2011) proposed a system based on geometric characteristics of the tile surface to evaluate tile alignment. A mask R-CNN technique and a stereo vision camera were used to detect steel bars and generate information such as quantity, spacing, diameter and length of steel bars (Kardovskyi et Moon 2021). The low number of publications for object detection applied to on-site quality control shows that it is a topic that needs to be further explored in the future, in particular, for the proactive detection of non-compliance to quality protocols.

6. CHALLENGES & OPPORTUNITIES

6.1 Data Acquisition Challenges

There is a lack of datasets specific to construction environments. Given that an object-detection algorithm requires a repository of visual data, the unavailability of domain datasets causes difficulties in training and evaluating these types of algorithms. Therefore, researchers must create their own datasets, which mostly results in incomplete datasets, which can cause an overfitting problem for the trained models. Moreover, data are mainly gathered in a specific construction site and requiring specific camera settings, all of which limits the reuse of datasets. Thus, the trained algorithm cannot be deployed in other construction site without being retrained with the data of the new site.

Additionally, maintaining the high quality of input data is challenging. Even if an object-detection algorithm has a well-structured architecture and has been trained on well-annotated data, if the inputted data is not of high quality (e.g., blurry or having some loose pixels) good results cannot be attained. Therefore, it is necessary to use cameras of sufficient quality and to define the specification for camera placement (angles, distance from the desired object, etc.).

6.2 Construction Environment Challenges

Construction sites are very dynamic with a multitude of simultaneous activities, which usually translates to cluttered backgrounds and occlusions in the camera's FoV. Thus, object-detection algorithms can be affected by a variety of object appearance variations such as viewpoint, scale or posture (for worker detection). Moreover, the object-detection systems are generally used to analyze outdoor scenes, which make them susceptible to weather conditions and illumination variations.

6.3 Performances Issues

For some monitoring tasks, such as safety and productivity monitoring, and quality control, real-time object detection is a necessity. This is quite a challenge because currently, most of the algorithms have large numbers of parameters which leads to the use of a substantial amount of memory to perform their computations and requires a large amount of space in disk storage. It is quite a concern when such an object-detection algorithm needs to be deployed on a limited-capacity server (only use CPUs) on a construction site or on an embedded system (such as a UAV). Some reviewed papers mentioned that their object-detection system worked in real-time but most of them took advantage of GPUs, which accelerates inference (predictions), to make that statement. It is not guaranteed that this type of component will be available on the machine that is present at the construction site. It is therefore preferable to give the inference time on a CPU. Moreover, to reduce the number of parameters of an object-detection model, the exploration of a lighter backbone can be done. This can reduce the inference time on the CPU (and GPU). Nevertheless, it will impact the precision of the model. The challenge is in finding a trade-off between inference time and precision to produce an efficient solution that can be deployed on most machines. Reducing the number of parameters of an object-detection model can also reduce the overfitting when the number of data is limited and there is a small number of classes to detect.

7. CONCLUSION

Object-detection techniques are very important for visual based framework applied to construction. Indeed, these techniques are the basis for the other computer vision techniques such as action recognition, object tracking, etc. In this review, the applications of object detection in on-site construction processes were explored. Traditional image processing and Deep Learning techniques were discussed as well as the type of object detected for each on-site processes and if the developed system perform in real-time. It was shown that object-detection algorithms are more used in safety and productivity monitoring, followed by progress monitoring. However, for on-site quality control it has not really been explored. Among the different Deep Learning techniques, the YOLO family is the most used technique in the construction industry because of its robustness and fast inference.

Challenges in implementing object detection for on-site construction processes were identified. One of the most important challenges is collecting a sufficient amount of data. The various reviewed papers showed that most of the time, researchers needed to collect data on their own which led to not having enough annotated data to adequately train and evaluate the developed frameworks. Another important challenge is the use of good quality data. Indeed, the dynamic context of a construction site makes it difficult to always input good quality images into the developed object-detection method. Also, using object-detection methods on construction sites can be affected by the hardware at hand. Indeed, the developed methods need to be precise enough but also lightweight enough to be applicable in real conditions.

8. REFERENCES

Azar, E.R. et B.Y. McCabe. 2013. « A visual sensing approach to estimate material hauling cycles in heavy construction and surface mining jobsites ». In ISARC 2013 - 30th International Symposium on Automation and Robotics in Construction and Mining, Held in Conjunction with the 23rd World Mining Congress. (2013), p. 774-782.

Bochkovskiy, Alexey, Chien-Yao Wang et Hong-Yuan Mark Liao. 2020. « YOLOv4: Optimal Speed and Accuracy of Object Detection ». arXiv:2004.10934v1 [cs].

Bügler, M., G. Ogunmakin, J. Teizer, P.A. Vela et A. Borrmann. 2014. « A comprehensive methodology for vision-based progress and activity estimation of excavation processes for productivity assessment ». In EG-ICE 2014, European Group for Intelligent Computing in Engineering - 21st International Workshop: Intelligent Computing in Engineering 2014. (2014).

Chen, S. et K. Demachi. 2021. « Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph ». Automation in Construction, vol. 125.

Dalal, N. et B. Triggs. 2005. « Histograms of Oriented Gradients for Human Detection ». In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). (San Diego, CA, USA, 2005), p. 886893. IEEE.

Delhi, V.S.K., R. Sankarlal et A. Thomas. 2020. « Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques ». Frontiers in Built Environment, vol. 6.

Deng, H., H. Hong, D. Luo, Y. Deng et C. Su. 2020. « Automatic Indoor Construction Process Monitoring for Tiles Based on BIM and Computer Vision ». Journal of Construction Engineering and Management, vol. 146, nº 1.

Ekanayake, Biyanka, Johnny Kwok-Wai Wong, Alireza Ahmadian Fard Fini et Peter Smith. 2021. « Computer vision-based interior construction progress monitoring: A literature review and future research directions ». Automation in Construction, vol. 127, p. 103705.

Fan, Z., C. Peng, L. Dai, F. Cao, J. Qi et W. Hua. 2020. « A deep learning-based ensemble method for helmet-wearing detection ». PeerJ Computer Science, vol. 6, p. 1-21.

Fang, Q., H. Li, X. Luo, L. Ding, H. Luo, T.M. Rose et W. An. 2018a. « Detecting non-hardhat-use by a deep learning method from far-field surveillance videos ». Automation in Construction, vol. 85, p. 1-9.

Fang, Weili, Lieyun Ding, Hanbin Luo et Peter E.D. Love. 2018b. « Falls from heights: A computer vision-based approach for safety harness detection ». Automation in Construction, vol. 91, p. 53-61.

Fang, Weili, Botao Zhong, Neng Zhao, Peter E.D. Love, Hanbin Luo, Jiayue Xue et Shuangjie Xu. 2019. « A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network ». Advanced Engineering Informatics, vol. 39, p. 170-177.

Fang, W., L. Ma, P.E.D. Love, H. Luo, L. Ding et A. Zhou. 2020. « Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology ». Automation in Construction, vol. 119.

Felzenszwalb, Pedro F, Ross B Girshick, David McAllester et Deva Ramanan. 2010. « Object Detection with Discriminatively Trained Part Based Models ». p. 20.

Girshick, Ross, Jeff Donahue, Trevor Darrell et Jitendra Malik. 2014. « Rich feature hierarchies for accurate object detection and semantic segmentation ». arXiv:1311.2524 [cs].

Girshick, Ross. 2015. « Fast R-CNN ». arXiv:1504.08083 [cs].

Goh, Y.M., J. Tian et E.Y.T. Chian. 2022. « Management of safe distancing on construction sites during COVID-19: A smart real-time monitoring system ». Computers and Industrial Engineering, vol. 163.

Golcarenarenji, G., I. Martinez-Alpiste, Q. Wang et J.M. Alcaraz-Calero. 2021. « Machine-learning-based top-view safety monitoring of ground workforce on complex industrial sites ». Neural Computing and Applications.

Gong, J. et C.H. Caldas. 2010. « Computer vision-based video interpretation model for automated productivity analysis of construction operations ». Journal of Computing in Civil Engineering, vol. 24, nº 3, p. 252-263.

Han, Kevin K. et Mani Golparvar-Fard. 2017. « Potential of big visual data and building information modeling for construction performance analytics: An exploratory study ». Automation in Construction, vol. 73, p. 184-198.

Hamledari, Hesam, Brenda McCabe et Shakiba Davari. 2017. « Automated computer vision-based detection of components of under-construction indoor partitions ». Automation in Construction, vol. 74, p. 78-94.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren et Jian Sun. 2014. « Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition ». arXiv:1406.4729 [cs], vol. 8691, p. 346-361.

Hevesi, Peter, Ramprasad Chinnaswamy Devaraj, Matthias Tschöpe, Oliver Petter, Janis Nikolaus Elfert, Vitor Fortes Rey, Marco Hirsch et Paul Lukowicz. 2021. « Towards Construction Progress Estimation Based on Images Captured on Site ». In Industrial IoT Technologies and Applications, sous la dir. de Peñalver, Lourdes et Lorena Parra, p. 141-161. Coll. « Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering », vol. 365. Cham : Springer International Publishing.

Hu, Q., Y. Bai, L. He, Q. Cai, S. Tang, G. Ma, J. Tan et B. Liang. 2020. « Intelligent Framework for Worker-Machine Safety Assessment ». Journal of Construction Engineering and Management, vol. 146, nº 5.

Ibrahim, Y.M., T.C. Lukins, X. Zhang, E. Trucco et A.P. Kaka. 2009. « Towards automated progress assessment of workpackage components in construction projects using computer vision ». Advanced Engineering Informatics, vol. 23, nº 1, p. 93-103.

Joachims, Thorsten. 1999. *Making large scale SVM learning practical*. Universität Dortmund, Fachbereich Informatik der Universität Dortmund; 24 p.

Kang, K.-S., Y.-W. Cho, K.-H. Jin, Y.-B. Kim et H.-G. Ryu. 2022. « Application of one-stage instance segmentation with weather conditions in surveillance cameras at construction sites ». Automation in Construction, vol. 133.

Kardovskyi, Yurii et Sungwoo Moon. 2021. « Artificial intelligence quality inspection of steel bars installation by integrating mask R-CNN and stereo vision ». Automation in Construction, vol. 130, p. 103850.

Kim, H., K. Kim et H. Kim. 2016. « Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects ». Journal of Computing in Civil Engineering, vol. 30, nº 4.

Kim, D., S. Lee et V.R. Kamat. 2020. « Proximity Prediction of Mobile Objects to Prevent Contact-Driven Accidents in Co-Robotic Construction ». Journal of Computing in Civil Engineering, vol. 34, nº 4.

Kim, H., S. Bang, H. Jeong, Y. Ham et H. Kim. 2018. « Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation ». Automation in Construction, vol. 92, p. 188-198.

Kim, H., Y. Ham, W. Kim, S. Park et H. Kim. 2019. « Vision-based nonintrusive context documentation for earthmoving productivity simulation ». Automation in Construction, vol. 102, p. 135-147.

Kim, K., S. Kim et D. Shchur. 2021. « A UAS-based work zone safety monitoring system by integrating internal traffic control plan (ITCP) and automated object detection in game engine environment ». Automation in Construction, vol. 128.

Li, N., X. Lyu, S. Xu, Y. Wang, Y. Wang et Y. Gu. 2021a. « Incorporate Online Hard Example Mining and Multi-Part Combination into Automatic Safety Helmet Wearing Detection ». IEEE Access, vol. 9, p. 139536-139543.

Li, J., X. Zhao, G. Zhou, M. Zhang, D. Li et Y. Zhou. 2021b. « Evaluating the work productivity of assembling reinforcement through the objects detected by deep learning ». Sensors, vol. 21, nº 16.

Li, J., X. Zhao, G. Zhou et M. Zhang. 2022a. « Standardized use inspection of workers' personal protective equipment based on deep learning ». Safety Science, vol. 150.

Li, J., G. Zhou, D. Li, M. Zhang et X. Zhao. 2022b. « Recognizing workers' construction activities on a reinforcement processing area through the position relationship of objects detected by faster R-CNN ». Engineering, Construction and Architectural Management.

Lin, Kuo-Liang et Jhih-Long Fang. 2011. « APPLICATIONS OF COMPUTER VISION ONTILE ALIGNMENT INSPECTION ». 2011 Proceedings of the 28th ISARC, p. 6.

Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu et Alexander C. Berg. 2016. « SSD: Single Shot MultiBox Detector ». arXiv:1512.02325 [cs], vol. 9905, p. 21-37.

Luo, X., H. Li, D. Cao, F. Dai, J. Seo et S. Lee. 2018. « Recognizing Diverse Construction Activities in Site Images via Relevance Networks of Construction-Related Objects Detected by Convolutional Neural Networks ». Journal of Computing in Civil Engineering, vol. 32, nº 3.

Luo, H., J. Liu, W. Fang, P.E.D. Love, Q. Yu et Z. Lu. 2020. « Real-time smart video surveillance to manage safety: A case study of a transport mega-project ». Advanced Engineering Informatics, vol. 45.

Meng, L., Z. Peng, J. Zhou, J. Zhang, Z. Lu, A. Baumann et Y. Du. 2020. « Real-time detection of ground objects based on unmanned aerial vehicle remote sensing with deep learning: Application in excavator detection for pipeline safety ». Remote Sensing, vol. 12, nº 1.

Mneymneh, B.E., M. Abbas et H. Khoury. 2019. « Vision-Based Framework for Intelligent Monitoring of Hardhat Wearing on Construction Sites ». Journal of Computing in Civil Engineering, vol. 33, nº 2.

Nath, N.D., A.H. Behzadan et S.G. Paal. 2020. « Deep learning for site safety: Real-time detection of personal protective equipment ». Automation in Construction, vol. 112.

Naoum, Shamil George. 2016. « Factors influencing labor productivity on construction sites: A state-of-the-art literature review and a survey ». International Journal of Productivity and Performance Management, vol. 65, nº 3, p. 401-421.

Pan, Z., C. Su, Y. Deng et J. Cheng. 2022. « Image2Triplets: A computer vision-based explicit relationship extraction framework for updating construction activity knowledge graphs ». Computers in Industry, vol. 137.

Papageorgiou, C.P., M. Oren et T. Poggio. 1998. « A general framework for object detection ». In IEEE 6th International Conference on Computer Vision. (Bombay, India, 1998), p. 555-562. Narosa Publishing House.

Peng, G., Y. Lei, H. Li, D. Wu, J. Wang et F. Liu. 2021. « CORY-Net: Contrastive Res-YOLOv5 Network for Intelligent Safety Monitoring on Power Grid Construction Sites ». *IEEE Access*, vol. 9, p. 160461-160470.

Piao, Y., W. Xu, T.-K. Wang et J.-H. Chen. 2021. « Dynamic Fall Risk Assessment Framework for Construction Workers Based on Dynamic Bayesian Network and Computer Vision ». Journal of Construction Engineering and Management, vol. 147, nº 12.

Piccardi, M. 2004. « Background subtraction techniques: a review ». In 2004 IEEE International Conference on Systems, Man and Cybernetics. (The Hague, Netherlands, 2004), p. 3099-3104. IEEE.

Pour Rahimian, Farzad, Saleh Seyedzadeh, Stephen Oliver, Sergio Rodriguez et Nashwan Dawood. 2020. « On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning ». Automation in Construction, vol. 110, p. 103012.

Ranaweera, K., J. Ruwanpura et S. Fernando. 2013. « Automated real-time monitoring system to measure shift production of tunnel construction projects ». Journal of Computing in Civil Engineering, vol. 27, nº 1, p. 68-77.

Redmon, Joseph, Santosh Divvala, Ross Girshick et Ali Farhadi. 2016. « You Only Look Once: Unified, Real-Time Object Detection ». In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (Las Vegas, NV, USA, juin 2016), p. 779-788. IEEE.

Redmon, Joseph et Ali Farhadi. 2017. « YOLO9000: Better, Faster, Stronger ». In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (Honolulu, HI, juillet 2017), p. 6517-6525. IEEE.

Redmon, Joseph et Ali Farhadi. 2018. « YOLOv3: An Incremental Improvement ». arXiv:1804.02767 [cs].

Ren, Shaoqing, Kaiming He, Ross Girshick et Jian Sun. 2016. « Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks ». arXiv:1506.01497 [cs].

Rezazadeh Azar, E. et B. McCabe. 2012a. « Automated visual recognition of dump trucks in construction videos ». Journal of Computing in Civil Engineering, vol. 26, nº 6, p. 769-781.

Rezazadeh Azar, E. et B. McCabe. 2012b. « Part based model and spatial-temporal reasoning to recognize hydraulic excavators in construction images and videos ». Automation in Construction, vol. 24, p. 194-202.

Roberts, D. et M. Golparvar-Fard. 2019. « End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level ». Automation in Construction, vol. 105.

Roh, Seungjun, Zeeshan Aziz et Feniosky Peña-Mora. 2011. « An object-based 3D walk-through model for interior construction progress monitoring ». Automation in Construction, vol. 20, nº 1, p. 66-75.

Shen, J., X. Xiong, Y. Li, W. He, P. Li et X. Zheng. 2021. « Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning ». Computer-Aided Civil and Infrastructure Engineering, vol. 36, nº 2, p. 180-196.

Son, H. et C. Kim. 2021. « Integrated worker detection and tracking for the safe operation of construction machinery ». Automation in Construction, vol. 126.

Swain M.J. et Ballard D.H. 1991. « Color indexing ». International Journal of Computer Vision, vol. 7, n°1, p.11-32.

Tang, S., D. Roberts et M. Golparvar-Fard. 2020. « Human-object interaction recognition for automatic construction site safety inspection ». Automation in Construction, vol. 120.

Wang, Z., Q. Zhang, B. Yang, T. Wu, K. Lei, B. Zhang et T. Fang. 2021. « Vision-Based Framework for Automatic Progress Monitoring of Precast Walls by Using Surveillance Videos during the Construction Phase ». Journal of Computing in Civil Engineering, vol. 35, nº 1.

Wang, D., X. Wang, B. Ren, J. Wang, T. Zeng, D. Kang et G. Wang. 2022. « Vision-Based Productivity Analysis of Cable Crane Transportation Using Augmented Reality-Based Synthetic Image ». Journal of Computing in Civil Engineering, vol. 36, nº 1.

Wu, J., N. Cai, W. Chen, H. Wang et G. Wang. 2019. « Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset ». Automation in Construction, vol. 106.

Wu, Y., M. Wang, X. Liu, Z. Wang, T. Ma, Z. Lu, D. Liu, Y. Xie, X. Li et X. Wang. 2021. « Monitoring the work cycles of earthmoving excavators in earthmoving projects using UAV remote sensing ». Remote Sensing, vol. 13, nº 19.

Xiao, B. et S.-C. Kang. 2019. « Deep learning detection for real-time construction machine checking ». In Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019. (2019), p. 1136-1141.

Xiong, R., Y. Song, H. Li et Y. Wang. 2019. « Onsite video mining for construction hazards identification with visual relationships ». Advanced Engineering Informatics, vol. 42.

Zeng, T., J. Wang, B. Cui, X. Wang, D. Wang et Y. Zhang. 2021. « The equipment detection and localization of large-scale construction jobsite by far-field construction surveillance video based on improving YOLOv3 and grey wolf optimizer improving extreme learning machine ». Construction and Building Materials, vol. 291.

Zhang, Bingfei, Zhenhua Zhu, Amin Hammad et Walid Aly. 2018. « Automatic matching of construction onsite resources under camera views ». Automation in Construction, vol. 91, p. 206-215.

Zhang, M., Z. Cao, Z. Yang et X. Zhao. 2020. « Utilizing Computer Vision and Fuzzy Inference to Evaluate Level of Collision Safety for Workers and Equipment in a Dynamic Environment ». Journal of Construction Engineering and Management, vol. 146, nº 6.

Zhu, Zhenhua, Xiaoning Ren et Zhi Chen. 2017. « Integrated detection and tracking of workforce and equipment from construction jobsite videos ». Automation in Construction, vol. 81, p. 161-171.

Zhu, D., H. Wen et Y. Deng. 2020. « Pro-active warning system for the crossroads at construction sites based on computer vision ». Engineering, Construction and Architectural Management, vol. 27, nº 5, p. 1145-1168.

Zou, Zhengxia, Zhenwei Shi, Yuhong Guo et Jieping Ye. 2019. « Object Detection in 20 Years: A Survey ». arXiv:1905.05055 [cs].

9. APPENDIX

Table : Object-detection techniques and detected object for on-site construction processes

Groups	*Methods*	*Detected objects*	*References*
		Safety monitoring	
Static hazards	Faster R-CNN	Person,PPE	(Fan *et al.*, 2020)
	SSD	PPE	(Wu *et al.*, 2019)
	YOLOv5	Person,PPE,Const equipment	(Peng *et al.*, 2021)

	YOLOv3	Person, PPE	(Nath, Behzadan et Paal, 2020)
	Faster R-CNN	Person	(Fang *et al.*, 2018a)
	Custom DL model	Person, PPE	(Shen *et al.*, 2021)
	YOLOv3	PPE	(Delhi, Sankarlal et Thomas, 2020)
	Faster R-CNN	Person, PPE	(Li *et al.*, 2021a)
	YOLOv4	Person	(Son et Kim, 2021)
	Mask R-CNN	Person, Building element	(Fang *et al.*, 2019)
	Custom DL model	Person	(Golcarenarenji *et al.*, 2021)
	CenterNet	Person	(Goh, Tian et Chian, 2022)
	YOLOv5	PPE	(Li *et al.*, 2022a)
	YOLOv3	PPE	(Chen et Demachi, 2021)
	Traditional	Person, PPE	(Mneymneh, Abbas et Khoury, 2019)
	YOLOv4	Person, Const equipment	(Kim, Kim et Shchur, 2021)
	Faster R-CNN	Person, PPE	(Fang *et al.*, 2018b)
Dynamic hazards	Faster R-CNN	Person, PPE, Const vehicle	(Piao *et al.*, 2021)
	Traditional	Person, Const vehicle	(Hu *et al.*, 2020)
	Mask R-CNN	Person, PPE, Const vehicle	(Fang *et al.*, 2020)
	SSD	Person, PPE, Const vehicle	(Xiong *et al.*, 2019)
	Traditional	Person, Const vehicle, vehicle	(Zhu, Wen et Deng, 2020)
	YOLOv3	Person, Const vehicle	(Kim, Lee et Kamat, 2020)
	YOLOv3	Const vehicle	(Meng *et al.*, 2020)
	YOLOv2	Person, Const vehicle	(Luo *et al.*, 2020)
	YOLOv3	Const vehicle	(Zeng *et al.*, 2021)
	Faster R-CNN	Person, Const vehicle	(Zhang *et al.*, 2020)
	Traditional	Person, Const vehicle	(Kim, Kim et Kim, 2016)
	YOLACT	Person, PPE, Const vehicle	(Kang *et al.*, 2022)
		Progress monitoring	
Interior construction	Traditional	Material	(Roh, Aziz et Peña-Mora, 2011)
	Traditional	Material	(Deng *et al.*, 2020)
	Traditional	Material	(Hamledari, McCabe et Davari, 2017)
Exterior construction	Traditional	Building element	(Ibrahim *et al.*, 2009)
	Mask R-CNN	Building element	(Wang *et al.*, 2021)
	SSD	Const vehicle, Person, Const equipment	(Zhang *et al.*, 2018)

	Custom DL model	Person, Building element	(Pour Rahimian *et al.*, 2020)
	Traditional	Person, Const vehicle	(Zhu, Ren et Chen, 2017)
	YOLOv3-SPP	Const vehicle, Const equipment	(Hevesi *et al.*, 2021)
		Productivity monitoring	
Human productivity	YOLOv2	Person, Tool, Material, Const equipment	(Pan *et al.*, 2022)
	Faster R-CNN	Person	(Li *et al.*, 2022b)
	CenterNet	Person, Building element	(Li *et al.*, 2021b)
	Faster R-CNN	Person,Const equipment,Const vehicle,Material	(Luo *et al.*, 2018)
Equipment/vehicle productivity	Faster R-CNN	Const equipment	(Wang *et al.*, 2022)
	SSD	Const vehicle	(Wu *et al.*, 2021)
	RetinaNet	Const vehicle	(Roberts et Golparvar-Fard, 2019)
	R-FCN	License plate	(Kim *et al.*, 2019)
	YOLOv3	Const vehicle	(Xiao et Kang, 2019)
	R-FCN	Const vehicle	(Kim *et al.*, 2018)
	Traditional	Const vehicle	(Bügler *et al.*, 2014)
	Traditional	Const vehicle	(Azar et McCabe, 2013)
	Traditional	Material	(Ranaweera, Ruwanpura et Fernando, 2013)
	Traditional	Const vehicle	(Rezazadeh Azar et McCabe, 2012a)
	Traditional	Const vehicle	(Rezazadeh Azar et McCabe, 2012b)
	Traditional	Const equipment	(Gong et Caldas, 2010)
		Quality monitoring	
	Mask R-CNN	Material	(Kardovskyi et Moon, 2021)
	Traditional	Material	(Lin et Fang, 2011)