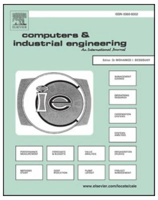
Contents lists available at ScienceDirect

Computers & Industrial Engineering

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



Management of safe distancing on construction sites during COVID-19: A smart real-time monitoring system

Yang Miang Goh [a,*], Jing Tian [b], Eugene Yan Tao Chian [a]

[a] Department of the Built Environment, School of Design & Environment, National University of Singapore, 4 Architecture Drive, Singapore 117566, Singapore
[b] Institute of Systems Science, National University of Singapore, 25 Heng Mui Keng Terrace, Singapore 119615, Singapore

ARTICLE INFO

Keywords:
COVID-19
Construction safety
Computer vision
Video surveillance
Safe distancing

ABSTRACT

The outbreak of Coronavirus Disease 2019 (COVID-19) poses a great threat to the world. One mandatory and efficient measure to prevent the spread of COVID-19 on construction sites is to ensure safe distancing during workers' daily activities. However, manual monitoring of safe distancing during construction activities can be toilsome and inconsistent. This study proposes a computer vision-based smart monitoring system to automatically detect worker breaching safe distancing rules. Our proposed system consists of three main modules: (1) worker detection module using CenterNet; (2) proximity determination module using Homography; and (3) warning alert and data collection module. To evaluate the system, it was implemented in a construction site as a case study. This study has two key contributions: (1) it is demonstrated that monitoring of safe distancing can be automated using our approach; and (2) CenterNet, an anchorless detection model, outperforms current state-of-the-art approaches in the real-time detection of workers.

1. Introduction

Pandemic such as the prevailing Coronavirus Disease 2019 (COVID-19) poses a significant threat to public health worldwide (Wiersinga et al., 2020). As of January 22, 2021, a total of 38 million confirmed cases have been reported around the world with 70.51 million recovered and 2.10 million deaths (Johns Hopkins University, 2020). In the face of the pandemic, governments throughout the world mandated or promoted measures such as maintaining safe distance and preventing overcrowding to prevent the spread of COVID-19 (Ministry of Health, 2020).

With the construction industry having one of the highest COVID-19 infection rates during the pandemic (National Statistic, 2020), safe distancing and crowd control, among other measures, need to be dutifully implemented to allow work to continue. To reduce the risk of infection during COVID-19, numerous safety policies and procedures have been established, including reduced physical interaction, and ensuring safe distance of one to 1.5 m between people at all times, including working in outdoor environment (e.g., OSHA, 2020; BCA, 2020). Chu et al. (2020) also emphasized the importance of physical distancing of one metre or more in reducing the risk of human-to-human transmission. In addition, public health authorities consistently advised maintaining a safe distance between people, including outdoor activities (e.g., CDC, 2021; Ministry of Health, 2021). However, people may unconsciously violate the mandatory safe distancing and overcrowding rules. Furthermore, traditional human supervision of safety measures and behaviour observation during construction can be toilsome and inconsistent. As a result, an automatic real-time monitoring system to monitor safe distancing on construction site is needed. Even though there had been many computer vision-based monitoring systems developed in the past, this study is the first to propose an automatic real-time monitoring system to monitor safe distancing on construction sites.

Non-visual sensors such as radio frequency identification (RFID) tags, and global positioning system (GPS) sensors can be used to automatically track workers' locations in real-time. However, sensor-based approach requires workers to wear sensors at all time and this is difficult to implement. Furthermore, many sub-contractors work across multiple sites, and they are frequently short-term workers, hence it may not be feasible to issue a sensor to every worker. In contrast, computer vision approaches alleviate some of these issues and have been used for automatic safety inspection and unsafe behavior recognition in the construction industry (e.g., Chian, Fang, Goh, & Tian, 2021; Fang, Love, Luo, & Ding, 2020; Fang, Ding, et al., 2020; Luo, Li, Wang, et al., 2019; Luo, Li, Yang, Yu, & Cao, 2019; Yu, Guo, Ding, Li, & Skitmore, 2017; Seo, Han, Lee, & Kim, 2015). For example, Ding et al. (2018) developed

* Corresponding author.
E-mail address: bdggym@nus.edu.sg (Y.M. Goh).

https://doi.org/10.1016/j.cie.2021.107847
Received 19 May 2021; Received in revised form 9 November 2021; Accepted 24 November 2021
Available online 7 December 2021

a hybrid deep learning system by integrating conventional neural networks (CNN) and long short-term memory (LSTM) to detect people's unsafe behaviour (e.g., abnormal climb) from videos.

With this in mind, a computer vision-based real-time monitoring system is proposed to automatically detect workers who violate safe distancing rules and send warning alerts to relevant site personnel for necessary actions. In addition, statistics generated by the system can be used to facilitate behaviour-based safety (BBS) management (Goh, Ubeynarayana, Wong, & Guo, 2018; Fang, Love, et al., 2020; Fang, Ding, et al., 2020). However, a limited amount of research has been undertaken that has applied computer vision to examine the social distance for prevention of COVID-19 transmission (Yang, Yurtsever, Renganathan, Redmill, & Özgüner, 2020; Ahmed, Ahmad, Rodrigues, Jeon, & Din, 2021). The most important studies that have been undertaken to identify social distance violation are Yang et al. (2020) and Ahmed et al. (2021). In Yang et al. (2020)'s work, Faster R-CNN and YOLOv4-based pedestrian detection method is proposed to measure social distancing. The effectiveness of the proposed method was tested in publica database (e.g., Oxford town center database, mall dataset, and train station dataset). In Ahmed et al. (2021)'s work, a computer vision with YOLOv3 model is proposed to identify social distance violation in indoor environment.

Despite the success of the work that has been undertaken to identify safety distancing violation in public database, there is no research focus on in construction industry. The construction industry poses many challenges to computer vision-based systems including varying sizes of objects, and cluttered background compared with public database. These challenges limit the performance of current anchor-based object detection approaches (e.g., Faster R-CNN, YoLov3) in construction research. To achieve higher level of accuracy, an anchorless model, CenterNet, is employed to detect workers in video surveillance for the real-time detection of safety distancing violations. To the best of our knowledge, this is the first time that CenterNet had been used in construction-related research. To validate our developed system's feasibility and effectiveness, a construction project in Singapore is used as a case study.

Our paper commences by providing a review of computer vision technologies in construction sites (Section 2). Then, it describes our developed computer vision-based real-time safe distancing monitoring system (Section 3). Next, a case study is used to validate our developed system (Section 4). The contributions, limitations, and conclusions are discussed subsequently.

2. Related works

2.1. Deep learning-based object detection in construction sites

A plethora of deep learning-based computer vision approaches have been used and developed to detect objects (e.g., people, plants, and equipment) in construction sites (Fang, Ding, Luo, & Love, 2018; Fang, Ding, Zhong, Love, & Luo, 2018; Fang, Li, et al., 2018; Fang, Love, et al., 2020; Fang, Ding, et al., 2020; Guo, Zou, Fang, Goh, & Zou, 2021). Fang, Ding, Luo, et al. (2018), Fang, Ding, Zhong, et al. (2018), Fang, Li, et al. (2018) employed an improved Faster R-CNN model to detect people and heavy equipment in construction sites. In their work, the deep learning model achieved higher accuracy than approaches that rely on hand-crafted features (e.g., histogram of oriented gradient (Dalal & Triggs, 2005), scale-invariant feature transform (Lowe, 2004)) in the detection of objects on images. Table 1 presents examples of prior works on deep learning-based object detection in construction sites.

Table 1 demonstrates that computer vision and deep learning technologies have been successfully used to detect construction objects. However, it should be acknowledged that the dynamic and complex nature of construction sites (e.g., cluttered background, varying size of objects, occlusion, and variation in human pose) affects the detection accuracy of deep learning models (Fang, Ding, Luo, et al., 2018; Fang, Ding, Zhong, et al., 2018; Fang, Li, et al., 2018; Son, Seong, Choi, & Kim, 2019). These prior works are based on anchor-based object detection architectures that make use of anchor boxes to initiate the detection process. The image to be processed is reduced in size and tiled across with anchor boxes. The anchor box nearest to the target object in the image becomes the detected object bounding box. These anchor-based object detection approaches can be further categorized into two types: (1) one-stage detector (e.g., Yolov2); and (2) two-stage detector (e.g., Faster R-CNN).

Table 1
Prior works on deep learning and computer vision-based object detection in construction sites.

Author (Year)	Algorithms		Target objects
Nath, Behzadan, and Paal (2020)	One-stage	YOLOv3	Personal protective equipment (e.g., hard hat and vest) and people
Luo, Liu, et al. (2020), Luo, Wang, et al. (2020)		YOLOv2	People and excavator
Wu, Cai, Chen, Wang, and Wang (2019)		Single Shot MultiBox Detector (SSD)	Hardhat and people
Luo, Liu, et al. (2020), Luo, Wang, et al. (2020)	Two-stage	Stacked Hourglass Network (HG), Cascaded Pyramid Network (CPN), ensemble model of HG and CPN	Excavators, trucks, cranes, and bulldozers)
Fang et al. (2019)		Mask R-CNN	People and structural support
Fang, Ding, Luo, et al. (2018), Fang, Ding, Zhong, et al. (2018), Fang, Li, et al. (2018)		Faster R-CNN	People and excavator
Fang, Ding, Luo, et al. (2018), Fang, Ding, Zhong, et al. (2018), Fang, Li, et al. (2018)		Faster R-CNN	People and hardhat

Two-stage detectors extract feature maps for each possible bounding box and then classify and regress those extracted features (Li, Peng, Yu, Zhang, Deng, & Sun, 2017). Compared with a two-stage detector, the architecture of one-stage detector is simpler and has faster detection speed. One-stage detectors slide a complex arrangement of possible bounding boxes and then regresses them directly (Tian et al., 2019; Bochkovskiy, Wang, & Liao, 2020). The accuracy of anchor-based detection methods can be increased by adjusting the anchor boxes' parameters (e.g., size and aspect ratio). However, these anchor-based approaches need a large number of anchors to ensure a sufficiently high Intersection-over-Union (IOU) with the ground-truth and anchor boxes' parameters (e.g., size and aspect ratio) are typically manually designed. In doing so, such anchor-based approaches may not be useful for multi-scale object detection, as it is difficult to specify suitable values for all possible anchor boxes' parameters (Duan et al., 2019).

CenterNet, an anchorless and key point-based object detection approach developed by Zhou, Wang, and Krähenbühl (2019), can be used to address the drawbacks of anchor-based approaches. The CenterNet outperformed current state-of-the-art object detection approaches when tested on the MS COCO database[1]. To achieve higher accuracy on the detection of people in construction sites, CenterNet is employed to detect people in this study. To our knowledge, this study is the first time that CenterNet had been applied in construction-related research.

[1] COCO is a large-scale object detection, segmentation, and captioning dataset. https://cocodataset.org/#home

2.2. Vision-based distance measurement

Several approaches have been used to determine the distance between objects in images and videos. Initially, most research utilized a single vision camera to measure the distance. For example, Rahman et al. (2009) developed a system to measure the distance between people and camera from a single camera. Similarly, Wahab, Sivadev, and Sundaraj (2011) developed a system with Hough transforms to determine the distance between objects. Likewise, Kim, Kim, and Kim (2017) applied a homography matrix to warp a camera view to another camera view whereby the reference pixel distance to a real distance ratio is known, and hence it is able to estimate the distance between objects in an image. Despite their success on distance measurement, it is unable to achieve good performance as it loses depth information from a single camera.

To address this problem, some research used stereo cameras to determine the distance between objects. For example, Mustafah, Noor, Hasbi, and Azma (2012) developed a stereo system to accurately measure the distance and its size of object in images. Despite the success of this approach, the accuracy is sensitive to binocular baseline and image resolution. In addition, it needs more computing resources to measure the distance from the stereo camera. Other studies focus on using depth camera (e.g., Kinect) to measure distance between objects. However, the depth camera has a limited range and is sensitive to lighting condition, it is not able to be used in outdoor construction sites. It is noted that most construction sites do not currently install stereo and depth cameras. To facilitate implementation, researchers frequently have to determine ways to estimate distances based on single cameras.

2.3. Computer vision-based proximity warning system for construction sites

High-resolution video cameras, increased storage capacity of databases, and high throughput of the internet have increased the capacity to document operations in the construction industry (Szeliski, 2010; LeCun, Bengio, & Hinton, 2015). Computer vision-based approaches have tapped on this increased capacity in recent years to automatically perform supervision tasks, such as workers' unsafe behavior identification (Fang, Ding, Luo, et al., 2018; Fang, Ding, Zhong, et al., 2018; Fang, Li, et al., 2018; Ding et al., 2018; Fang et al., 2019), workers' activities recognition (Luo et al., 2018), and object detection (Fang, Ding, Luo, et al., 2018; Fang, Ding, Zhong, et al., 2018; Fang, Li, et al., 2018).

Studies have used computer vision to detect people's proximity to dangerous objects to identify hazardous situations (Kim, Kim, & Kim, 2016; Son et al., 2019; Luo, Liu, et al., 2020; Luo, Wang, Wong, & Cheng, 2020). For example, Luo, Liu, et al. (2020), Luo, Wang, et al. (2020) developed a computer vision system that alert supervisors if workers enter an excavator's working radius when the excavator is being operated. In their work, a Yolov2 model was employed to detect people and excavators from videos, and a transformation matrix was applied to calculate the distance between people and excavators to determine unsafe behaviour. Table 2 presents some prior works on computer vision-based proximity warning systems in construction sites. As can be seen in Table 2, a computer vision-based proximity system can be used to identify hazards (e.g., stuck-by) and improve safety in construction. To the best of our knowledge, there is no existing research that used computer vision to automatically identify people breaching safe distancing rules for prevention of disease transmission in construction.

It should be noted that the performance (e.g., accuracy and speed) of these computer vision systems depends on the accuracy of object detection and distance measurement. For example, Luo, Liu, et al. (2020), Luo, Wang, et al. (2020) noted that a cluttered construction site could hinder the model's ability to recognize people. In Luo, Liu, et al. (2020), Luo, Wang, et al. (2020), small worker could not be accurately detected by the Yolov2 model. Existing computer vision-based proximity detection systems have performance issues when applied in practical applications with dynamic and changing environment and objects of varying sizes. CCTV cameras are frequently installed on top of tower cranes to have a bird eye's view of the construction site and construction floor. However, as the tower crane gets jacked up, the distance between the construction floor and camera can increase, resulting in decreased sizes of worker images, as shown in Fig. 1. Therefore, developing a computer vision system to identify workers breaching safe distancing rules accurately remains a challenge for the construction industry.

Table 2
Prior works on computer vision-based proximity warning system.

Author (Year)	Descriptions	Algorithms/ methods	Limitation
Luo, Liu, et al. (2020), Luo, Wang, et al. (2020)	A smart video surveillance system with Yolov2 is proposed to real-time detect people entering excavator's working area.	Yolov2 and transformation matrix	The proposed system is not able to detect small worker images in video.
Kim, Liu, Lee, and Kamat (2019)	An unmanned aerial vehicle (UAV) system with Yolov3 was developed to prevent people from being stuck by plants in construction sites.	Yolov3	The developed UAV system is not able to: 1) detect hazards in real-time; and 2) the plant's status.
Son et al. (2019)	A computer vision system is developed for detection of collisions between people and equipment in construction sites.	Faster R-CNN and Homography matrix	The developed computer vision system has a limited field of view for detection of collisions.
Kim et al. (2016)	Determination of safety levels on-site based on detected accidents in construction sites	Gaussian mixture model (GMM)	The developed approach has the limitations that 1) the plant's status was not considered; and 2) the actual accuracy did not meet the desired requirement because of the dynamic nature of construction sites (cluttered background, and occlusion).

3. Research approach

A design science research approach is adopted to develop a real-time computer vision-based smart monitoring system that can automatically detect workers breaching safe distancing and overcrowding rules during COVID-19. Design science is a research approach that describes and predicts the current natural or social world by understanding problems and designing solutions to improve human performance (van Aken, 2005, Geerts, 2011). Therefore, design science approach can be used to design and implement the proposed smart monitoring system. The research process used to develop the real-time computer vision-based smart monitoring system for detecting people breaching safe distancing rules during COVID-19 is presented in Fig. 2.

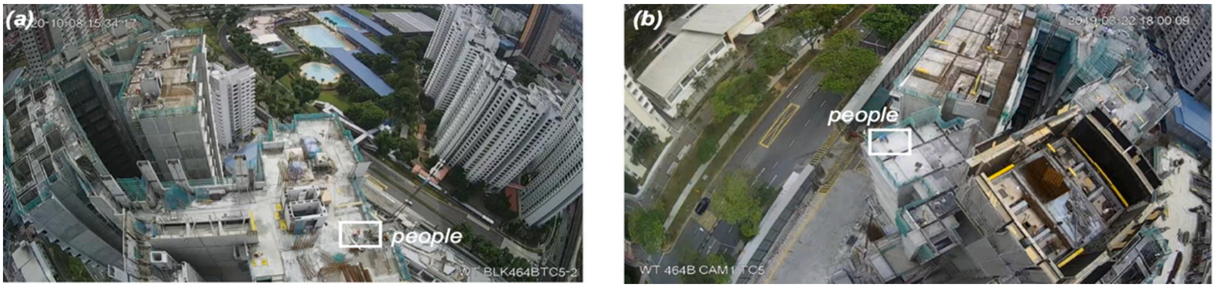


Fig. 1. Examples of CCTV camera footage installed on tower crane in construction site.

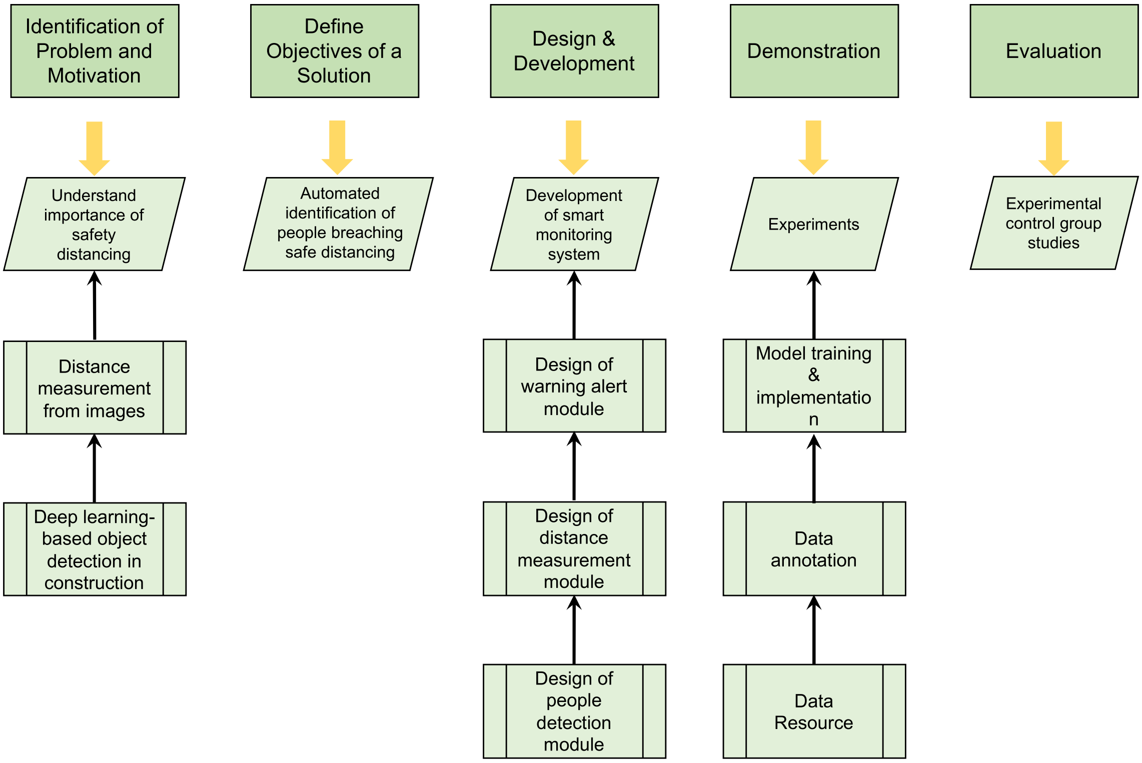

Identification of Problem and Motivation
Define Objectives of a Solution
Design & Development
Demonstration
Evaluation
Understand importance of safety distancing
Automated identification of people breaching safe distancing
Development of smart monitoring system
Experiments
Experimental control group studies
Distance measurement from images
Design of warning alert module
Model training & implementatio n
Deep learning-based object detection in construction
Design of distance measurement module
Data annotation
Design of people detection module
Data Resource

Fig. 2. Workflow of design science approach (adapted from Chu, Matthews, and Love (2018) and Luo et al. (2018)).

3.1. Design development of smart monitoring system

3.1.1. People detection module

Compared with anchor-based approaches (e.g., Faster R-CNN), anchorless approach has fewer number of anchor parameters and does not need non-maximum suppression (NMS)[2]. In this case, the models developed using the anchorless approach is more generalizable and have faster detection speed. Based on the comparison of performances of different object detection methods, the CenterNet object detection approach is selected to detect construction workers due to its better performance. As mentioned, CenterNet is an anchor-free and key point-based detection model. It detects an object first as a center point and then regresses the object bounding box's height and width with respect to the center point.

In this research, Deep Layer Aggregation-34 (DLA-34) with hierarchical skip connections are used as the backbone network. Following Zhou et al. (2019), the fully convolutional upsampling version of the DLA network used for dense prediction is employed. Then, the dense prediction with iterative aggregation is used to increase the resolution of extracted feature maps. Following Zhou et al. (2019), the original convolution of DLA-34 is replaced with 3 × 3 deformable convolutions at each upsampling layer, a 3 × 3 convolutional layer with 256 channels is added before each output head, and a 1 × 1 convolution is added. More details on CenterNet can be found in Zhou et al. (2019). The structure of CenterNet is presented in Fig. 3.

In this approach, we assume that I is the input image, and $(x_1^{(k)}, y_1^{(k)}, x_2^{(k)}, y_2^{(k)}$ is the coordinate of bounding box of object k with category c_k. As presented in Fig. 3, we firstly input images (I) to DLA-34 network for extraction of feature maps. Then, our model outputs a heatmap (see Fig. 4(a)) containing the center points of the objects as per Eq. [1]. Examples of bounding boxes are presented in Fig. 4(b).

$$\widehat{Y} \in [0,1]^{\frac{W}{R}\times\frac{H}{R}\times c} \tag{1}$$

where, R is the output stride, C is the number of key point types, W is the width of input image, and H is the height of image. In this research, we followed Cao, Simon, Wei, and Sheikh (2017) where the output stride, R, is set to 4.

The peaks in the generated heatmap correspond to the center of the

[2] NMS is a technique used in computer vision algorithms to select one bounding box out of many overlapping bounding boxes according to intersection over union (IOU).

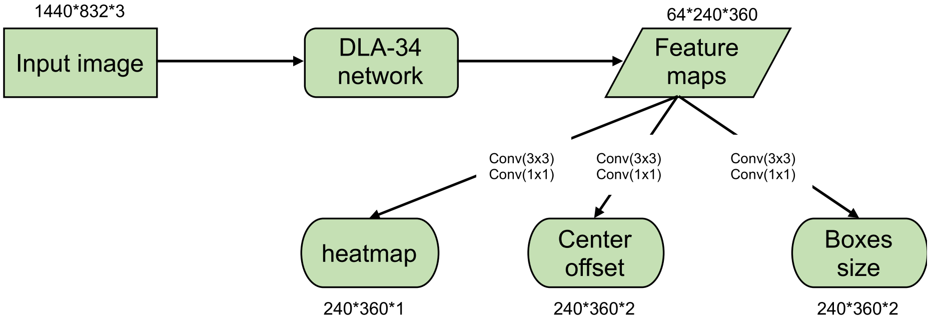


Fig. 3. Structure of CenterNet.

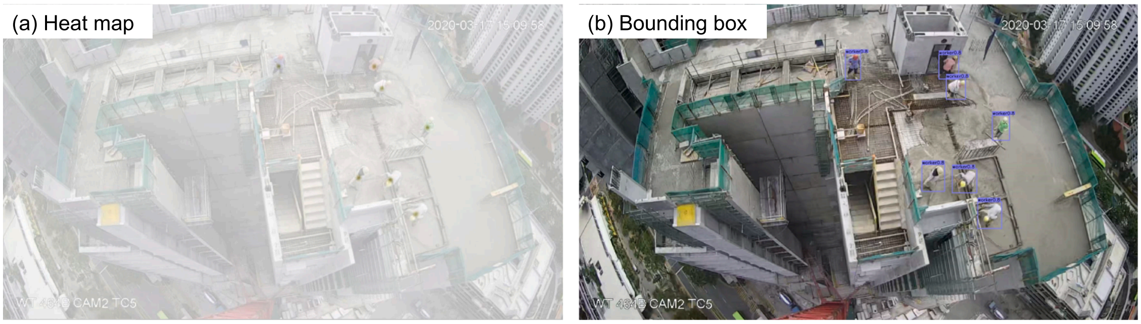


Fig. 4. Examples of output from model: (a) heatmap; (b) bounding boxes.

object, which can be used to predict the width and height of the object's bounding box (Fig. 4(b)). To extract the peak of each heatmap, all responses with value higher than or equal to its 8-connected neighbors are detected and the top 100 peaks are retained.

To address the issue of discretization error caused by the output stride, a local offset for each center point is predicted.

Let $\widehat{P}_c$ be the set of n detected center points $\widehat{P} = \left\{ \left(\widehat{x}_i, \widehat{y}_i\right) \right\}_{i=1}^{n}$ of class c. We use the key points value $\widehat{Y}_{x_i y_i c}$ as a measure of its detection confidence, and produce a bounding box at a location as per Eq. [2]:

$$\begin{bmatrix} \widehat{x}_i + \delta\widehat{x}_i - \widehat{w}_i/2 & \widehat{y}_i + \delta\widehat{y}_i - \widehat{h}_i/2 \\ \widehat{x}_i + \delta\widehat{x}_i + \widehat{w}_i/2 & \widehat{y}_i + \delta\widehat{y}_i + \widehat{h}_i/2 \end{bmatrix} \tag{2}$$

where, $\left(\delta\widehat{x}_i, \delta\widehat{y}_i\right) = \widehat{O}_{\widehat{x}_i,\widehat{y}_i}$ is the offset prediction, (x_i, y_i) is the integer coordinates of key point, and $\left(\widehat{w}_i, \widehat{h}_i\right) = \widehat{S}_{\widehat{x}_i,\widehat{y}_i}$ is the size prediction.

The overall training loss include three parts: (1) heatmap; (2) center offset; and (3) box size, which is included in Eq. [3]:

$$L_{det} = L_k + \lambda_{size} L_{size} + \lambda_{off} L_{off} \tag{3}$$

where, L_k is loss of heatmap, L_{size} is the loss of boxes size, and L_{off} is center offset. Here, we set $\lambda_{size} = 0.1$ and $\lambda_{off} = 1$ in this research.

(1) ***Heatmap***

The loss of heatmap is based on Eq. [4].

$$L_k = \frac{-1}{N} \sum_{xyc} \begin{cases} \left(1 - \widehat{Y}_{xyc}\right)^{\alpha} log\left(\widehat{Y}_{xyc}\right) & if Y_{xyc} = 1 \\ \left(1 - Y_{xyc}\right)^{\beta} \left(\widehat{Y}_{xyc}\right)^{\beta} & otherwise \\ log\left(1 - \widehat{Y}_{xyc}\right) & otherwise \end{cases} \tag{4}$$

where, α *and* β are hyper-parameters of the focal loss.

(2) ***Center Offset***

The loss of center offset L_{off} is noted in Eq. [5]

$$L_{off} = \frac{1}{N} \sum_{p} \left| \widehat{O}_p - \left(\frac{p}{R} - p \right) \right| \tag{5}$$

(3) ***Box Size***

The loss of bounding box size is noted in Eq. [6]

$$L_{size} = \frac{1}{N} \sum_{k=1}^{N} \left| \widehat{S}_{pk} - s_k \right| \tag{6}$$

Here, we assume that the center point is $p_k = \left(\frac{x_1^{(k)} + x_2^{(k)}}{2}, \frac{y_1^{(k)} + y_2^{(k)}}{2} \right)$. Key point estimator $\widehat{Y}$ predicts all center points. Furthermore, the object size $s_k = \left(x_2^{(k)} - x_1^{(k)}, y_2^{(k)} - y_1^{(k)} \right)$ for each object k is regressed.

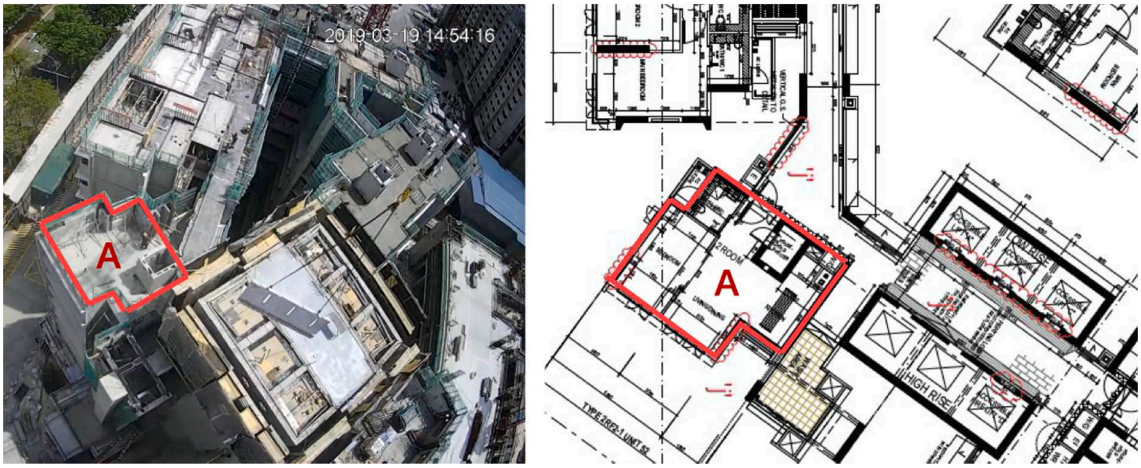
To reduce the computational burden, a single size $\widehat{S} \in R^{\frac{W}{R} \times \frac{H}{R} \times 2}$ is employed for the prediction of people in this study.

3.1.2. *Distance measurement module*

To measure the distance among people captured in a video footage, a Homography approach (Belongie & Kriegman, 2007) is used to project the plane of interest (e.g., construction floor) in the camera view to the corresponding construction floor plan (Fig. 5). Euclidean Distance Matrix (Dokmanic, Parhizkar, Ranieri, & Vetterli, 2015) is then used to determine the distance among people on construction floor plan (Fig. 5). These two approaches are briefly introduced as follows.

(1) Project plane of area of interest to construction floor plan

Homography is a transformation that is used to project a plane of interest (e.g., construction floor) in the camera view to construction floor plan, as shown in Fig. 5. Given a plane of interest in camera view (Fig. 5(a)) and the construction floor plan (Fig. 5(b)), the homography approach defines a 3 × 3 matrix that can be used to warp the same plane between these two images as noted in Eq. [7]:



(a) CCTV camera view

(b) Construction floor plan

Fig. 5. Calibration between camera image plane (left image) and Construction floor plan (right image).

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \tag{7}$$

where, p and q are coordinates of a point of a plane of interest in camera view and x and y are the corresponding coordinates of the point on the construction floor plan. h_{ij} ($i = 1,2,3$; $j = 1,2,3$) are transformation coefficients, which can be calculated by using *Direct Linear Transformation* (Shapiro, 1978) using OpenCV Python Package in a one-time offline calibration step by specifying at least four points of the plane of interest in camera view and four corresponding points on the construction floor plan.

(2) Distance measurement

In this study, the bottom center point of a bounding box is used to represent the position of worker in image plane. After projecting the image plane to the construction floor plan, the proximity among people can be estimated by calculating the Euclidean distance on construction floor plan. The scale bar on the construction floor plan is used as a reference, and the pixel distance in the image is transformed to the meter metric unit. The distance between people can be calculated based on Eq. [8].

$$proximity_{meter} = proximity_{CAD} \times S_{scale\_bar} \tag{8}$$

where, $proximity_{CAD}$ is the distance of objects in construction floor plan.

3.1.3. *Warning alert generation and data collection module*

If the distance between construction workers is less than a predetermined threshold ($L_{\min}$), i.e., safe distancing rule had been breached, a warning alert will be generated and sent to site personnel using Telegram. Similarly, if the number of workers in a zone exceeds a pre-determined threshold (D_{max}), i.e., overcrowding occurs, a Telegram alert can be sent to the relevant site personnel. The site personnel to receive the Telegram can be site managers, supervisors, and workers. It is also possible for onsite alarms to be triggered when either the proximity or density thresholds are exceeded. Furthermore, statistics about violations of safe distancing and overcrowding rules can be collected to help managers assess the effectiveness of their BBS and COVID-19 interventions.

4. Experimental control group studies

4.1. *Data resource*

In our study, a public housing construction site in Singapore was used as a case study to evaluate the feasibility and effectiveness of our proposed computer vision system. The construction contract requires CCTV cameras to be installed on the tower cranes for safety and security reasons. Hence, this study makes use of the existing system and video data. Accordingly, privacy issues related to recording human work activities (Rashwan, Solanas, Puig, & Martínez-Ballesté, 2015) are less of a concern. In addition, the work area being monitored is at the construction level, so the likelihood of contravening the workers' privacy is also minimized.

Fifteen cameras were mounted on seven tower cranes for daily monitoring of the construction site. It is important to note that we made use of existing CCTV cameras installed on tower cranes and there was no need for the contractor to install any new cameras on site, which made it easier to implement. In this case study, we focus on two CCTV cameras covering one construction block. The monitoring of overcrowding is essentially a subset of safe distancing monitoring because it is based on the detection of workers and is more easily implemented. Thus, in this study, we are focusing on the safe distancing function.

As construction progress and the tower crane is jacked up, the distance between the CCTV camera and construction level changes. Based on the samples collected in Table 3, the approximate distance between the camera and the construction floor varies between 6.6 m and 17.8 m (see Table 3 and Fig. 6).

In our study, the image dataset created for model training is collected from the site's CCTV system. We firstly extracted images from the CCTV video footage by collecting 36 images from every eight consecutive days of each month (1–8, 9–16, 17–24, and 25–31). The 36 images were collected randomly from three periods: (1) 8:00am to 11:00am, (2) 11:00am to 16:00 pm and (3) 16:00 pm to 19:00 pm, while ensuring that each period has 12 images. The resolution of each image is 1080 × 1920. A database of 5616 CCTV images, which contain 16,285 worker images, was created. The images are divided into training set and testing set with a ratio of 8:2. In other words, 4492 CCTV images were used for training and 1124 images were used for testing.

Due to the varying distances between the workers on the construction levels and the CCTV cameras, the pixel size of people in the images varies significantly. With reference to Fig. 7, and in accordance to COCO database's definition of small, medium, and large images (Lin et al., 2014), we find that most of the worker images in our database falls in the

Table 3
Distance between camera and construction level during January 2019-September 2019.

Month	Jan	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.
Distance/m	10	7.2	4.4	16.8	11.2	8.4	17.8	12.2	6.6

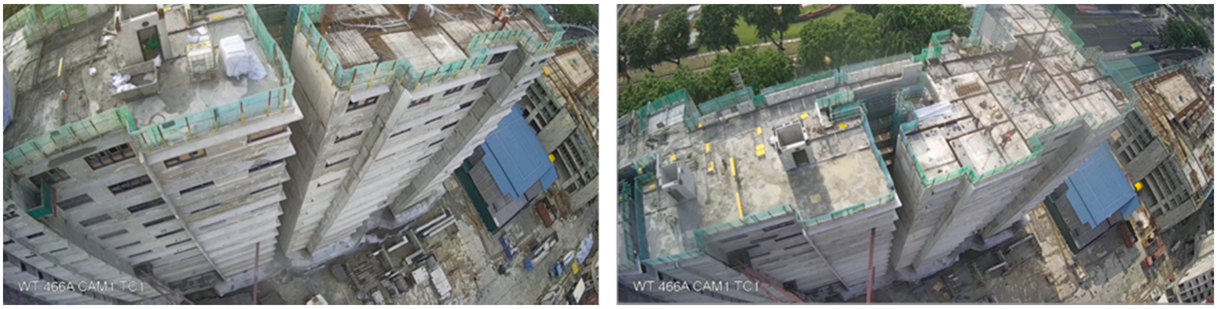


Fig. 6. Snapshot of CCTV camera system.

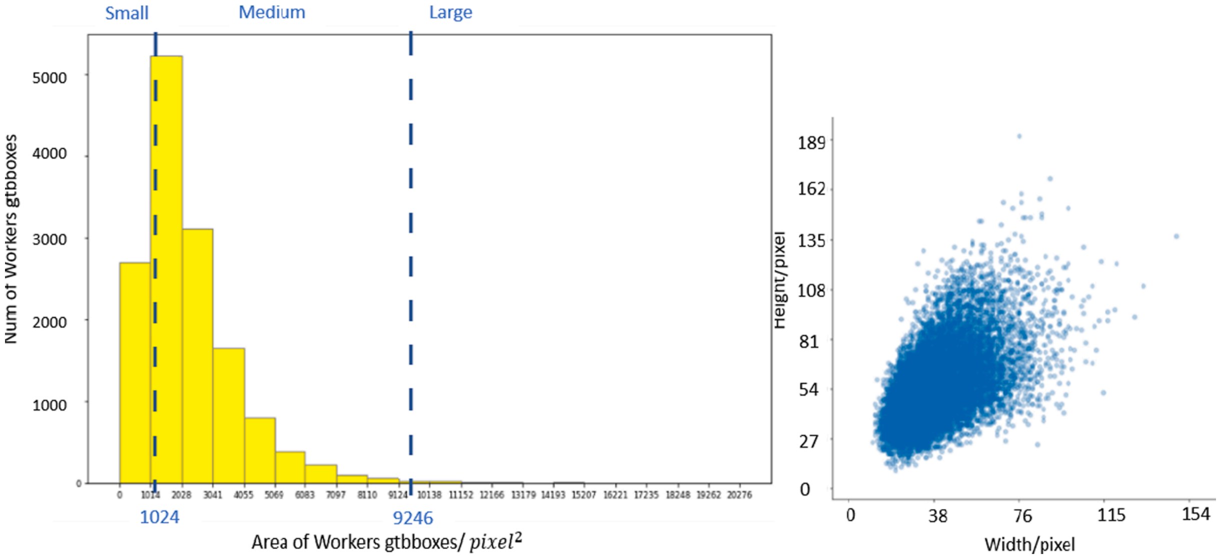


Fig. 7. Pixel size distribution of database.

range of small and medium size.

4.2. Data annotation

The following exclusion rules are defined in this study to improve consistency of labelling of worker images:

1. Worker occluded by objects (e.g., heavy equipment and temporary structure) will not be labelled (see Fig. 8(a));
2. Worker in blurred images will not be labelled (see Fig. 8(b)); and
3. Worker with very small pixel size will not be labelled (see Fig. 8(c)).

In this study, the annotation tool, 'LabelImg', is used. LabelImg is a graphical image annotation tool written in Python. Each bounding box is a rectangular box denoted by the top left corner point and bottom right corner point, enclosing the object of interest. Fig. 9 shows an example of the creation of annotated worker bounding boxes on the original image using LabellImg.

5. Experimental results

5.1. Detection of construction workers

Our smart monitoring system is implemented on a server equipped with Intel i7 9th Generation CPU Computer with Nvidia GeForce RTX 2070 graphics card. Following Zhou et al. (2019), a standard dense supervised learning approach (Newell, Huang, & Deng, 2017) is adopted for training. In terms of hyper parameters of the CenterNet model, the batch-size was set at 2 and the learning rate of 5e−5 was used for 140 epochs. In this study, the learning rate dropped 10× at 90 and 120 epochs.

To assess performance, two key performance indicators (KPIs) including Average Precision (AP), and Average Recall (AR) are used to evaluate the CenterNet detection performance. AP measures the area under the precision and recall curve and AR measures recall of detection. Precision and recall are calculated using Eq. [9] and Eq. [10] respectively.

$$precision = \frac{TP}{TP + FP} \tag{9}$$

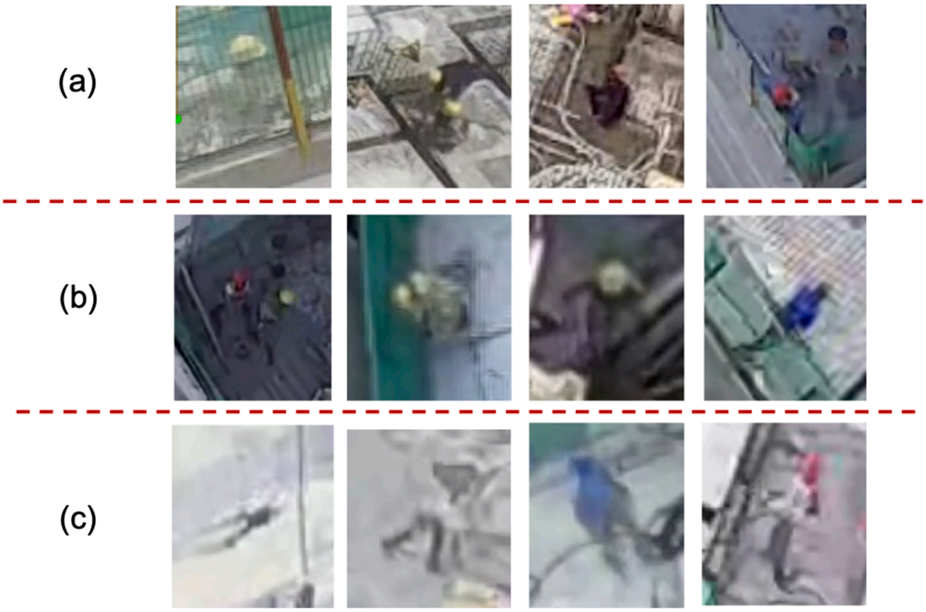
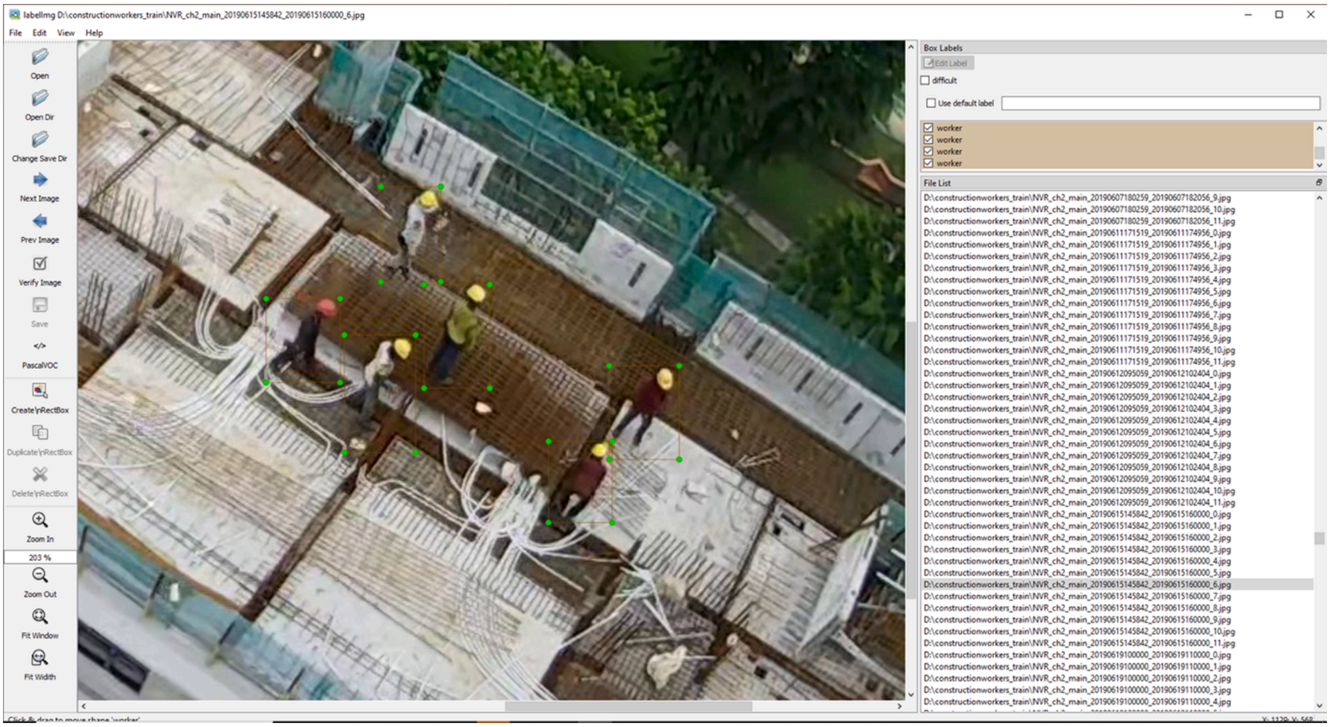


Fig. 8. Examples of images not labelled.



☑ worker
☑ worker
☑ worker
☑ worker

```
D:\constructionworkers_train\NVR_ch2_main_20190607180259_20190607182056_9.jpg
D:\constructionworkers_train\NVR_ch2_main_20190607180259_20190607182056_10.jpg
D:\constructionworkers_train\NVR_ch2_main_20190607180259_20190607182056_11.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_0.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_1.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_2.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_3.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_4.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_5.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_6.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_7.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_8.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_9.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_10.jpg
D:\constructionworkers_train\NVR_ch2_main_20190611171519_20190611174956_11.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_0.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_1.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_2.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_3.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_4.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_5.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_6.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_7.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_8.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_9.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_10.jpg
D:\constructionworkers_train\NVR_ch2_main_20190612095059_20190612102404_11.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_0.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_1.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_2.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_3.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_4.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_5.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_6.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_7.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_8.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_9.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_10.jpg
D:\constructionworkers_train\NVR_ch2_main_20190615145842_20190615160000_11.jpg
D:\constructionworkers_train\NVR_ch2_main_20190619100000_20190619110000_0.jpg
D:\constructionworkers_train\NVR_ch2_main_20190619100000_20190619110000_1.jpg
D:\constructionworkers_train\NVR_ch2_main_20190619100000_20190619110000_2.jpg
D:\constructionworkers_train\NVR_ch2_main_20190619100000_20190619110000_3.jpg
D:\constructionworkers_train\NVR_ch2_main_20190619100000_20190619110000_4.jpg
```

Fig. 9. Example of manual annotation of workers with ‘LabelImg’.

$$recall = \frac{TP}{TP + FN} \tag{10}$$

where, ‘*precision*’ refers to the ratio of correctly detected workers to the total number of objects classified as workers. ‘*recall*’ refers to the ratio of correctly detected workers to all the actual workers in the images. A ‘true positive’ (TP) refers to a worker that is correctly detected. A ‘false positive (FP)’ occurs when a detected worker is actually some other object. A ‘false negative (FN)’ refers to a failure to detect a worker in the image.

In this study, following the work of Everingham, Van Gool, Williams, Winn, and Zisserman (2010), a threshold value of ‘true positive’ is set to 0.5. A true positive occurs when the detected bounding box overlaps with the ground truth annotated box by more than 0.5 IOU (Intersection

Table 4
CenterNet-based worker detection results based on testing dataset.

Resolution	time	batch	AP0.5 (All)	AP0.5 (small)	AR0.5 (All)	AR0.5 (small)
512 × 512	0.085	4	0.595	0.283	0.825	0.569
960 × 544	0.110	4	0.782	0.486	0.948	0.892
1440 × 832	0.175	2	0.800	0.555	0.950	0.906

over Union). If the IOU is less than 0.5 then it is considered a false positive. Table 4 presents the test results on the use of the trained CenterNet model to detect workers. It also demonstrates that our applied CenterNet can accurately detect construction workers. Fig. 10 presents examples of correct detections and errors, respectively.

To compare the performance of CenterNet to anchor-based detection network, two anchor-based detection approaches including Faster R-CNN and SSD were selected for comparison. The Faster R-CNN was proposed by Ren, He, Girshick, and Sun (2015) and it consists of three essential parts: (1) detection network; (2) region proposal network (RPN); and (3) fully connected layers for classification and bounding box regression. The Faster R-CNN achieved an accuracy of 73.2% mAP on PASCAL VOC 2007 and 70.4% mAP on PASCAL VOC 2012. The anchor boxes with different sizes and aspect ratios are essential to the detection performance with Faster R-CNN. In this study, Faster RCNN was implemented based on Google TensorFlow Object Detection API. We set the parameters as follows: (1) three different aspect ratios (0.5, 1.0, and 2.0); (2) one base size (256 × 256); and (3) four scale size (0.25, 0.5, 1.0, and 2.0).

SSD, the one-stage detector approach proposed by Liu et al. (2016), achieved 74.3% mAP on VOC2007 test at 59 FPS. SSD extracted feature maps from images and then produced bounding boxes and scores to detect object class instances. In this study, the SSD was implemented based on Google TensorFlow Object Detection API. Two critical parameters including anchor scale (4 × 4) and aspect ratio (0.5, 1.0, and 2.0) are set.

In our evaluation of the different object detection networks, we use the same training and testing dataset that was used for our CenterNet model. The hyper parameters (e.g., each model's learning rate) were based on the values used in the relevant references for the respective algorithm. Table 5 presents the detection performance, which shows that CenterNet is more accurate and has faster detection speed than the two anchor-based approaches. The evaluation also showed that the CenterNet model has the highest detection speed when the resolution of images is 512 × 512.

5.2. *Detection of violation of safe distancing rules*

After correctly detecting workers from the CCTV video in real-time, we then sought to determine the proximity between workers. In Singapore and many other countries, safe distancing rules by the government stipulates a minimum distance of at least one meter must be enforced at all times to reduce the risk of disease transmission during COVID-19 and this applies to all public places and workplaces, including construction sites. By applying the method presented in Section 3.1.2, each pixel represents a distance of 30 mm on the construction level and the distance between each pair of workers is determined based on the Euclidean distance matrix. When the safety distance of one metre was continuously breached for a predefined period of three seconds, a Telegram alert will be triggered. It is noted that the predefined period can be adjusted if necessary. Fig. 11 presents an example of a positive detection of workers breaching safe distancing rules in the case study. The warning alert was immediately sent to site personnel via Telegram and the data is captured on the management dashboard.

While our approach can accurately detect workers violating safe distancing rule, there are still some misdetections and errors. Fig. 12 presents examples of erroneous detection results (see red circles in the Figure).

6. Discussion

One of the challenges for site management during COVID-19 is to ensure that workers comply with safe distancing rules in a sustained manner. Many policies and regulations have been made in different countries to reduce physical interaction and ensure compliance to safe distancing and overcrowding rules. Our smart monitoring system provides site management with a real-time mechanism to proactively detect workers who have violated safe distancing rules during operations. As discussed earlier, the system can easily include monitoring of overcrowding, where a certain number of workers is congregating in a small area for more than certain amount of time. The system can alert any site personnel, but it will be especially useful to site managers and supervisors who can take immediate actions to modify workers' behaviours and emphasize the importance of compliance to safety rules. Our system also helps to identify trades and groups of workers with higher tendency for close proximity work. In the event that it is not possible for these workers to comply with the safe distancing rule continuously, the management can pay more attention to their temperature, hygiene, and segregation with other workers.

Thus, the contributions of this research are as follows. First, this study developed a smart real-time monitoring system that can be used to automatically monitor construction workers' compliance with safe distancing requirements. To reduce the risk of transmission of COVID-19, it is important to implement continuous supervision to encourage safe behaviour and identify safety violations and errors. However, behaviour supervision is time-consuming and there could be different observation biases between different observers. Furthermore, requiring human observers to be on site exposes the observers and workers to additional disease transmission risk. Thus, instead of human observers observing workers for extended period of time, the proposed smart real-time monitoring system provides an automated means to monitor safe distancing between workers and identify overcrowding situations. The developed system will help to minimize the observation biases and obtain consistent behaviour observation results. The observation data collected will also be more extensive and helps site management produce more reliable BBS indicators that can act as leading indicators for safety management. Second, as compared to previous studies, the CenterNet model used herein can better detect small worker images in CCTV footage. CenterNet can also address the drawbacks of anchor-based approaches (e.g., Faster R-CNN) in detecting small worker images in construction site video footages. Our experiments have demonstrated that CenterNet has better performance than popular state-of-the-art

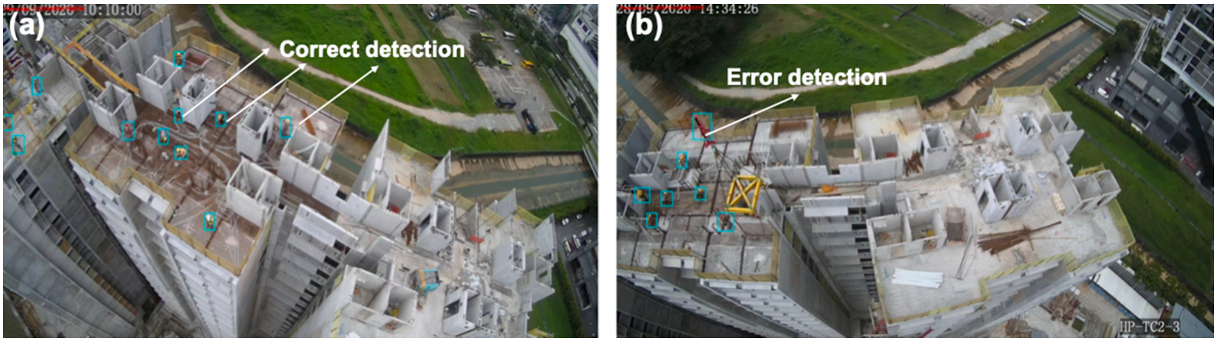
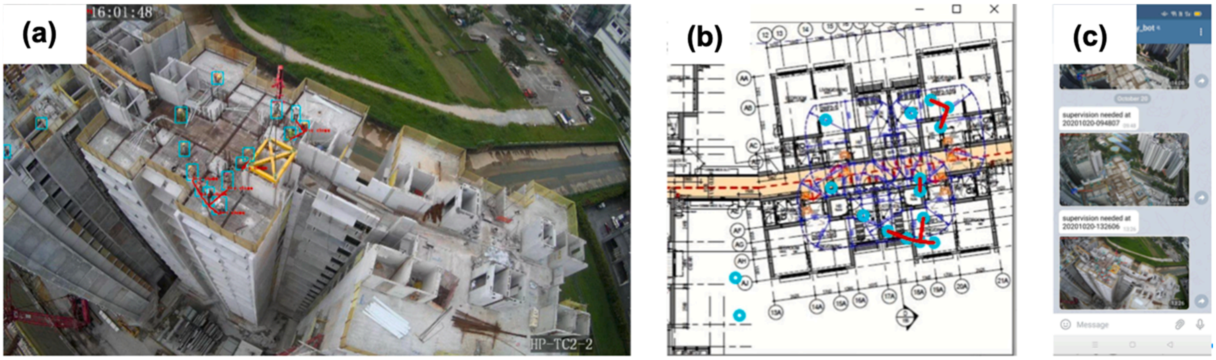


Fig. 10. Examples of worker detection using CenterNet.

Table 5
Comparison of state-of-the-art approaches.

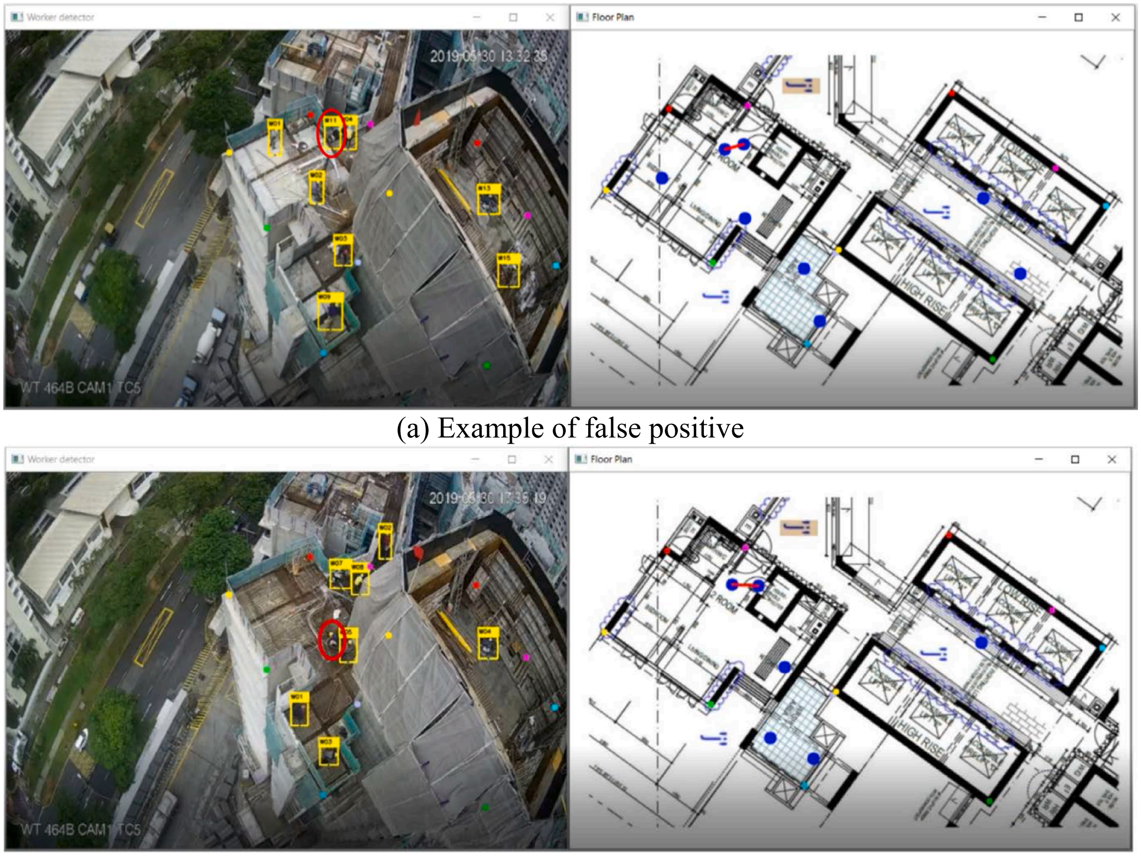
Model	Resolution	Speed (sec)	AP0.5 (All)	AP0.5 (Small)	AP0.5 (Medium)	AR0.5 (All)	AR0.5 (Small)	AR0.5 (Medium)
Faster RCNN	1024 × 600	0.108	0.585	0.231	0.640	0.703	0.448	0.780
SSD	640 × 640	0.106	0.601	0.310	0.638	0.639	0.363	0.730
CenterNet	512 × 512	**0.085**	0.595	0.283	0.648	0.825	0.569	0.869
CenterNet	960 × 544	0.110	0.782	0.486	0.832	0.948	0.892	0.958
CenterNet	1440 × 832	0.175	**0.800**	**0.555**	**0.835**	**0.950**	**0.906**	**0.958**



supervision needed at 20201020-094807

supervision needed at 20201020-132606

Fig. 11. Example of detection result with warning alert.



(a) Example of false positive

(b) Example of false negative

Fig. 12. Examples of detection errors.

object detector like Faster-RCNN and SSD when applied on the same dataset. We note that CenterNet had not been used in construction research, and based on the findings of this study, it is a promising algorithm that can be applied more widely in the construction industry.

Furthermore, this paper suggests that future studies on BBS management should consider the use of computer vision. One of the critical success factors of BBS approaches is the availability of reliable behaviour observation data, but traditional BBS implementation relies heavily on manual observation, which can be inconsistent, resource intensive and provides inadequate amount of data (Fang, Love, et al., 2020; Fang, Ding, et al., 2020). Instead of manual observation, computer vision can automatically identify, and measure safety behavior of workers, which can support the effective implementation of BBS. For example, based on Guo, Goh, and Le Xin Wong (2018) the percentage of safe or unsafe behavior can be determined for different time intervals (e.g., morning, afternoon, week, and month) and for different locations in a worksite. Then, based on the more detailed indicator of safety behaviour, different interventions (e.g., feedback, and training) can be implemented in a more targeted manner to improve safety performance.

7. Limitations and future works

There are four limitations in our study. Firstly, the Homography approach used in this research to determine distance is sensitive to the point selections, and the estimation error is linearly proportional to the distance from the camera to the workers. With further improvement of distance algorithms (Rodríguez-Quiñonez et al., 2017), the distance errors in the developed system can be reduced. However, it must be noted that in comparison to manual observations, the developed system is able to monitor the site continuously, consistently, and more accurately. Secondly, due to the lack of depth information from a regular CCTV camera, the distance measurement using Homography are estimated. In terms of safe distancing, such estimates are acceptable. Furthermore, our developed system is more implementable as it uses existing CCTV cameras and do not require any additional specialized cameras. However, our further work will consider the use of stereo camera to improve accuracy on distance measurement. Thirdly, the calibration for distance measurement needs to be conducted every time the tower crane is jacked up. The calibration is time-consuming and can contribute to errors. We are currently developing an automated calibration method that can self-adjust the calibration parameters. Fourthly, occlusion is still a major issue that affects most computer vision-based systems, and our developed system is not an exception. To address this limitation, we will consider the use of multi-cameras approaches to detect and track target workers across different cameras. Lastly, the developed smart monitoring system was only tested on two construction sites in Singapore. The generalizability of our proposed smart monitoring system still needs to be improved further. Currently, more data are being collected from different sites to improve the performance of the model across different construction sites. Concurrently, self-learning algorithm (Le, Sugimoto, Ono, & Kawasaki, 2020) is being explored to facilitate efficient improvement of the models.

In future, the proposed computer vision-based system can also be integrated with Internet of Things (IOT), Cloud platform and wearables (e.g., wrist band), person-ReID detection method. By integrating these technologies, a more robust and responsive system can be developed and the worker who is breaching the safety distancing can be detected and tracked.

8. Conclusion

This study proposes an automatic real-time monitoring system for detecting workers violating safe distancing rules on construction sites during the on-going COVID-19 pandemic. The developed system integrates the recent advances in computer vision and deep learning, including object detection with CenterNet, proximity determination with Homography and Euclidean distance matrix, and warning alert generation. A case of a public housing project in Singapore is used to validate the effectiveness and feasibility of our developed system. The experimental results show that CenterNet performs better than other popular object detection networks, including Faster RCNN and SSD. Our proposed computer vision-based smart monitoring system supports efficient implementation of behaviour-based safety (BBS) such that behavioural observation is continuous and automated. In doing so, workers' safe behaviour can be reinforced, and unsafe behaviour can be corrected. Considering the observation biases, time and effort required for human supervision of people's behaviour, our developed system is a favourable alternative or supplement. The key contributions of our paper are twofold: (1) it is demonstrated that monitoring of safe distancing on construction sites can be automated using the proposed computer vision-based smart monitoring system; and (2) CenterNet, an anchorless detection model, outperforms current state-of-the-art approaches (e.g., Faster R-CNN, SSD) in the real-time detection of construction workers.

CRediT authorship contribution statement

Yang Miang Goh: Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Jing Tian:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Eugene Yan Tao Chian:** Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-100E-2019-026) and the Singapore Housing and Development Board (HDB).

References

Ahmed, I., Ahmad, M., Rodrigues, J. J. P. C., Jeon, G., & Din, S. (2021). A deep learning-based social distance monitoring framework for COVID-19. *Sustainable Cities and Society, 65*, 102571. https://doi.org/10.1016/j.scs.2020.102571

Belongie, S., & Kriegman, D. (2007). *Explanation of Homography Estimation from Department of Computer Science and Engineering*. San Diego: University of California.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.

Building and Construction Authority. (2020). Built environment sector COVID-19 info. https://www1.bca.gov.sg/COVID-19.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291–7299).

CDC, Centers for Disease Control and Prevention. (2021). How to Protect Yourself & Others. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html. Accessed on November 9, 2021.

Chian, E., Fang, W., Goh, Y. M., & Tian, J. (2021). Computer vision approaches for detecting missing barricades. *Automation in Construction, 131*, 103862. https://doi.org/10.1016/j.autcon.2021.103862

Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., … Schünemann, H. J. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and meta-analysis. *The lancet, 395*(10242), 1973–1987.

Chu, M., Matthews, J., & Love, P. E. D. (2018). Integrating mobile building information modelling and augmented reality systems: An experimental study. *Automation in Construction, 85*, 305–316.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.

Dokmanic, I., Parhizkar, R., Ranieri, J., & Vetterli, M. (2015). Euclidean distance matrices: Essential theory, algorithms, and applications. *IEEE Signal Processing Magazine, 32*(6), 12–30.

Ding, L., Fang, W., Luo, H., Love, P. E. D., Zhong, B., & Ouyang, X. (2018). A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automation in Construction, 86*, 118–124.

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). Centernet: Keypoint triplets for object detection. In *In Proceedings of the IEEE international conference on computer vision* (pp. 6569–6578).

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision, 88*(2), 303–338.

Fang, W., Ding, L., Luo, H., & Love, P. E. D. (2018). Falls from heights: A computer vision-based approach for safety harness detection. *Automation in Construction, 91*, 53–61.

Fang, W., Ding, L., Zhong, B., Love, P. E. D., & Luo, H. (2018). Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach. *Advanced Engineering Informatics, 37*, 139–149.

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction, 85*, 1–9.

Fang, W., Zhong, B., Zhao, N., Love, P. E. D., Luo, H., Xue, J., & Xu, S. (2019). A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Advanced Engineering Informatics, 39*, 170–177.

Fang, W., Love, P. E. D., Luo, H., & Ding, L. (2020). Computer vision for behaviour-based safety in construction: A review and future directions. *Advanced Engineering Informatics, 43*, 100980. https://doi.org/10.1016/j.aei.2019.100980

Fang, W., Ding, L., Love, P. E. D., Luo, H., Li, H., Peña-Mora, F., … Zhou, C. (2020). Computer vision applications in construction safety assurance. *Automation in Construction, 110*, 103013. https://doi.org/10.1016/j.autcon.2019.103013

Geerts, G. L. (2011). A design science research methodology and its application to accounting information systems research. *International Journal of Accounting Information Systems, 12*(2), 142–151.

Goh, Y. M., Ubeynarayana, C. U., Wong, K. L. X., & Guo, B. H. W. (2018). Factors influencing unsafe behaviors: A supervised learning approach. *Accident; Analysis and Prevention, 118*, 77–85. https://doi.org/10.1016/j.aap.2018.06.002

Guo, B. H. W., Goh, Y. M., & Le Xin Wong, K. (2018). A system dynamics view of a behavior-based safety program in the construction industry. *Safety Science, 104*, 202–215.

Guo, B. H. W., Zou, Y., Fang, Y., Goh, Y. M., & Zou, P. X. W. (2021). Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Safety Science, 135*, 105130. https://doi.org/10.1016/j.ssci.2020.105130

Kim, H., Kim, K., & Kim, H. (2016). Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects. *Journal of Computing in Civil Engineering, 30*(4), 04015075. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000562

Kim, K., Kim, H., & Kim, H. (2017). Image-based construction hazard avoidance system using augmented reality in wearable device. *Automation in Construction, 83*, 390–403.

Kim, D., Liu, M., Lee, S., & Kamat, V. R. (2019). Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. *Automation in Construction, 99*, 168–182. LabelImg, https://github.com/tzutalin/labelImg.

Le, T. N., Sugimoto, A., Ono, S., & Kawasaki, H. (2020). Toward Interactive Self-Annotation For Video Object Bounding Box: Recurrent Self-Learning And Hierarchical Annotation Based Framework. In *In The IEEE winter conference on applications of computer vision* (pp. 3231–3240).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., & Sun, J. (2017). Light-head r-cnn: In defense of two-stage object detector. arXiv preprint arXiv:1711.07264.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37). Cham: Springer.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–755). Cham: Springer.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

Luo, H., Liu, J., Fang, W., Love, P. E. D., Yu, Q., & Lu, Z. (2020). Real-time smart video surveillance to manage safety: A case study of a transport mega-project. *Advanced Engineering Informatics, 45*, 101100. https://doi.org/10.1016/j.aei.2020.101100

Luo, H., Wang, M., Wong, P.-Y., & Cheng, J. C. P. (2020). Full body pose estimation of construction equipment using computer vision and deep learning techniques. *Automation in Construction, 110*, 103016. https://doi.org/10.1016/j.autcon.2019.103016

Luo, H., Xiong, C., Fang, W., Love, P. E. D., Zhang, B., & Ouyang, X. (2018). Convolutional neural networks: Computer vision-based workforce activity assessment in construction. *Automation in Construction, 94*, 282–289.

Luo, X., Li, H., Wang, H., Wu, Z., Dai, F., & Cao, D. (2019). Vision-based detection and visualization of dynamic workspaces. *Automation in Construction, 104*, 1–13.

Luo, X., Li, H., Yang, X., Yu, Y., & Cao, D. (2019). Capturing and understanding workers' activities in far-field surveillance videos with deep action recognition and Bayesian nonparametric learning. *Computer-Aided Civil and Infrastructure Engineering, 34*(4), 333–351.

Ministry of Health. (2020). https://www.moh.gov.sg/news-highlights/details/stricter-safe-distancing-measures-to-prevent-further-spread-of-covid-19-cases, 20th May, 2020.

Ministry of Health. (2021). COVID-19 Phase Advisory. https://www.moh.gov.sg/covid-19-phase-advisory. Accessed on November 9, 2021.

Mustafah, Y. M., Noor, R., Hasbi, H., & Azma, A. W. (2012). Stereo vision images processing for real-time object distance and size measurements. In *2012 international conference on computer and communication engineering* (pp. 659–663). IEEE.

Nath, N. D., Behzadan, A. H., & Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment. *Automation in Construction, 112*, 103085. https://doi.org/10.1016/j.autcon.2020.103085

National Statistic. (2021). Coronavirus (COVID-19) related deaths by occupation, England and Wales, https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/coronaviruscovid19relateddeathsbyoccupationenglandandwales/deathsregistereduptoandincluding20april2020. Access January 2021.

Newell, A., Huang, Z., & Deng, J. (2017). Associative embedding: End-to-end learning for joint detection and grouping. In *Advances in neural information processing systems* (pp. 2277–2287).

Occupational Safety and Health Administration. (2020). https://www.osha.gov/SLTC/covid-19/construction.html.

Rahman, K. A., Hossain, M. S., Bhuiyan, M. A. A., Zhang, T., Hasanuzzaman, M., & Ueno, H. (2009). Person to camera distance measurement based on eye-distance. In *2009 Third International Conference on Multimedia and Ubiquitous Engineering* (pp. 137–141). IEEE.

Rashwan, H. A., Solanas, A., Puig, D., & Martínez-Ballesté, A. (2015). Understanding trust in privacy-aware video surveillance systems. *International Journal of Information Security, 15*(3), 225–234. https://doi.org/10.1007/s10207-015-0286-9

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

Rodríguez-Quiñonez, J. C., Sergiyenko, O., Flores-Fuentes, W., Rivas-lopez, M., Hernandez-Balbuena, D., Rascón, R., & Mercorelli, P. (2017). Improve a 3D distance measurement accuracy in stereo vision systems using optimization methods' approach. *Opto-Electronics Review, 25*(1), 24–32.

Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics, 29*(2), 239–251.

Shapiro, R. (1978). Direct linear transformation method for three-dimensional cinematography. *Research Quarterly. American Alliance for Health, Physical Education and Recreation, 49*(2), 197–205.

Son, H., Seong, H., Choi, H., & Kim, C. (2019). Real-time vision-based warning system for prevention of collisions between workers and heavy equipment. *Journal of Computing in Civil Engineering, 33*(5), 04019029. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000845

Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Springer Science & Business Media.

Tian, Z., Shen, C., Chen, H., & He, T. (2019). Fcos: Fully convolutional one-stage object detection. In *In Proceedings of the IEEE international conference on computer vision* (pp. 9627–9636).

van Aken, J. E. (2005). Management research as a design science: Articulating the research products of mode 2 knowledge production in management. *British Journal of Management, 16*(1), 19–36.

Johns Hopkins University, Coronavirus resource center, https://coronavirus.jhu.edu/map.html. Access October 15th, 2020.

Wahab, M. N. A., Sivadev, N., & Sundaraj, K. (2011). Target distance estimation using monocular vision system for mobile robot. In *2011 IEEE Conference on Open Systems* (pp. 11–15). IEEE.

Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J., & Prescott, H. C. (2020). Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): A review. *JAMA, 324*(8), 782–793.

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. *Automation in Construction, 106*, 102894. https://doi.org/10.1016/j.autcon.2019.102894

Yang, D., Yurtsever, E., Renganathan, V., Redmill, K. A., & Özgüner, Ü. (2020). A Vision-based Social Distance and Critical Density Detection System for COVID-19. arXiv preprint arXiv:2007.03578.

Yu, Y., Guo, H., Ding, Q., Li, H., & Skitmore, M. (2017). An experimental study of real-time identification of construction workers' unsafe behaviors. *Automation in Construction, 82*, 193–206.

Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as points. arXiv preprint arXiv:1904.07850.