This is the Pre-Published Version.
Y. Ding, Y. Zhang, X. Huang (2023) *Intelligent Emergency Digital Twin System for Monitoring Building Fire Evacuation*, **Journal of Building Engineering**, 77, 107416. https://doi.org/10.1016/j.jobe.2023.107416

Intelligent Emergency Digital Twin System for Monitoring Building Fire Evacuation

Yifei Ding[a], Yuxin Zhang[a,b,c,*], Xinyan Huang[a,*]

[a]*Research Centre for Fire Safety Engineering, Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong*

[b]*State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China*

[c]*Department of Geotechnical Engineering, Tongji University, Shanghai, China*

* *Corresponding to* yx.zhang@polyu.edu.hk *(YZ);* xy.huang@polyu.edu.hk *(XH)*

Abstract:

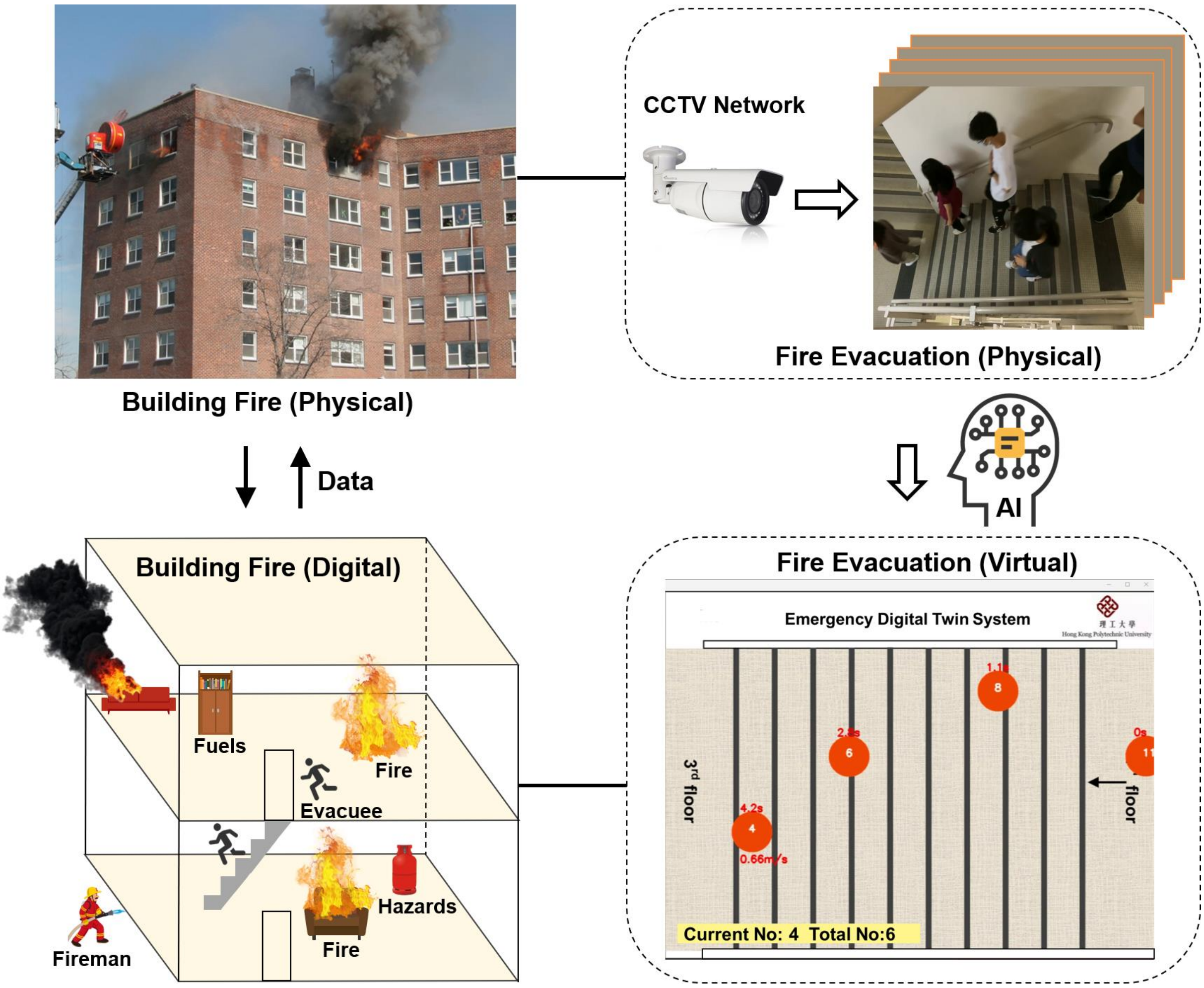
The provision of real-time and detailed evacuation information feedback is vitally significant for the formulation and adaptation of the onsite evacuation strategy. The conventional surveillance system relying on surveillance cameras is limited to processing video and fails to extract human behaviour or provide privacy protection. This work proposes an Intelligent Emergency Digital Twin system based on computer vision and deep learning. The system comprises (1) CCTV network, (2) YOLOv4 evacuee detector, (3) DeepSORT evacuee tracker, (4) Perspective transformer, and (5) Digital Twin interface. It enables the detection and tracking of evacuees, the calculation of their egress speed, and the protection of their privacy in a digital interface. The proposed system was evaluated in a staircase of an office building through two types of tests with positive results: ratio of successfully detecting is 100% in individual objects test and about 90% in multiple objects test. The evacuation data generated by the digital twin system would be useful for guiding evacuations out of fire scenarios. This proposed digital twin framework can lay the foundation for the implementation of smart human monitoring in fire scenarios for buildings.

Keywords: *fire emergency; computer vision; human behaviour; digital twin; building evacuation*

Abbreviation

AI	Artificial intelligence	NMS	Non-maximum suppression
CCTV	Closed-circuit television	PAN	Path aggregation network
CNN	Convolutional neural network	R-CNN	Regions with CNN features
CSP	Cross stage partial (Network)	RSET	Required safe regress time
CV	Computer vision	SORT	Simple online and real-time tracking
HRR	Heat release rate	SSD	Single shoot multi-box detector
IOU	Intersection-over-union	SPP	Spatial pyramid pooling
MOT	Multiple objects tracking	YOLO	You Only Look Once

Graphic Abstract



1. Introduction

Due to rapid globalisation and urbanisation, a tremendous increase in population density induces greater performance and function demand for construction. In modern architectural design, the quest for height and structural complexity provides greater usable space and building aesthetics, while increasing the higher safety risk and corresponding evacuation challenges once fire occurs [1]. For instance, estimated 2,092 occupants in World Trade Centre failed to egress during the "9/11 attacks", which exposed several issues concerning evacuation (Fig. 1a). The 2016 Hong Kong mini storage fire caused serious casualty of firefighters (Fig. 1b). In most building fires, the safe and timely evacuation can prevent casualties (Fig. 1c). However, most reasons for death and injury in building fire are related to the lack of real-time and dynamic evacuation information in fire, such as occupants' positions, their distributions and firefighters' conditions, etc., at the escape and rescue phase [2].

Fig. 1. (a) WTC fire and collapse, 2001, (b) Hong Kong mini-storage fire, 2016, and (c) occupants evacuating from a building fire [3,4][5].

Recently, there has been an increasing awareness of the importance of evacuation strategy and human behaviours in fire research. In those research, the evacuation challenge and strategy in confined construction environments [6], the smart numerical model for pedestrian evacuation [7], and human egress characteristics in building staircase [8] are studied in detail. In modern fire research and management, obtaining and conveying real-time data on evacuees can be of great help in evacuation process monitoring and making dynamic evacuation decisions. For example, early perception of incident location such as fire source information makes human evacuation more efficient [9]. Moreover, in enclosure fire scenarios, technical installations are the only approach to deliver fire and evacuee information [10]. However, the current CCTV system cannot provide sufficient feedback on evacuation sites. In addition, manually observing CCTV monitors to acquire human behaviours during or after the evacuation process would cost more time and be labour intensive. Therefore, to obtain more real-time information on human fire evacuation behaviours for a faster and safer evacuation, we need a more intelligent monitoring system.

In recent years, artificial intelligence (AI) is widely used in fire safety science and fire protection engineering and has become a critical engine driving the development of smart firefighting [11]. For example, AI methods have been applied to forecast flashover [12,13], fire source location [14–16], fire hazards [17–19], and temperature fields [20–22]. In these studies, deep learning models have been widely used to generate accurate fire modelling and predictions. Computer vision (CV), as another important field of artificial intelligence, has enormous scope for application in fire calorimetry and human behaviours research. Recently, with the popularity of object detection and tracking algorithms based on CV and deep learning, scholars are also focusing on their value in fire science and safety engineering. For instance, several academics developed AI models based on object detection algorithms for smoke [23] and fire flame detection [24].

In addition, object detection algorithms, especially YOLO (You Only Look Once) [25], are also used in the study of evacuation models and evacuation systems. Chen *et al.* [26] applied a YOLO-based recursive neural network to the evacuation model in the design of public construction. Zhang *et al.* [27] used YOLO to extract the individual's position and velocity data in the crowd for the development of an evacuation simulation system. Li *et al.* [28] applied an object detecting algorithm to extract and model human movement characteristic parameters during evacuation. While past research utilised YOLO to extract human features for evacuation research, they have not applied YOLO's strong and fast video processing potential combined with other CV tricks to obtain and analyse real-time evacuation information, which would be the critical point in this research.

Moreover, the existing CCTV monitoring network suffers the concerns of privacy protection of occupants during the evacuation while the worldwide emphasis focus on privacy and ethical protection [29,30]. With the support of CV, the digital twin framework is introduced to establish a connection between reality and the virtual world and present real-world evacuees' data on the virtual side to improve privacy protection. The concept of the digital twin, presented by Grieves *et al.* [31], contains three components, a physical product, a corresponding virtual representation, and mutual data connections. The recent concept of a digital twin has been maturely applied to building life-cycle management [32–34] and fire safety [15,35–38]. For example, our previous work has combined sensor data and AI model to develop digital twin systems to quantify the fire location, size and hazards inside room and tunnel [14–16]. Jiang *et al.* [37] proposed a digital twin system combined with semantic web technologies, which integrates domain knowledge, BIM information and IoT data to implement intelligent system control for building fire protection. Khajavi *et al.* [38] presented an implementation structure of building lifecycle digital twin and applied sensor package, 3D visualisation and AI to detect and predict potential building fire threats. Elhami-Khorasani *et al.* [35,36] proposed and demonstrated the use of machine vision in identifying the fuel and estimating the fuel load.

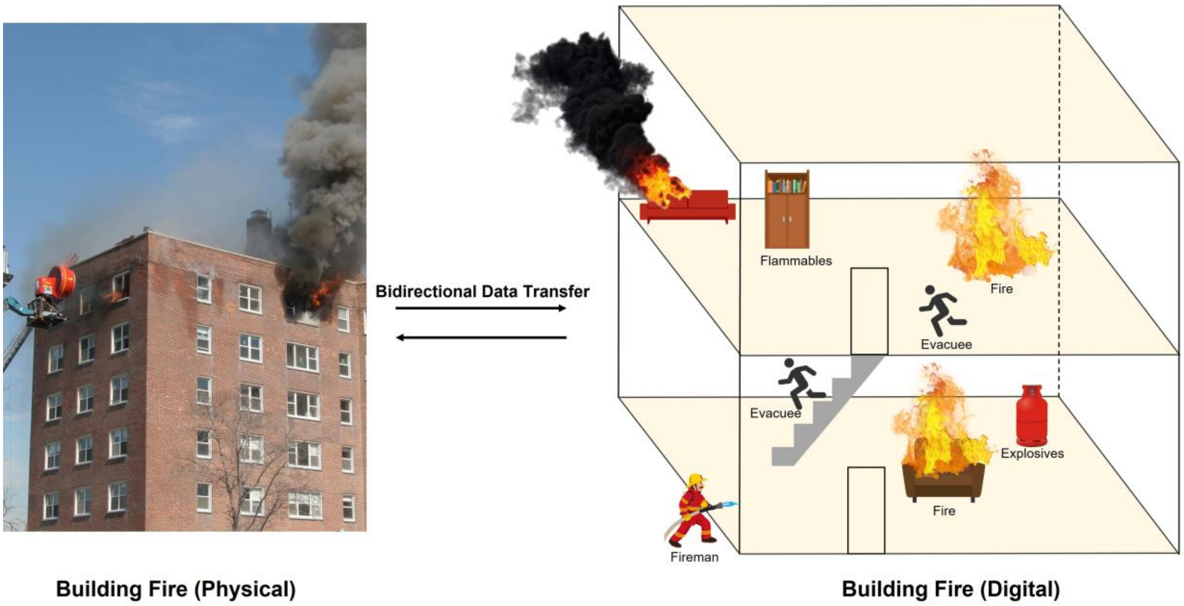


Fig. 2. Demonstration of digital twin structure for building fire.

These proposed building fire digital twin systems involve new technologies, such as IoT networks, computer vision, data fusion, and deep learning. Fig. 2 illustrates a framework for the use of digital twin in building fire safety management. It enables real-time virtual reconstruction of building fire scenario, prediction and controlling of fire risks, containing (1) detection and prediction of fire and smoke motions, fuel and flammables distribution by IoT sensors and AI model, and (2) the track of fire evacuation and rescue by cameras and human detection model. However, few related framework and demonstration of a digital twin for building fire evacuation monitoring systems have been presented, requiring an in-depth study.

This work proposes a framework for an Intelligent Emergency Digital Twin monitoring system, and it mainly contains five components: (1) CCTV network, (2) YOLOv4 evacuee detector, (3) DeepSORT evacuee tracker, (4) Perspective transformer, and (5) Digital Twin interface for displaying evacuee information in the virtual edge. In this study, a multiple object tracking algorithm architecture is developed to detect and track evacuees displayed by the digital twin interface. In addition, a speed estimation algorithm is developed to calculate occupants moving speed regarded as the critical feedback of the digital twin system. To identify the performance of the presented system, pedestrian experiments and fire drill were performed to test the system and demonstrate its potential in fire evacuation research.

2. Methodology

2.1. Deep learning evacuee detection algorithm

Evacuees in emergency scenarios can be detected by a deep learning object detection algorithm based on CV and deep convolutional neural network (CNN). In recent years, several well-known object detection algorithms, including R-CNN families [39], Single Shoot Multibox Detector (SSD) [40], and YOLO series [25], have evolved into accurate and lightweight detection systems that have been widely applied to academia and industry. Among all those approaches, YOLO is a high-efficient method that combines the detected components into a single convolutional neural network in the full image. It [25] was presented first by Joseph Redmon in 2016 and caused quite a sensation in computer vision and machine learning fields. In this work, we select the 4th version of YOLO (YOLOv4) [41] as an evacuee detector to balance the trade-off between detection speed and precision of monitoring.

The YOLOv4 frame (Fig. 3) of object detection is regarded as a regression problem to spatially separated bounding boxes and associated class probabilities directly from full images in one evaluation. YOLO unifies the separate components of localisation and classification into a single CNN and uses features from the entire image to predict each bounding box. It also predicts all bounding boxes across all classes for an image simultaneously. This means the network reasons globally about the full image and all the objects in the image. The YOLO design enables end-to-end training and real-time speeds while maintaining high precision. Moreover, YOLO divides the input image into $\mathbf{S} \times \mathbf{S}$ grids. If the centre of an object falls into a grid cell, that grid cell is responsible for detecting that object. Each grid cell predicts B bounding boxes and confidence scores for those boxes.

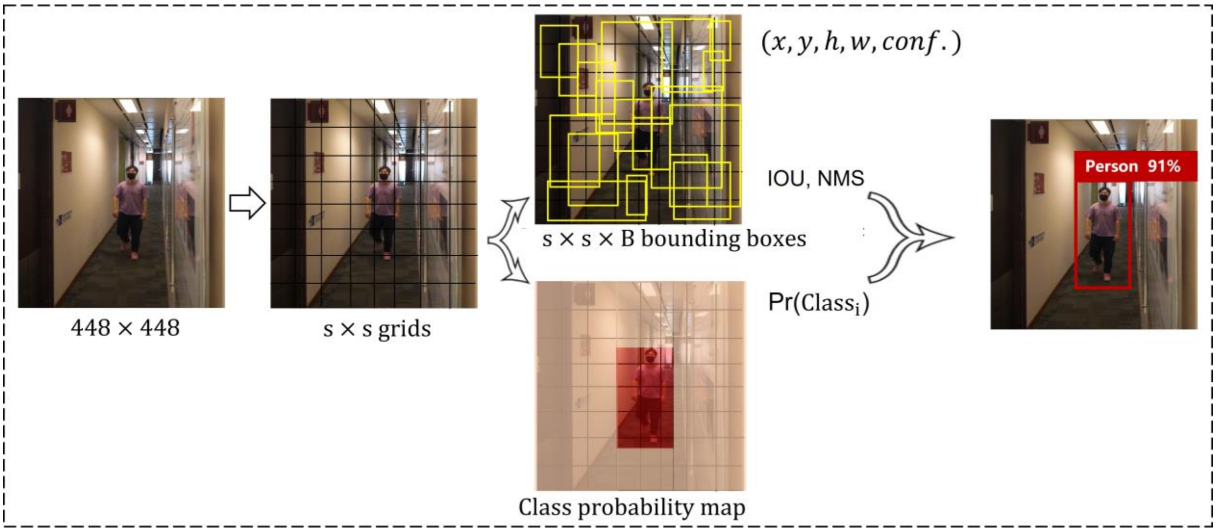


Fig. 3. The model diagram of YOLO applied in the digital twin system.

In this work, we select YOLOv4 [41], one of the most widely used versions of the YOLO series, and combine it with DeepSORT [42], a well-known multiple objects tracking (MOT) algorithm as the evacuee detection and tracking architecture. YOLOv4 inherits the fast-processing speed advantage of previous versions of YOLO families and applies various other tricks to improve the performance of the deep learning model. As shown in Fig.4, the architecture of YOLOv4 frameworks consists of the backbone: CSPDarknet53 [43], the neck: SPP [44] +PAN [45] and the detecting head: YOLOv3 [46]. The backbone is pre-trained and mainly contains convolutional layers that turn the input frames in the raw video (Fig.4) into the feature space.

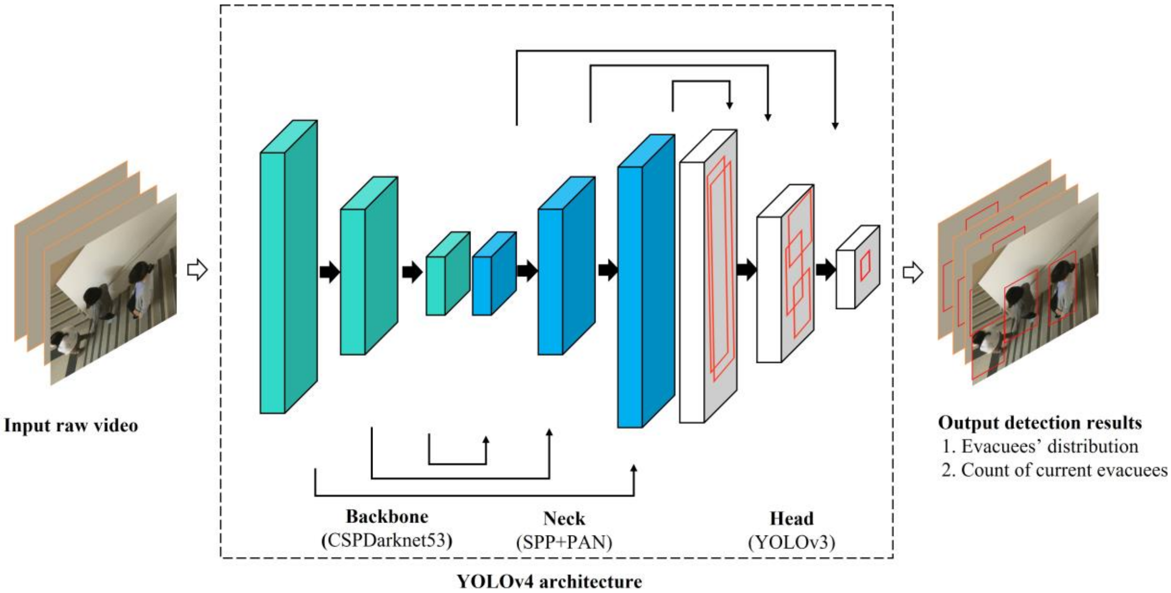


Fig. 4. Demonstration of YOLOv4 architecture in the detector of the digital twin system.

In the backbone, the CSP (Cross-Stage Partial network) plays the role of shortcut connections among convolutional layers to enhance the efficiency and precision of feature extractors. The combination of SPP (Spatial Pyramid Pooling) and PAN (Path Aggregation Network) is a huge innovative progress for the later versions of YOLO, and it forms the feature enhancers of the neck. The neck is usually inserted between the backbone and head to collect feature maps from the different stages and is composed of bottom-up paths and top-down paths. The head part of YOLOv4 is YOLOv3 detector and is used to perform one-stage dense prediction. The final prediction is composed of a vector containing the category and corresponding confidence score as well as the bounding boxes (for example the red detection boxes in Fig.4).

2.2. Evacuee tracking algorithm

When processing a continuous video stream, the YOLO detector would re-identify and relabels objects at each frame, which results in an evacuee would be given a new ID in each frame. Frequent ID switching is useless for evacuee trajectory surveillance, so multiple objects tracking (MOT) algorithms are needed in the digital twin system to settle this issue. DeepSORT is a typical MOT algorithm based on a deep convolutional neural network, which is the upgraded version of SORT [47]. DeepSORT inherits the general framework and features of the original version that associates the inter-frame identity documents (IDs) by Hungarian algorithm [48] executing linear assignment based on the Intersection-Over-Union (IOU) distance between the Kalman Filter [49] prediction and tracking boxes.

DeepSORT improves the tracking precision and reduces the ID switching frequency significantly by incorporating the object appearance information into the matching calculation so that the IDs of objects can also be matched accurately in case the object is delayed or blurred. In addition, another innovation of DeepSORT is the introduction of a cascade matching scheme ensuring that each detector assigns a single tracker to improve the tracking performance. For example (Fig. 5), the predicted box and the detected bounding box in frame $\boldsymbol{i}$ are inputted in a neural network to extract the features and conduct the cascade matching in prediction status. If the updated bounding box in frame $\boldsymbol{i+1}$ matched successfully with the predicted box, the assigned ID would be constant to represent continuous tracking.

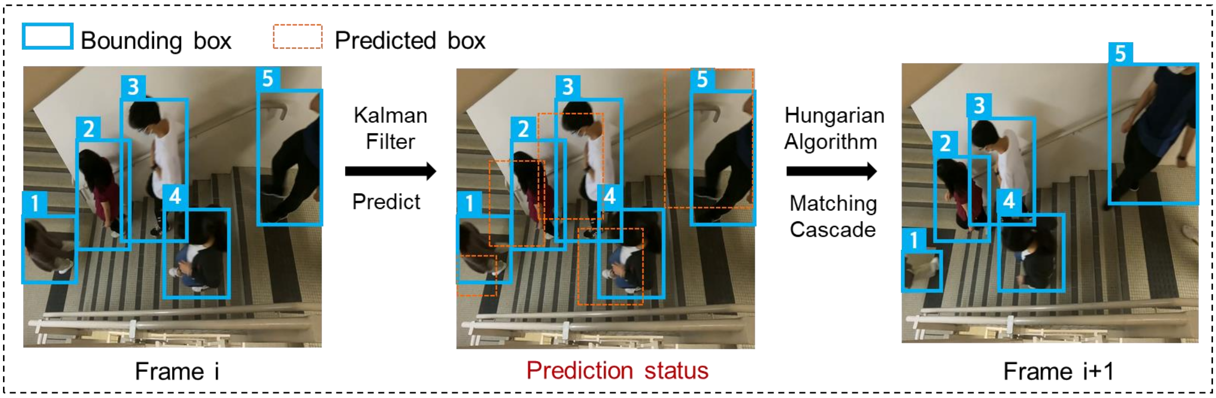


Fig. 5. The illustration of the flow chart of DeepSORT.

2.3. *Inverse perspective mapping*

Traditional CCTV surveillance can induce a range of ethical and personal privacy issues and the detail of the evacuee's appearance is not a critical concern in fire evacuation studies. Therefore, we present a mapping method based on an inverse perspective view into bird's eye view that maps the tracking data of the CCTV stream into a two-dimensional plane and represents evacuees as simplified balls. To make the bird's eye view transformation as the virtual side, the essential point pairs are to calculate the transformation matrix, which is computed using a function in OpenCV Library, which is defined as "*getPerspectiveTransform*".

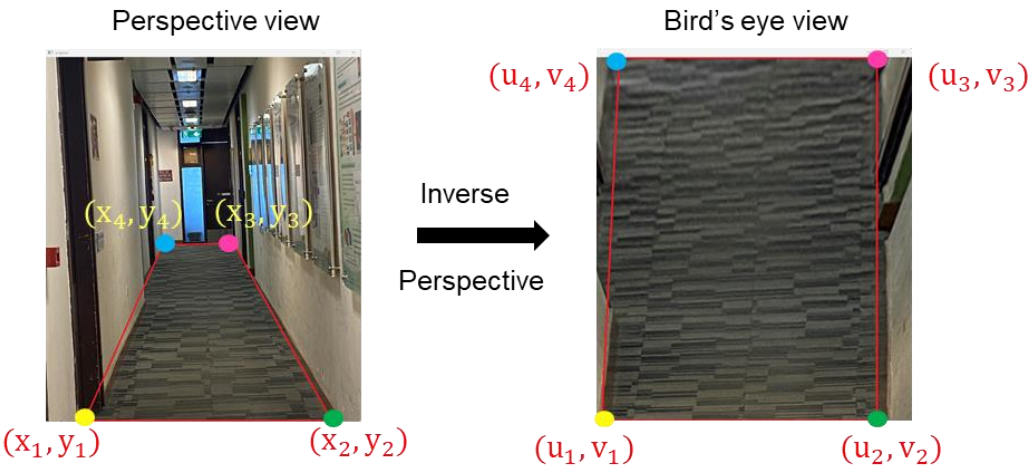


Fig. 6. Perspective transformation to bird'-eye view. The colourful points represent chosen point pairs for the calculation of the transformation matrix.

The calculation is performed by choosing the source and the target points in the monitoring scenarios, see Fig.6 and Eq. (1-3). In addition to that, the bird's-eye view matrix is achieved by multiplying each element matrix from the source of images using the function "*warpperspective*" in OpenCV Library, see Eq. (4).

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1u_1 & -y_1u_1 & -u_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1v_1 & -y_1v_1 & -v_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2u_2 & -y_2u_2 & -u_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2v_2 & -y_2v_2 & -v_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3u_3 & -y_3u_3 & -u_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3v_3 & -y_3v_3 & -v_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4u_4 & -y_4u_4 & -u_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4v_4 & -y_4v_4 & -v_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{1}$$

where, (x_i, y_i) and (u_i, v_i) are coordinates of source points and target points respectively, and h_{ij} is transformation coefficient. We denote M as the left 9×9 matrix and denote H as the 9×1 h vector, denote Y as the right vector.

$$H = M^{-1}Y \tag{2}$$

H can also denote the 3×3 matrix:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_{11} & \boldsymbol{h}_{12} & \boldsymbol{h}_{13} \\ \boldsymbol{h}_{21} & \boldsymbol{h}_{22} & \boldsymbol{h}_{23} \\ \boldsymbol{h}_{31} & \boldsymbol{h}_{32} & \boldsymbol{h}_{33} \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} \boldsymbol{u}_j \\ \boldsymbol{y}_j \\ \boldsymbol{1} \end{bmatrix} = \boldsymbol{H} \begin{bmatrix} \boldsymbol{x}_j \\ \boldsymbol{y}_j \\ \boldsymbol{1} \end{bmatrix} \tag{4}$$

where, $(\boldsymbol{x}_j, \boldsymbol{y}_j)$ and $(\boldsymbol{u}_j, \boldsymbol{v}_j)$ are coordinates of evacuees presented on perspective view and bird's eye view, respectively. The list containing coordinates and size of each bounding box, ID, and speed is mapped at the same time. Overall, the artificial intelligence model and computer vision model are applied to process the video sequence obtained by the CCTV network from the physical side and inverse perspective mapping algorithm simulates virtual interface.

2.4. Speed estimation

Evacuee's movement speed especially when occupants pass through corridors or go down staircases is an important research object and the critical parameter to estimate the required safe egress time (RSET). The speed of evacuee is greatly influenced by the walkways slope and occupants' physical conditions [50]. Therefore, the commander of the fire evacuation could roughly evaluate the physical characteristics of the evacuees and the environmental conditions in fire scenario according to evacuee's speed and distribution, and further decide to whether provide additional assistance.

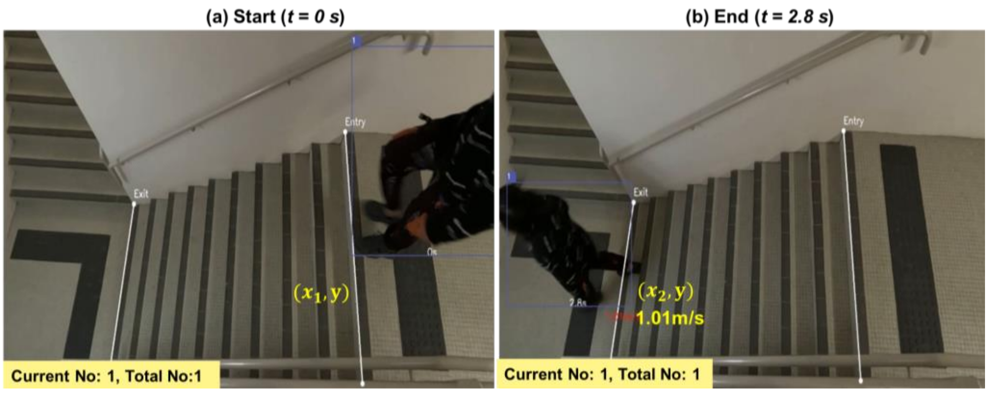


Fig. 7. Illustration of the principle of timer and speed estimator for one evacuee, (a) start, and (b) end.

Previously, the moving speed of occupants was counted manually from the video and calculated using the distance and time between camera positions for each occupant [47]. time-consuming This method is labour-intensive and prone to human error. To make it easier and more accurate to measure pedestrian speed, a vision-based speed estimation algorithm was proposed and embedded in the digital twin system, as shown in Fig. 7.

The timer starts when the distance between the lower left point of the bounding box and the entry line, denotes $\boldsymbol{D}_1(\boldsymbol{x}_1, \boldsymbol{y}) = \boldsymbol{0}$ and stops when the distance between the lower right point of the bounding box and the exit line, and denotes $\boldsymbol{D}_2(\boldsymbol{x}_2, \boldsymbol{y}) = \boldsymbol{0}$. The time of each evacuee will be stored, and the moving speed is calculated by the distance between two lines over the crossing time. This method automatically and conveniently estimates the speed and is mapped to the digital twin interface by AI algorithm and is not constrained by the camera's shooting angle.

3. The framework of the intelligent digital twin system

The framework of the proposed intelligent monitoring system mainly contains five components (see Fig. 8): (1) CCTV network, (2) YOLOv4 evacuee detector, (3) DeepSORT evacuee tracker, (4) Perspective transformer, and (5) Digital Twin interface. The proposed evacuation digital twin is designed to operate and survive in real fire scene and achieve three main tasks, that is, real-time data acquisition, data processing, and digital presentation.

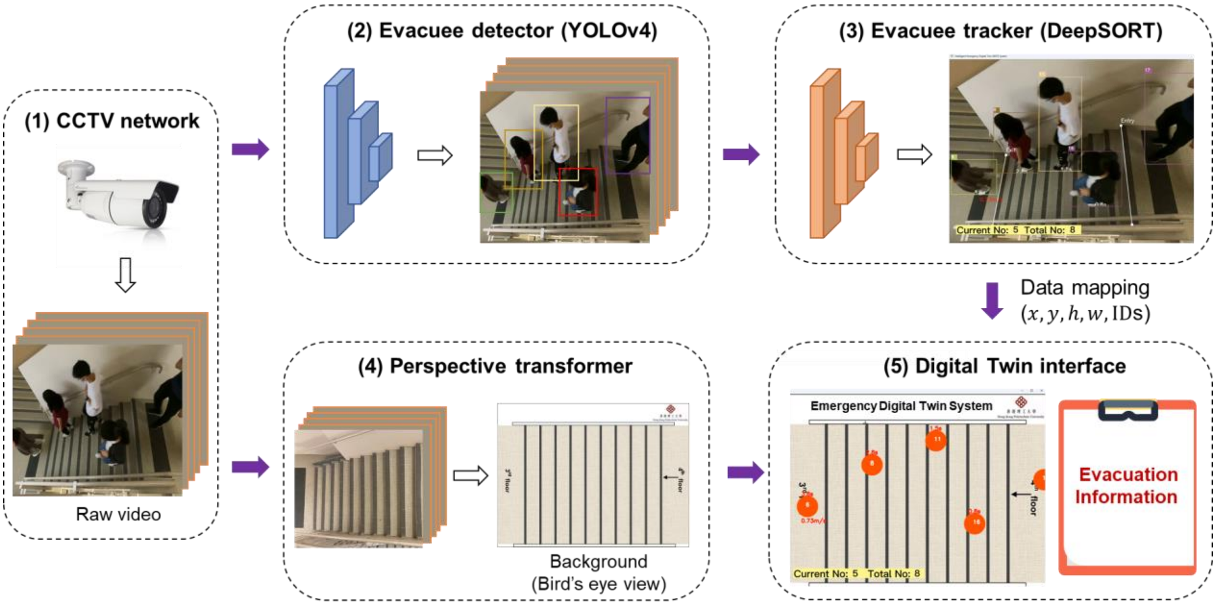


Fig. 8. Overall workflow and five key components of the Intelligent Emergency Digital Twin system for monitoring fire evacuation in real-time.

3.1. *Data obtaining*

CCTV devices are pre-installed in critical positions such as in the ceilings and handrails of the staircase along egress routes to record original videos of evacuation. The cameras and corresponding LAN-based wireless connectivity form the CCTV networks to obtain raw video data. The raw data will be stored in the central server of the digital twin and the real-time video stream will be read by the YOLOv4 detector and the DeepSORT tracker for further processing and analysis. The low latency data acquisition and transmission is a key prerequisite for outputting evacuation information in real time.

3.2. *Data processing*

Data processing is conducted by an AI engine comprised of the YOLOv4 detector and the DeepSORT tracker. The YOLOv4 detector is a deep learning object detection algorithm for positioning the evacuees with bounding boxes in the real-time video and analysing the distribution information of evacuees as well as the pixel coordinates of locations. The number of evacuees detected in each frame will also be recorded by the YOLOv4 detector. During the frame-by-frame detection analysis of the input video, the DeepSORT tracker is introduced to ensure that the evacuees' assigned IDs are constant in every frame. The speed estimator will calculate the passage time and speed of each evacuee based on the corresponding coordinates under each unique ID. After real-time data processing, the integrated evacuation information including the distribution of evacuees, the passing time and average speed of each evacuee and the count of current and total evacuees is mapped and presented on the digital twin interface. More detailed information about the digital twin presentation is given in Section 3.3.

3.3. Digital twin presentation

Digital twin interface is the core component of the proposed system to show the processed evacuation information. The background of the interface is designed as a virtual scene of an egress component in the building, for example, the staircase (see Fig. 9). Not only the dynamic traces of evacuees can be presented in the virtual interface, but it also demonstrates diverse crucial evacuation data, which helps engineers, policymakers, firefighters, etc., to acquire real-time information in evacuation site and benefits to further evacuation and rescue decision-making. Overall, the digital twin interface offers the following strengths:

(1) Real-time evacuation information assembly: Traditional surveillance interface can only transmit and record video without any data analysis capabilities. Important evacuation data and parameters of evacuee behaviours can only be processed at a later stage and are not available in providing sufficient dynamic evacuation guidance on site. In response to this issue, the proposed digital twin system combines deep learning algorithms and computer vision techniques to realise the real-time evacuation data acquisition and processing and presents the critical information in the digital twin interface, see Fig. 9. The dynamic information on the digital twin interface contains the distribution and trace of each evacuee, the time required to pass the building key egress component, the average passing speed as well as the current and total number of escaping evacuees. The monitor can visualize this information through the digital twin interface and use this information to inform the subsequent evacuation strategy development.

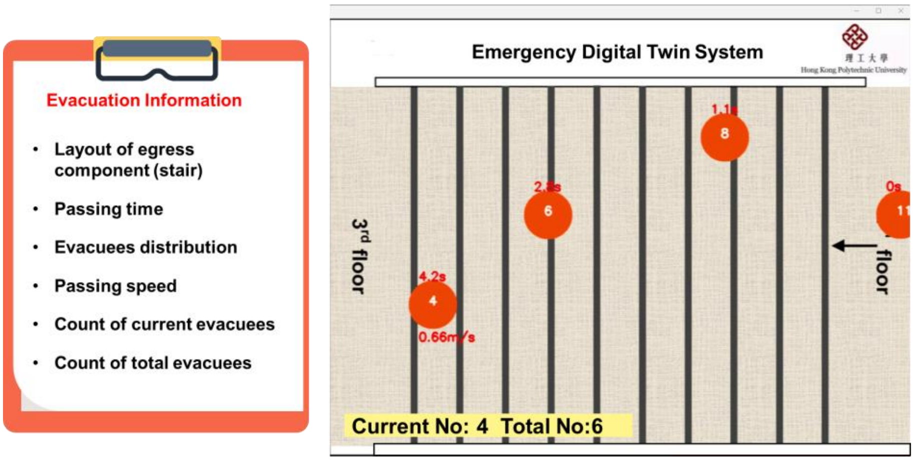


Fig. 9. Illustration of the Digital Twin interface for fire evacuation monitoring.

(2) Privacy protection: Compared to traditional monitoring systems, the evacuees' appearance would be overlooked in the proposed digital twin interface, and uniformly represented as many orange balls (see Fig. 9). Moreover, only the virtual interface would be presented as the visualisation output of the proposed digital twin system and the physical edge of original video captured by CCTV would be stored in the server rather than being played. The monitor can only observe the evacuees' traces and distribution through the virtual interface and personal information such as individual appearance would be invisible. Therefore, the occupants do not need to worry that their behaviours and privacy will be exposed if the proposed system is used during the emergency escape.

4. Experimental performance

4.1. Experiment setting

To demonstrate and test the performance of this intelligent digital twin system, three groups of pedestrian experiments were performed in a staircase of an office building located at the Hong Kong Polytechnic University, see Table 1 and Fig.10. Test 1-1 and Test 1-2 are set as repeated tests with the same conditions and only the order of participants entering the camera frame is adjusted. The experiment video streams are generated by the CCTV camera installed in the ceiling of the staircase and transferred to a local personal computer and input into the intelligent object detector and tracker.

Table 1. The experiment setting.

No.	Type	No. of people	Scenario	Description
Test 1-1	Individual object	12	Pedestrian test	Participants descended the staircase one by one and only one occupant appeared on each frame.
Test 1-2	Individual object	12	Pedestrian test	Same as Test 1-1 and only the order of participants entering the camera frame is adjusted.
Test 2-1	Multiple objects	164	Fire drill	Participants descended the staircase continuously and multiple occupants appeared on each frame.

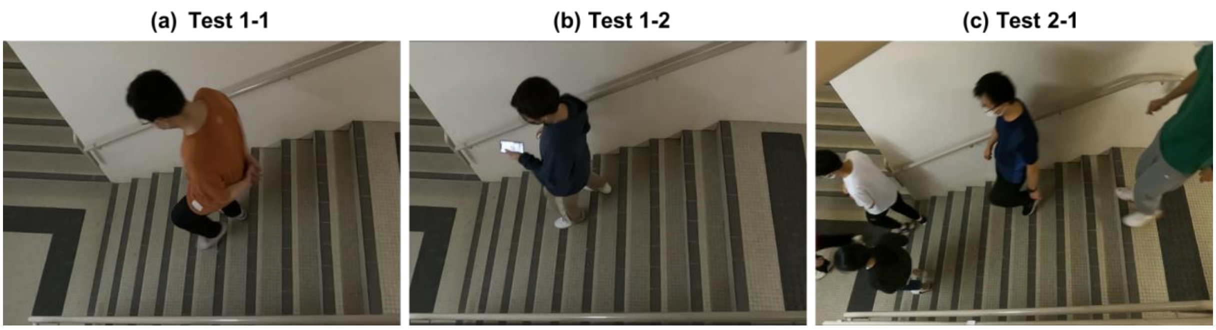


Fig. 10. Two types of experiments: Test 1-1 and Test 1-2 are individual object tests, and Test 2-1 is multiple objects test.

4.2. *Demonstration of individual object test*

In this section, we demonstrate how different components of the digital twin system work to realise individual occupant detecting and tracking as well as speed estimation, shown in Fig.11 and Video S1 in supplementary material. Firstly, the video stream was generated by a CCTV camera and transited to AI model, and the YOLOv4 detector marks the participant with blue bounding box and ID-1 precisely when he enters the CCTV vision field, and timing is not started in that time. Secondly, the timer starts immediately the participant touches the white entry line, and the orange ball appears on the visual interface. During the participant descending the stair, his ID remains constant under tracked by DeepSORT and the timer keeps on running. The timer records the time when the occupant crosses the exit line and the speed estimator calculates the average traveling speed and update interface.

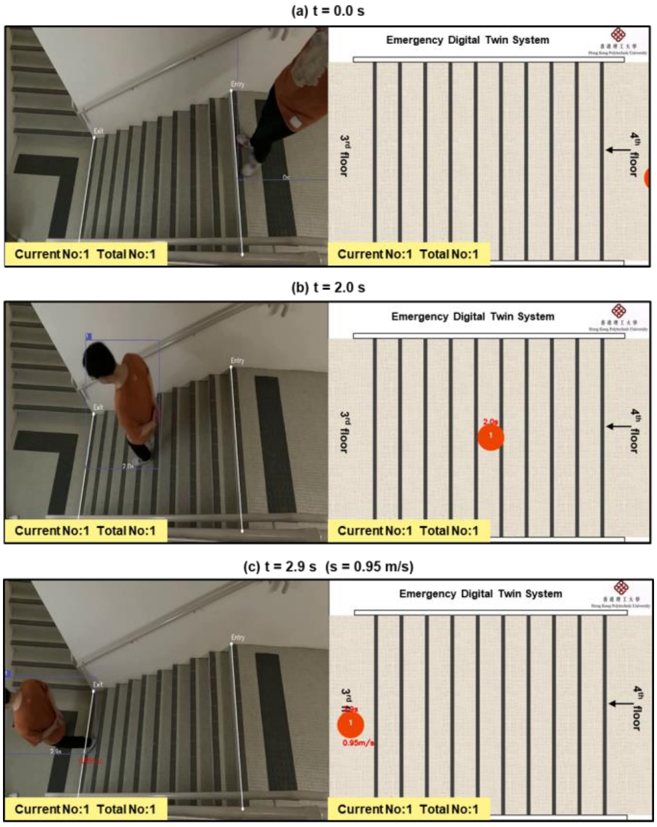


Fig. 11. Demonstration of the Digital Twin system. (a) Timer starts when the occupant touches the entry line; (b) Occupant is under-tracked; and (c) Speed is estimated when occupant crosses the exit line.

4.3. *Demonstration of multiple objects test*

The task of occupant tracking, timing and speed estimation is more difficult and complex when multiple objects are present in each frame compared to individual object tests. A combination of the YOLOv4 detector and DeepSORT multiple objects tracker can solve the problems of objects loss and IDs change. While the method of continuous timing in the whole experiment process to estimate the traveling speed of each occupant is not suitable. To address this issue, we assign a separate timer to each participant and associate it with the respective ID of each participant.

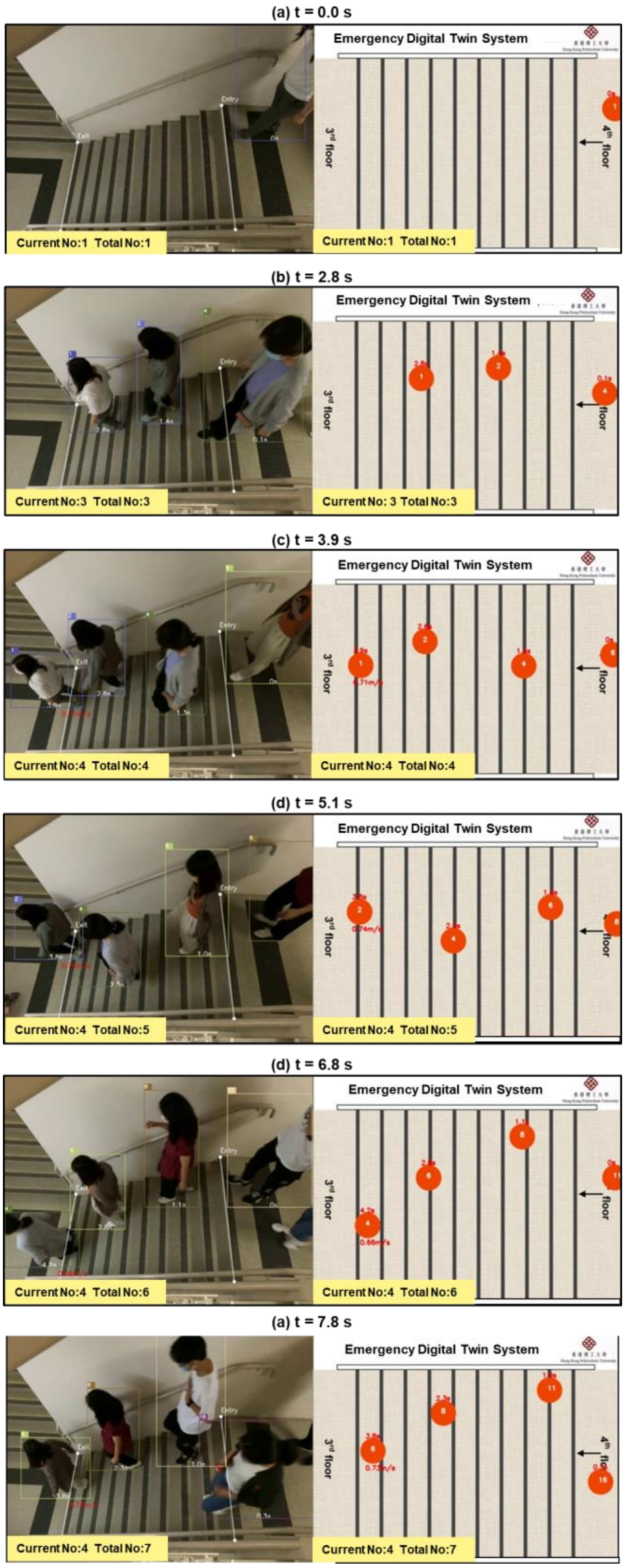


Fig. 12. Demonstration of the flowchart of the digital twin monitoring system for multiple objects test. (a) The detector marks the first occupant, and the timer starts once she touches the entry line; (b) Occupants are tracked and the respective timers are working separately; (c) 1st participant crosses the exit line in 3.9 s; (d) 2nd participant crosses the exit line in 5.1 s; (e) 3rd participant crosses the exit line in 6.8 s; (f) 4th participant crosses the exit line in 7.8 s.

In other words, the timer keeps track of each participant's individual traveling time, and it would not restart if their IDs have not switched. As each participant passes near the exit line, their respective timers stop and record individual traveling time and average traveling speed. The values of total traveling time and average speed will also be mapped to the right visual interface, shown in Fig. 12 and Video S2 in supplementary material. For example, when t=3.9 s, the first participant crosses the exit line with 3.9 s, and the speed estimator calculates the traveling speed as 0.71 m/s. When t=5.1 s, the second participant crosses the exit line with 3.8 s, and the speed estimator calculates the traveling speed as 0.74 m/s.

4.4. The performance of the digital twin system

Table 2 demonstrates the success ratios of two types of experiments. The ratio of successful detection equals the number of detected participants over the total number and illustrates the detecting performance of YOLOv4 detector. The ratio of tracking successfully illustrates how many occupants can be tracked during the whole test process. The tracking is considered as a failure once the ID switching occurs before the participant crosses the entry line. The ratio of estimating speed successfully is to evaluate how many participants' average speed can be calculated and shown on the virtual view.

Table 2. Success ratios or accuracy of evacuee detection, tracking, and speed estimation.

No.	Success ratios or accuracy		
	Detection	Tracking	Speed
Test 1-1	100%	92%	83%
Test 1-2	100%	83%	83%
Test 2-1	93%	84%	75%

This system performed high accuracy and tracking performance in individual object test. For Test 1-1, all Test 1-2 participants were detected and tracked by the AI model during the experiment, and the accuracy is 100%. While two participants' data of average traveling speed were not calculated by the speed estimator because the model did not store the coordinate data of occupants' position when the participant entered the staircase which leads the timer failure, or the timer closed in advance before the participant passed the staircase caused by the coordinate data loss. For Test 1-2, all 12 participants were also detected successfully, while two participants failed to be tracked, which resulted in speed data loss.

Though the performance of the proposed system for the multiple-objects test was also high with a 93% detection ratio, 84% tracking ratio and 75% speed estimation ratio. it was reduced compared to the individual test. In specific, 11 participants were not detected, 26 participants were not tracked, and 40 participants' speeds were not estimated. The main reason for the decrease in performance was that these four participants' body parts were overlapped by other evacuees during the experiment. The overlapped phenomenon causes the convolutional neural network model in the YOLO detector loses features of potential objects which is further likely to cause inaccurate detection and tracking.

Note that the criterion for determining whether a pedestrian touches the entry or exit line is that the distance between the pedestrian coordinates and the white line is less than a certain threshold. There are cases where even though the pedestrian crosses the entry or exit line, but the distance is not within the determination interval. In that case, the timer would not work, and the speed estimation would be a failure. As a result, how to solve the problem of occlusion during detection and determine the distance threshold is the key to improving the performance of the digital twin system. In the future developments of the digital twin system, solving the overlapping problem and increasing the accuracy of tracking and speed estimation would be the primary points.

5. Conclusions

In this paper, a framework of an intelligent evacuation digital twin system is proposed based on deep learning and computer vision and demonstrated with two types of experiments. The proposed framework is mainly constructed with a CCTV network, YOLOv4 detector and DeepSORT tracker and a digital twin interface. Overall, the digital twin monitoring system can provide more detailed and real-time feedback compared with the conventional monitoring system relying on CCTV network and the information would be likely to instruct the evacuation strategy making and update. The main contributions and conclusions of this work are as follows:

(1) The proposed digital twin monitoring system can detect and track evacuees based on the deep learning YOLO-DeepSORT algorithm.

(2) A speed estimation algorithm is developed and embedded in the proposed system to automatically calculate the evacuee's moving speed during traveling the critical building component such as the staircase.

(3) A digital twin interface is designed based on inverse perspective transformation, and the evacuation data extracted by the smart detector, tracker and speed estimator can be mapped into the virtual interface to integrate diverse real-time evacuation information and the virtual representation of evacuee's trace allowing for occupants' privacy protection to some extent.

(4) The performance of the digital twin monitoring system is tested with two types of pedestrian experiments. The results of the three tests achieved a success rate: ratio of successfully detecting is 100% in individual objects test and 93% in multiple objects test; ratio of successfully tracking both in two types of test achieved over 83%. The result shows that both individual and multiple occupants' scenarios are well adapted to this system.

The evacuation data generated by the digital twin system would be useful for guiding evacuations out of fire scenarios. This proposed digital twin framework lays the foundation for the implementation of smart monitoring of human behaviours in fire scenarios for buildings. In the future work, more fire drills with crowed and mixed population and real fire rescue scenarios with the fire service department will be used to test the proposed digital twin system. Moreover, more advanced AI algorithm and evacuation strategies to focus on undetected and untracked peoples, and capture evacuee's irrational emotions and mixed human groups and behaviours during fire evacuation will be developed and applied to upgrade the digital twin system.

CRediT authorship contribution statement

Yifei Ding: Investigation, Methodology, Writing - original draft, Formal analysis.

Yuxin Zhang: Investigation, Conceptualization, Writing - review & editing.

Xinyan Huang: Conceptualization, Supervision, Writing - review & editing, Funding acquisition.

Acknowledgments:

This work is funded by the Hong Kong Research Grants Council Theme-based Research Scheme (T22-505/19-N) and the National Natural Science Foundation of China (52204232).

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.

References

[1] M. Kobes, I. Helsloot, B. de Vries, J.G. Post, Building safety and human behaviour in fire: A literature review, Fire Safety Journal. 45 (2010) 1–11. https://doi.org/10.1016/j.firesaf.2009.08.005.

[2] D. Tong, D. Canter, The decision to evacuate: a study of the motivations which contribute to evacuation in the event of fire, Fire Safety Journal. 9 (1985) 257–265. https://doi.org/10.1016/0379-7112(85)90036-0.

[3] E. Galea, High-Rise Building Evacuation Post 911 - Addressing the Issues, in: Tall Building Fire Safety Conference, London, 2014: pp. 75–89. https://fseg.gre.ac.uk/fire/fseg_tall_building_paper_ed4_distrib.pdf.

[4] Remembering 9/11 on 16th anniversary:These images of Twin Towers attack still haunt the world - IBTimes India, (n.d.). https://www.ibtimes.co.in/remembering-9-11-16th-anniversarythese-images-twin-towers-attack-still-haunt-world-741628.

[5] H. Leung, Deaths of 2 Hong Kong firefighters in 2016 storage unit blaze ruled as a result of misadventure, (2022). https://hongkongfp.com/2022/05/11/deaths-of-2-hong-kong-firefighters-in-2016-storage-unit-blaze-ruled-as-a-result-of-misadventure/.

[6] Y. Zhang, X. Huang, A Review of Tunnel Fire Evacuation Strategies and State-of-the-Art Research in China, Fire Technology. (2022) 1–34. https://doi.org/10.1007/S10694-022-01357-5/FIGURES/17.

[7] Y. Zhang, W. Li, Y. Rui, S. Wang, H. Zhu, Z. Yan, A modified cellular automaton model of pedestrian evacuation in a tunnel fire, Tunnelling and Underground Space Technology. 130 (2022) 104673. https://doi.org/10.1016/J.TUST.2022.104673.

[8] F. Huo, W. Song, L. Chen, C. Liu, K.M. Liew, Experimental study on characteristics of pedestrian evacuation on stairs in a high-rise building, Safety Science. 86 (2016) 165–173. https://doi.org/10.1016/j.ssci.2016.02.025.

[9] K. Aghabayk, A. Soltani, N. Shiwakoti, Investigating Pedestrians’ Exit Choice with Incident Location Awareness in an Emergency in a Multi-Level Shopping Complex, Sustainability. 14 (2022). https://doi.org/10.3390/su141911875.

[10] Y. Zhang, H. Zhu, Q. Guo, R. Carvel, Z. Yan, The effect of technical installations on evacuation performance in urban road tunnel fires, Tunnelling and Underground Space Technology. 107 (2021). https://doi.org/10.1016/j.tust.2020.103608.

[11] X. Huang, X. Wu, A. Usmani, Perspectives of Using Artificial Intelligence in Building Fire Safety, in: M.Z. Naser (Ed.), Handbook of Cognitive and Autonomous Systems for Fire Resilient Infrastructures, Springer, New York, 2022. https://doi.org/10.1007/978-3-030-98685-8_6.

[12] J. Wang, W.C. Tam, Y. Jia, R. Peacock, P. Reneke, E.Y. Fu, T. Cleary, P-Flash – A machine learning-based model for flashover prediction using recovered temperature data, Fire Safety Journal. 122 (2021) 103341. https://doi.org/10.1016/j.firesaf.2021.103341.

[13] T. Zhang, Z. Wang, H.Y. Wong, W.C. Tam, X. Huang, F. Xiao, Real-time forecast of compartment fire and flashover based on deep learning, Fire Safety Journal. 130 (2022) 103579. https://doi.org/10.1016/j.firesaf.2022.103579.

[14] X. Wu, Y. Park, A. Li, X. Huang, F. Xiao, A. Usmani, Smart Detection of Fire Source in Tunnel Based on the Numerical Database and Artificial Intelligence, Fire Technology. 57 (2021) 657–682. https://doi.org/10.1007/s10694-020-00985-z.

[15] T. Zhang, Z. Wang, Y. Zeng, X. Wu, X. Huang, F. Xiao, Building Artificial-Intelligence Digital Fire (AID-Fire) system: A real-scale demonstration, Journal of Building Engineering. 62 (2022) 105363. https://doi.org/10.1016/J.JOBE.2022.105363.

[16] X. Wu, X. Zhang, Y. Jiang, X. Huang, G.G.Q. Huang, A. Usmani, An intelligent tunnel firefighting system and small-scale demonstration, Tunnelling and Underground Space Technology. 120 (2022) 104301. https://doi.org/10.1016/j.tust.2021.104301.

[17] C. Xiong, Z. Wang, Y. Huang, F. Shi, X. Huang, Smart evaluation of building fire scenario and hazard by attenuation of alarm sound field, Journal of Building Engineering. 51 (2022) 104264. https://doi.org/10.1016/j.jobe.2022.104264.

[18] Z. Wang, T. Zhang, X. Wu, X. Huang, Predicting transient building fire based on external smoke images and deep learning, Journal of Building Engineering. 47 (2022) 103823. https://doi.org/10.1016/j.jobe.2021.103823.

[19] Z. Wang, T. Zhang, X. Huang, Predicting real-time fire heat release rate by flame images and deep learning, Proceedings of the Combustion Institute. 39 (2023) 4115–4123. https://doi.org/10.1016/j.proci.2022.07.062.

[20] X. Wu, X. Zhang, X. Huang, F. Xiao, A. Usmani, A real-time forecast of tunnel fire based on numerical database and artificial intelligence, Building Simulation. 15 (2022) 511–524. https://doi.org/10.1007/s12273-021-0775-x.

[21] M. Li, Y. Jiang, Z. Wu, R. Fan, Real-Time Prediction of Smoke Spread Affected by Multiple Factors in Subway Tunnel Using CAERES-DNN Model, Fire Technology. 57 (2021) 2025–2059. https://doi.org/10.1007/s10694-021-01109-x.

[22] C. Guo, Q. Guo, T. Zhang, W. Li, H. Zhu, Z. Yan, Study on real-time heat release rate inversion for dynamic reconstruction and visualization of tunnel fire scenarios, Tunnelling and Underground Space Technology. 122 (2022) 104333. https://doi.org/10.1016/j.tust.2021.104333.

[23] M. Jiang, Y. Zhao, F. Yu, C. Zhou, T. Peng, A self-attention network for smoke detection, Fire Safety Journal. 129 (2022) 103547. https://doi.org/10.1016/J.FIRESAF.2022.103547.

[24] D.P. Lestari, R. Kosasih, T. Handhika, Murni, I. Sari, A. Fahrurozi, Fire Hotspots Detection System on CCTV Videos Using You only Look Once (YOLO) Method and Tiny YOLO Model for High Buildings Evacuation, Proceedings - 2019 2nd International Conference of Computer and Informatics Engineering: Artificial Intelligence Roles in Industrial Revolution 4.0, IC2IE 2019. (2019) 87–92. https://doi.org/10.1109/IC2IE47452.2019.8940842.

[25] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, 2016.

[26] Y. Chen, S. Hu, H. Mao, W. Deng, X. Gao, Application of the best evacuation model of deep learning in the design of public structures, Image and Vision Computing. 102 (2020) 103975. https://doi.org/10.1016/J.IMAVIS.2020.103975.

[27] G. Zhang, D. Lu, H. Liu, IoT-Based Positive Emotional Contagion for Crowd Evacuation, IEEE Internet of Things Journal. 8 (2021) 1057–1070. https://doi.org/10.1109/JIOT.2020.3009715.

[28] S. Li, L. Tong, C. Zhai, Extraction and modelling application of evacuation movement characteristic parameters in real earthquake evacuation video based on deep learning, International Journal of Disaster Risk Reduction. 80 (2022) 103213. https://doi.org/10.1016/J.IJDRR.2022.103213.

[29] E. Taylor, I Spy with My Little Eye: The Use of CCTV in Schools and the Impact on Privacy, The Sociological Review. 58 (2010) 381–405. https://doi.org/10.1111/J.1467-954X.2010.01930.X.

[30] B.J. Goold, Privacy rights and public spaces: CCTV and the problem of the “unobservable observer,” Criminal Justice Ethics. 21 (2010) 21–27. https://doi.org/10.1080/0731129X.2002.9992113.

[31] M. Grieves, J. Vickers, Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems, Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. (2016) 85–113. https://doi.org/10.1007/978-3-319-38756-7_4/FIGURES/7.

[32] X. Yuan, C.J. Anumba, Cyber-Physical Systems for Temporary Structures Monitoring, Cyber-Physical Systems in the Built Environment. (2020) 107–138. https://doi.org/10.1007/978-3-030-41560-0_7.

[33] D. Wang, J. Chen, D. Zhao, F. Dai, C. Zheng, X. Wu, Monitoring workers’ attention and vigilance in construction activities through a wireless and wearable electroencephalography system, Automation in Construction. 82 (2017) 122–137.

[34] H.H. Hosamo, P.R. Svennevig, K. Svidt, D. Han, H.K. Nielsen, A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics, Energy and Buildings. 261 (2022) 111988. https://doi.org/10.1016/j.enbuild.2022.111988.

[35] N. Elhami-Khorasani, J.G. Salado Castillo, T. Gernay, A Digitized Fuel Load Surveying Methodology Using Machine Vision, Fire Technology. 57 (2021) 207–232. https://doi.org/10.1007/s10694-020-00989-9.

[36] N. Elhami-Khorasani, J.G. Salado Castillo, E. Saula, T. Josephs, G. Nurlybekova, T. Gernay, Application of a Digitized Fuel Load Surveying Methodology to Office Buildings, Fire Technology. (2020). https://doi.org/10.1007/s10694-020-00990-2.

[37] L. Jiang, J. Shi, C. Wang, Z. Pan, Intelligent control of building fire protection system using digital twins and semantic web technologies, Automation in Construction. 147 (2023) 104728. https://doi.org/10.1016/j.autcon.2022.104728.

[38] S.H. Khajavi, M. Tetik, Z. Liu, P. Korhonen, J. Holmström, Digital twin for Safety and Security: Perspectives on Building Lifecycle, IEEE Access. (2023).

[39] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: pp. 580–587.

[40] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: European Conference on Computer Vision, Springer, 2016: pp. 21–37.

[41] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection, ArXiv Preprint ArXiv:2004.10934. (2020).

[42] N. Wojke, A. Bewley, D. Paulus, Simple online and realtime tracking with a deep association metric, in: 2017 IEEE International Conference on Image Processing (ICIP), 2017: pp. 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962.

[43] C.-Y. Wang, H.-Y.M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, I.-H. Yeh, CSPNet: A new backbone that can enhance learning capability of CNN, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: pp. 390–391.

[44] K. He, X. Zhang, S. Ren, J. Sun, Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence. 37 (2015) 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.

[45] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path Aggregation Network for Instance Segmentation, (2018) 8759–8768.

[46] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, ArXiv Preprint ArXiv:1804.02767. (2018).

[47] A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Upcroft, Simple online and realtime tracking, in: 2016 IEEE International Conference on Image Processing (ICIP), 2016: pp. 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003.

[48] H.W. Kuhn, The Hungarian method for the assignment problem, Naval Research Logistics Quarterly. 2 (1955) 83–97.

[49] R.E. Kalman, A new approach to linear filtering and prediction problems, (1960).

[50] K. Aghabayk, N. Parishad, N. Shiwakoti, Investigation on the impact of walkways slope and pedestrians physical characteristics on pedestrians normal walking and jogging speeds, Safety Science. 133 (2021) 105012.