Automation in Construction 174 (2025) 106108

Contents lists available at ScienceDirect

Automation in Construction

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

Enhancing worker monitoring and management on large-scale construction sites with UAVs and digital twin modeling

Mingqiao Han, Jihan Zhang *, Yijun Huang, Jiwen Xu, Xi Chen *, Ben M. Chen

Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, N.T, Hong Kong Special Administrative Region of China

ARTICLE INFO

Keywords:
Unmanned aerial vehicle (UAV)
Digital twin (DT)
Smart cities
Site monitoring
Project management

ABSTRACT

Monitoring large-scale work sites is challenging, particularly in vast outdoor areas. Unmanned aerial vehicles (UAVs) provide an effective solution for site monitoring and worker management. This paper introduces a UAV-based framework integrated with digital twin (DT) modeling to enhance real-time data management and worker authorization verification. The pretrained YOLO-LCA model improved detection accuracy from 31.5% to 96.4%. The framework combines multi-object tracking with 3D site reconstruction, enabling precise global registration and situational awareness. Cross-referencing UAV detections with GPS-enabled worker IDs ensures that only authorized personnel are present, effectively identifying unapproved workers. The proposed framework has undergone large-scale validation across multiple construction projects in Hong Kong, demonstrating significant potential for modernizing work site management. By integrating UAVs and DT technology, this framework supports efficient monitoring, operational safety, and informed decision-making, providing a scalable approach to addressing the demands of large-scale construction site management.

1. Introduction

In large-scale outdoor working environments within urban contexts, managing worker authorization presents significant challenges due to the vast and open nature of such sites. Ensuring that only authorized, properly trained personnel are present on-site is critical not only for operational efficiency but also for safety, especially in densely populated urban areas. Unauthorized workers typically lack the necessary safety training and may be unfamiliar with the site's specific hazards, increasing the risk of accidents. This is of particular concern in urban environments like open-pit mines [1], wind farms [2], and large infrastructure projects such as highways [3] and railways [4], where real-time monitoring of workers is essential to prevent incidents and maintain safety across vast and complex city landscapes. While smaller, enclosed sites allow for straightforward worker identity verification through controlled access points or physical security measures [5]. However, the sheer scale and openness of urban-level work sites make traditional methods like manual patrolling or closed-circuit television (CCTV) surveillance impractical and costly [6]. For example, mines situated in city outskirts or construction zones with dynamic excavation areas pose continuous worker monitoring challenges [7]. Similarly, wind farms and large-scale burial sites require real-time monitoring of workers, but the vast geographical spread of these projects limits the effectiveness of fixed surveillance methods [8].

The application of unmanned aerial vehicles (UAVs) in large-scale urban site monitoring has proven to be a flexible and sustainable solution, effectively addressing the limitations of traditional surveillance methods such as fixed CCTV systems, which are often restricted by their limited coverage areas and lack of adaptability to dynamic site layouts [9,10]. UAVs offer unmatched versatility in covering extensive and constantly evolving urban construction environments, enabling real-time aerial monitoring of site progress, worker activities, and environmental conditions. Unlike mobile 360-degree cameras, which require autonomous ground vehicles for mobility and are limited by terrain accessibility, UAV can operate fully autonomously without requiring operator intervention during routine inspections. Furthermore, UAVs are not constrained by terrain, making them highly adaptable to complex and uneven construction environments. This enhances both operational safety and compliance with urban planning regulations by facilitating automated worker verification through integrated Global Positioning System (GPS) tracking systems [11,12]. However, despite these advantages, the deployment of UAVs in construction projects introduces new challenges, particularly in managing and processing the vast amounts of image data generated by UAV flights. In addition, for larger-scale scenarios where single UAV battery life may pose a limitation, the site should be divided into smaller sub-regions for sequential inspections or employing multiple UAVs to enable simultaneous coverage of different zones, which may addressed the battery

* Corresponding authors.
E-mail addresses: 1155139089@link.cuhk.edu.hk (J. Zhang), xichen002@cuhk.edu.hk (X. Chen).

https://doi.org/10.1016/j.autcon.2025.106108
Received 6 November 2024; Received in revised form 20 February 2025; Accepted 28 February 2025

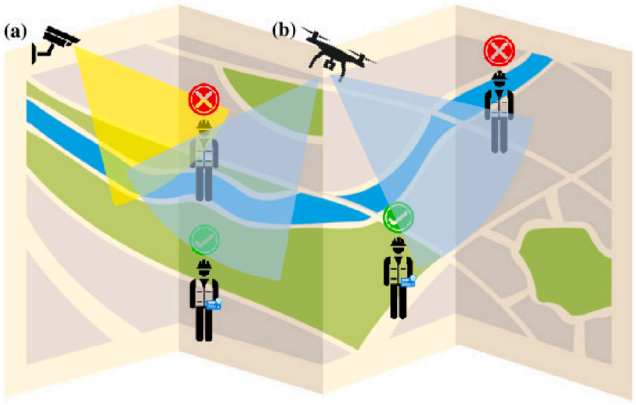

Fig. 1. Traditional monitoring methods.

limitation for single UAV. To meet the sustainability and efficiency goals emphasized in urban monitoring systems, sophisticated computer vision algorithms are necessary to process these large datasets effectively, requiring advancements in image processing techniques to ensure accurate detection and tracking of site activities and personnel in real time [13,14].

Advancements in computer vision and deep learning algorithms have greatly enhanced sustainable urban construction monitoring by automating the detection, tracking, and analysis of site activities in real-time. UAV-based systems, while offering flexibility, face significant challenges due to the dynamic nature of large, open environments. Unlike fixed-camera setups with known spatial coordinates, UAVs must contend with constantly changing positions, making real-time localization and accuracy difficult. Issues like GPS interruptions and signal interferences in dense urban environments further complicate the monitoring process, where precision is critical for tracking personnel and equipment across vast construction areas [15]. Moreover, one of the critical challenges in UAV-based site monitoring is the lack of specialized datasets for large-scale, open environments. Existing deep learning models are often trained on datasets that are not representative of the complex, ever-changing conditions encountered in outdoor construction environments, such as varying lighting and terrain. This discrepancy reduces the effectiveness of these models when applied to real-world scenarios [16,17]. The development of comprehensive datasets tailored for open-site object detection, like the SODA dataset, is necessary to improve the performance and generalizability of UAV-based monitoring systems in construction projects [18]. Addressing these issues requires further refinement of algorithms and robust datasets to better handle the complexities of UAV data in dynamic urban-level construction environments.

Traditional monitoring methods, as illustrated in Fig. 1(a), are often limited in coverage area, whereas drones, shown in Fig. 1(b), offer comprehensive site monitoring capabilities. However, challenges remain in verifying worker permissions solely through drone surveillance, highlighting the need for a more robust solution that can authenticate worker authorization while continuously updating worker distribution across the site. Recent digital twin (DT) research has gained increasing prominence in working site monitoring due to its ability to seamlessly integrate multi-source 2D and 3D data and provide real-time updates [19]. Faced with these conditions, we introduce a UAV-enabled surveillance and management framework that leverages digital twin (DT) technology for real-time site monitoring. In this context, surveillance refers to UAVs capturing real-time video streams of the construction site and transmitting them to a centralized management platform. This allows site managers to remotely monitor all work regions, eliminating the need for manual inspections in large-scale construction environments. However, real-time video surveillance alone does not guarantee that only authorized personnel are present on-site, necessitating an additional management mechanism that integrates worker authorization verification with automated data processing. Therefore, we incorporates a worker authorization system, which cross-references UAV-detected personnel with GPS-based electronic work permits to verify their legitimacy. Additionally, the system aggregates and visualizes statistical data on both authorized and unauthorized workers for better management, enabling managers to make informed workforce allocation and compliance decisions. By integrating real-time UAV surveillance with intelligent worker management, this approach not only ensures continuous situational awareness but also facilitates proactive interventions. Through on-site validation, this method proves to be a scalable and efficient solution for modern site management, supporting enhanced safety enforcement and automated workforce surveillance.

2. Related works

2.1. UAV in construction monitoring

UAVs have emerged as a transformative tool in construction monitoring due to their ability to provide real-time, comprehensive data collection, high flexibility, and extensive coverage across various construction environments [20]. UAVs offer significant advantages over traditional monitoring methods, such as manual inspections and fixed-camera surveillance systems, particularly in terms of efficiency, accessibility, and safety [21]. By enabling aerial views and integrating with advanced imaging technologies like LiDAR and photogrammetry, UAVs can rapidly collect large amounts of data that can be processed to monitor site progress, worker activity, and material usage [22].

In urban and confined construction projects, UAVs have achieved considerable success [23,24]. They are widely used for progress tracking [25], wet bulb globe temperature (WBGT) prediction for climate adaptation [26], safety inspections [13], and structural assessments [27], where high-resolution imagery and precise mapping are crucial [28]. By reducing human involvement in dangerous environments, UAVs enhance safety while promoting resilient, low-impact construction workflows. UAVs offer a safer and faster alternative to manual inspections, reducing the need for workers to access hazardous areas [29], further supporting urban sustainability goals by enhancing resource allocation and reducing environmental footprints.

As UAVs have proven effective in smaller, confined sites, their application has expanded into large-scale outdoor and open construction environments, such as open-pit mines [30], wind farms, and highway projects [31,32]. In these vast and dynamic environments, UAVs offer unparalleled coverage and flexibility, providing real-time monitoring of extensive areas that are otherwise impractical for traditional surveillance methods [33]. UAVs are now being used for topographical mapping, 3D modeling, environmental assessments, and real-time worker and equipment tracking [34], contributing to sustainable city management by integrating environmental monitoring with infrastructure development. The ability to monitor large distances, shifting terrain, and evolving site conditions makes UAVs essential for ensuring operational efficiency and maintaining safety standards across expansive construction sites, contributing to the resilience and sustainability of urban landscapes.

The deployment of UAVs in large-scale, open construction environments presents significant challenges. The vast size and complexity of such sites generate enormous volumes of data, demanding advanced processing capabilities and sophisticated algorithms for timely and accurate analysis. UAVs must operate in dynamic, uncontrolled environments where factors such as complex terrain, variable weather conditions, and shifting lighting can hinder data capture and compromise system reliability. Their mobility further complicates real-time localization and data alignment due to constant adjustments in position

and orientation. Integrating UAV data into broader monitoring frameworks is particularly challenging, as construction sites require continuous adaptation and robust data synchronization to maintain accuracy. While UAVs hold substantial potential, their effectiveness in large-scale construction relies on further advancements in data processing, environmental adaptability, and operational strategies.

2.2. Object detection in UAV application

Object detection methods in UAV applications can be broadly categorized into single-stage and two-stage approaches. Single-stage methods, such as YOLOv10 and SSD, are known for their real-time performance and low computational requirements, making them highly suitable for applications that demand fast inference [35,36]. In contrast, two-stage methods, such as Faster R-CNN, improve detection accuracy by using a region proposal network (RPN), although this comes at the cost of slower inference speeds [37,38]. Transformer-based models like DETR aim to enhance real-time performance by replacing the RPN with an end-to-end detection approach, offering high accuracy but requiring more computational resources [39,40].

While these algorithms perform well on standard benchmark datasets like Pascal VOC 2007 and COCO, they often face limitations when applied to UAV-based detection. UAV images tend to have large background areas with complex features, while the target objects (e.g., workers or equipment) occupy only a small portion of the image, making detection more challenging [41,42]. Furthermore, UAVs are constrained by limited endurance and computational capacity, which require detection models that are both computationally efficient and capable of fast inference [43,44]. The critical challenge in UAV-based detection tasks is finding the balance between detection precision and real-time performance. As such, models like YOLO, with their minimal computational overhead and fast inference speeds, are often preferred for UAV inspection tasks.

However, even YOLO models, typically pre-trained on conventional datasets, may not align well with the unique operational conditions of UAV-based inspections. UAVs capture images from an overhead perspective, which results in distinct object features for the same categories and challenges like object occlusion [41,44]. This leads to the need for retraining these models on specialized datasets tailored to the specific context of UAV inspections. Currently, datasets specifically designed for UAV perspectives or site inspection environments — such as UAVDT [45], VisDrone [46], and UAV123 [47] — remain limited, often lacking generalizability across diverse operational environments [48,49]. While these datasets provide important benchmarks for certain UAV tasks, their applicability in broader, more complex real-world scenarios remains constrained. As a result, there is a growing need to develop task-specific datasets that are customized to meet the unique challenges of UAV-based inspections.

In terms of applications, many studies have explored UAV-based object detection, though most have focused on specific tasks like crack detection. For instance, He et al. [50] developed MUENet based on YOLOX-S for road crack detection from a UAV perspective, while Ding et al. [51] deployed a transformer model on UAVs to detect building defects. Xiao et al. [52] employed point cloud segmentation and projection methods for bridge defect detection. On the other hand, while some studies have applied object detection for worker monitoring, these are often designed for indoor environments or rely on CCTV systems [53,54]. While these systems perform well in controlled indoor settings, their reliability in large-scale, open outdoor environments is limited. In conclusion, while UAV-based object detection has proven to be a powerful tool in construction and site monitoring, several challenges persist, particularly in large-scale outdoor scenarios. The need for specialized datasets, computational efficiency, and task-specific model adaptations remains critical to improving the reliability and performance of these systems in dynamic, real-world environments.

2.3. DT-based working site management

DT technology has emerged as an integral solution for sustainable urban site monitoring, offering a real-time, high-fidelity digital representation of physical sites by combining data from diverse sources such as building information modeling (BIM), sensors, and AI-driven systems [55]. DTs allow for comprehensive monitoring by integrating environmental, operational, and worker data into a cohesive model, providing a global overview of site activities [54]. This enables stakeholders to access real-time updates, track worker activities, and make informed decisions based on current site conditions [19]. By connecting UAV-generated aerial data with operational data — such as worker location and equipment usage — DT enhances not only site visualization but also the task assignment and resource allocation [56]. As a result, DTs facilitate more efficient and safer workflows, optimizing worker movements, material distribution, and machinery operation across large and dynamic working sites [57].

UAVs hold significant potential for enhancing DT applications in urban construction, but their integration presents several challenges. While UAVs provide real-time aerial data, capturing detailed images and videos of worker activities and site progress, the dynamic nature of construction sites complicates the integration of 2D visual data into 3D models for identity verification. The constantly changing perspective of the drone makes it difficult to align and track workers' precise locations and activities, complicating the matching of dynamic visual data with static site models [58]. Although AI-driven image analysis has been explored to manage the large volumes of UAV-generated data, integrating this information into a DT framework for real-time worker tracking and authorization remains a complex issue. Furthermore, synchronizing UAV imagery with other critical metrics — such as worker movements, real-time positioning, and equipment usage — poses additional challenges, particularly in converting discrete time-stamped data into continuous measures of worker activity for accurate verification [59,60]. These limitations restrict UAVs from delivering a fully integrated monitoring system for worker identity and safety, which is essential for ensuring compliance and security on large-scale construction sites.

2.4. Contribution

Despite advancements in UAV-based construction monitoring, object detection, and DT-driven site management, existing methods struggle with real-time data processing, adaptability, and seamless DT integration in large-scale, dynamic environments. UAV monitoring faces challenges in worker tracking, while object detection models often fail in complex outdoor settings due to occlusion and scale variation. Additionally, DT-based management lacks real-time worker tracking and authorization verification, limiting its effectiveness in site-wide compliance monitoring. These challenges, as discussed in related works, underscore the need for an integrated approach that enhances UAV surveillance and worker management within DT frameworks. To address the challenges of data integration and alignment in large-scale urban construction, this study presents a UAV-supported DT modeling method that advances construction site worker surveillance and management. In this framework, surveillance is achieved through UAVs capturing real-time video streams and transmitting them to a centralized platform, allowing site managers to oversee all work regions remotely without the need for manual patrols. Meanwhile, management is realized through a worker authorization system that cross-references UAV detections with GPS-based electronic work permits, ensuring that only authorized personnel are present on-site. By combining real-time monitoring with proactive worker supervision, our approach offers a more comprehensive solution for large-scale construction site management. The contribution of our research is reflected in the following aspects. (1) By leveraging AI-driven visual detection, our method continuously aligns 2D UAV imagery with the 3D site model, effectively

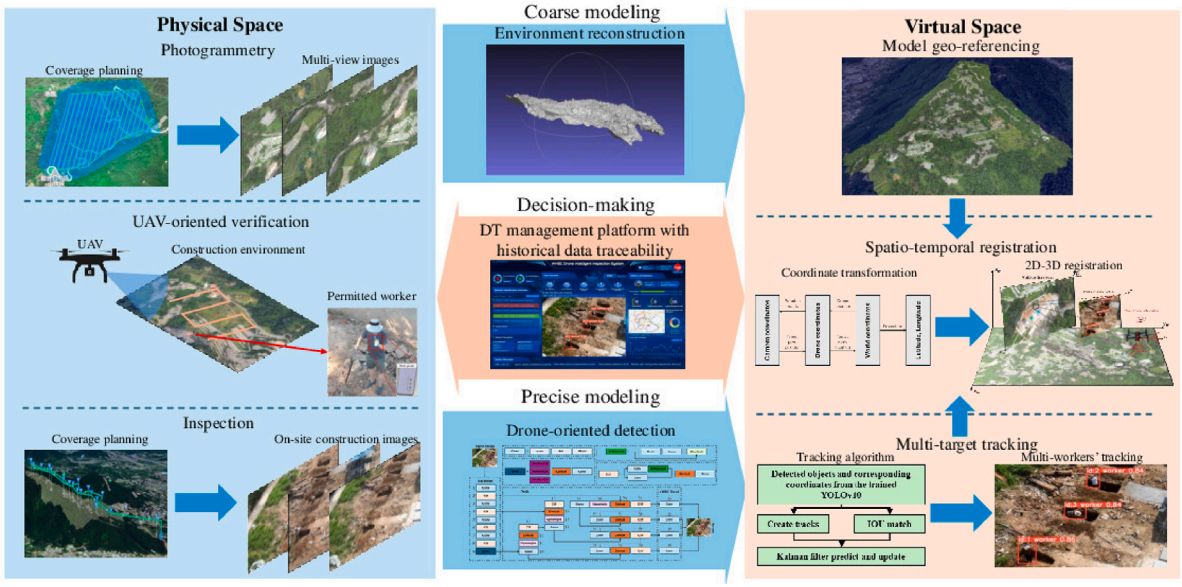

Fig. 2. Overall framework for monitoring and management of large-scale urban construction sites through interaction of physical and virtual spaces.

solving the 2D–3D matching problem to improve spatial registration accuracy in dynamic urban scenarios. This alignment resolves persistent temporal-spatial data issues in DT applications, ensuring real-time worker movements are accurately represented within the DT model. (2) We propose the pre-trained YOLOv10-LCA model based on YOLOv10-S model, which enhances large-scale worker authorization verification. This model overcomes the limitations of traditional UAV detection algorithms in outdoor environments, achieving substantial gains in detection precision and real-time performance to improve site safety and efficiency. (3) Our GPS-based worker authorization verification method identifies unauthorized personnel and integrates this information into a centralized smart city management platform, offering an intelligent, comprehensive solution for construction site management. Altogether, these advancements provide a cohesive and efficient decision-making tool, supporting modern site monitoring and management while significantly enhancing safety, operational efficiency, and adaptability across complex construction environments.

3. Methodology

This study presents the design and implementation of a framework for conducting inspection tasks using drones in large-scale outdoor environments, as shown in Fig. 2. This framework is achieved through a close integration between physical and virtual spaces. UAV are deployed with photogrammetry to create a 3D reconstruction of construction environments, generating a coarse DT model of the site and aligning it to geographical coordinates through georegistration. UAV inspection routes are then planned based on this model, incorporating terrain-following flight to optimize coverage and maximize inspection efficiency within the drone's endurance limits in physical space. At each inspection waypoint, we developed a specialized dataset for outdoor drone inspections and implemented the YOLO-based detection network, YOLOv10-LCA, for precise worker detection. Using the detection results, a multi-object tracking network is applied to ensure continuous and accurate tracking of each worker. By combining these tracking outputs with the georegistered model, we transform the camera-based coordinate data into geographic coordinates, visualizing all worker movement trajectories within the DT model. This approach to DT modeling effectively resolves spatial alignment issues, enabling accurate, real-time representation of worker activities within the virtual site model. Further, GPS information from locators is uploaded to our management platform, enabling the verification of worker authorization status and the identification of unauthorized activities. By integrating drones with vision-based detection and GPS locators, this framework bypasses the inefficiencies and unreliability of manual verification in large-scale environments. Ultimately, it significantly improves the efficiency and accuracy of large-scale site monitoring while providing a clear, intuitive visualization for enhanced site management.

3.1. 3D reconstruction and flight route design in the physical space

In large-scale outdoor construction environments, 3D models can significantly enhance the efficiency of robotic inspections and provide better visualization, thus improving project management [61]. In this task, the 3D model is used to plan detailed drone inspection routes, and the results of worker authorization verification are also visualized on the 3D model. To achieve this, we first need to ensure that the 3D model is accurate enough to identify key areas of interest, but without being excessively detailed as a coarse 3D model is sufficient to meet the framework's requirements. Therefore, efficient flight planning for 3D reconstruction is critical.

Conducting large-scale 3D reconstruction with drones requires the use of oblique photogrammetry, which captures more surface details and textures by photographing the same scene from different angles. To design an optimal flight path for capturing these details, we must calculate the area covered by the drone's camera during each flight segment. This can be described by the following formula:

$$A_{\text{coverage}} = 2h \cdot \tan\left(\frac{\theta}{2}\right) \cdot d \tag{1}$$

where h is the drone's flight altitude, θ is the camera's field of view (FOV), and d is the distance between flight lines. This formula helps ensure that we maintain sufficient overlap between images for the 3D reconstruction, improving the accuracy and resolution of the resulting model.

Additionally, to maximize the efficiency of the flight within the limited battery life of the drone, we need to calculate the maximum area the drone can cover during a single flight. This is essential to ensure that we gather all the necessary data without exceeding the drone's operational limits. The following formula accounts for both battery life and the area covered:

$$A_{\max} = \frac{T_{\text{battery}} \cdot v \cdot 2h \cdot \tan\left(\frac{\theta}{2}\right)}{d_{\text{overlap}}} \tag{2}$$

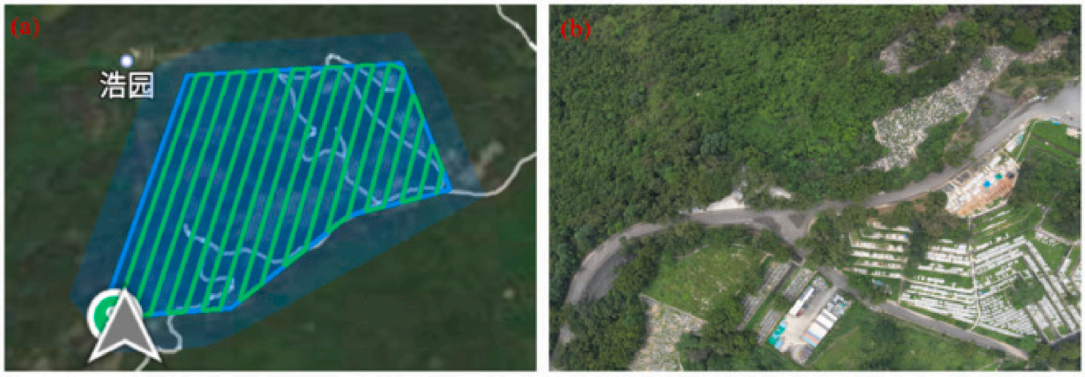

Fig. 3. (a) Flight path used for 3D reconstruction of the working site. (b) Sample image captured during the flight for 3D reconstruction.

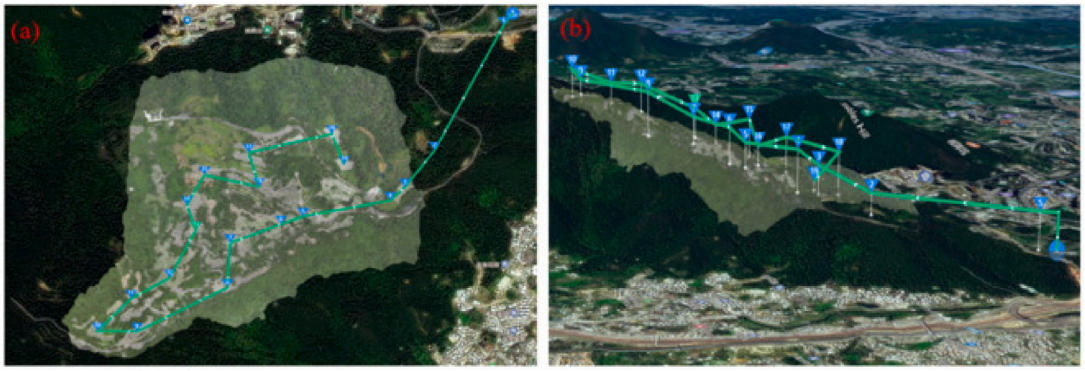

Fig. 4. (a) Top-down view of the inspection flight route using terrain-following flight. (b) Side view showing the drone maintaining a constant relative altitude of 30 m above the ground.

Here, T_{battery} is the total battery life, v is the drone's flight speed, and d_{overlap} is the required overlap distance between adjacent flight paths. This equation allows us to optimize the total area covered within the constraints of the drone's battery, ensuring efficient use of resources during large-scale 3D reconstruction. Based on these equations, we derived the oblique photography flight path as shown in Fig. 3(a). In this flight path, and image is captured every 3 s, with a sample image shown in Fig. 3(b). This flight path resulted in the collection of over 3000 images for 3D reconstruction, by inputting these images into the DJI Terra reconstruction software, a coarse model of this large-scale work environment was generated. Once the coarse 3D model is generated, we can identify areas that require detailed monitoring during the inspection process. For these areas, we employ terrain-following flight at an distance of 30 m, ensuring that the drone maintains a consistent altitude relative to the ground.

In large-scale outdoor construction sites, particularly in mountainous environments, construction managers typically designate key work zones based on terrain and safety considerations. Flat and stable terrain is generally preferred for construction activities, as it provides a safer and more controlled environment, reducing operational risks and improving efficiency. As a result, the selection of critical construction areas is inherently linked to both terrain conditions and safety requirements, with project managers prioritizing these locations for active construction. Based on this practice, the 19 inspection locations chosen in our study correspond to these predefined key work regions, ensuring that our UAV-based framework is verified in real-world scenarios where effective site surveillance is crucial. Given that the drone's maximum flight endurance is 41 min and there are 19 inspection locations, it is crucial to design an optimal route that minimizes total flight time while ensuring all locations are inspected within the time limit.

To formalize this optimization problem, we define the objective function as:

$$\text{Minimize} \sum_{i=1}^{n-1} d_{i,i+1} + \sum_{i \in N} T_{\text{hover},i} \tag{3}$$

where $d_{i,i+1}$ represents the distance between consecutive inspection locations, and $T_{\text{hover},i}$ is the hover time at location i. The UAV's total flight time remains constrained by:

$$\sum_{i=1}^{n-1} \frac{d_{i,i+1}}{v} + \sum_{i \in N} T_{\text{hover},i} + T_{\text{return}} \leq 41 \tag{4}$$

where v is the UAV's flight speed and T_{return} is the time required to return to the starting point. In our study, the hover time at each inspection location, $T_{\text{hover},i}$ is set to a default of 1 min. However, this duration can be adjusted based on predefined parameters and operational priorities. The decision regarding hover time is influenced by several factors, primarily determined by the construction manager's assessment of the area's importance and activity level. For example, in high-priority construction zones, such as areas with active work or critical operations, the UAV may hover for a longer duration to capture sufficient data and ensure thorough monitoring. On the other hand, in less critical areas, such as worker rest zones, the UAV's hover time may be minimized to conserve battery and optimize flight efficiency.

We further model this problem as a 3D Traveling Salesman Problem (3D-TSP) to determine the optimal sequence of visits while minimizing total travel time. Each inspection location i is represented by its 3D coordinates (x_i, y_i, z_i), and the direct distance between any two locations i, j is:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{5}$$

The UAV's route is optimized through the binary decision variable x_{ij}:

$$x_{ij} = \begin{cases} 1, & \text{if the UAV travels directly from location } i \text{ to location } j \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Minimizing the total travel distance is formulated as:

$$\min \sum_{i \in N} \sum_{j \in N, j \neq i} d_{ij} x_{ij} \tag{7}$$

subject to:

$$\sum_{j \in N, j \neq i} x_{ij} = 1, \quad \forall i \in N \tag{8}$$

$$\sum_{i \in N, i \neq j} x_{ij} = 1, \quad \forall j \in N \tag{9}$$

To prevent sub-tours, we introduce the Miller–Tucker–Zemlin (MTZ) constraints:

$$u_i - u_j + |N| x_{ij} \leq |N| - 1, \quad \forall i, j \in N, i \neq j \tag{10}$$

$$1 \leq u_i \leq |N|, \quad \forall i \in N \tag{11}$$

The total mission duration, incorporating adaptive hover times, is constrained by:

$$\sum_{i \in N} \sum_{j \in N, j \neq i} \frac{d_{ij} x_{ij}}{v} + \sum_{i \in N} T_{\text{hover},i} + T_{\text{return}} \leq 41 \tag{12}$$

Considering that obstacles may be encountered during the UAV's flight, we assume that the UAV used in our study is equipped with environmental awareness capabilities, such as multiple cameras or infrared sensors. To address this, we introduce a distance constraint to ensure that the UAV's flight path maintains a safe distance from any obstacles in the environment. Specifically, for each UAV position $\mathbf{p}(t) = (x(t), y(t), z(t))$, where $\mathbf{p}(t)$ represents the position of the UAV in its own coordinate system, we impose the following condition for every time step:

$$\|\mathbf{p}(t) - \mathbf{p}_{\text{obs}}(t)\| \geq d_{\text{safe}}, \quad \forall t \tag{13}$$

where $\mathbf{p}_{\text{obs}}(t)$ is the position of the obstacle at time t, and d_{safe} is the minimum safe distance between the UAV and the obstacle.

This optimization model ensures that the inspection route is designed to minimize total flight time while respecting the drone's battery life constraint, as well as avoiding potential collisions. By solving this 3D-TSP-based optimization problem, we obtain an efficient inspection route that not only meets the 41-min flight constraint but also minimizes redundant travel between locations, all while maintaining safe distances from obstacles. From another perspective, accurately estimating worker locations from the drone's video stream necessitates matching the 2D images with the 3D reconstructed model to infer the workers' geolocation. Traditional feature point matching methods prove ineffective, and ensuring the real-time performance and reliability of deep learning approaches poses challenges. To address this, we established a predetermined flight path, allowing the drone to capture the worksite from fixed positions, thus obtaining accurate metadata, including the drone's longitude, latitude, altitude, gimbal angle, and focal length during recording. This information enables the framework to accurately infer the geolocation of each target object in the video, effectively overcoming the challenges of automatic 2D–3D matching in large-scale outdoor environments.

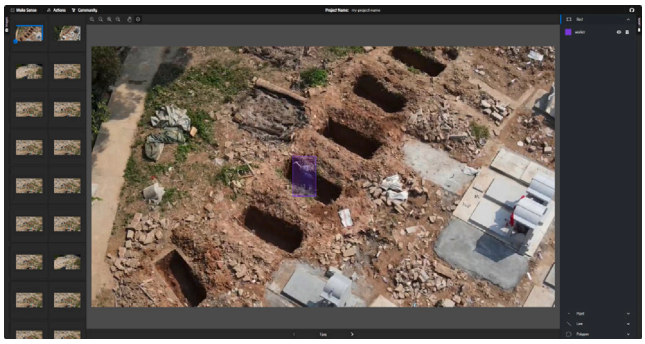

Fig. 5. Make sense annotator used for dataset construction.

3.2. UAV-oriented worker detection

Given the need for minimal computational overhead and fast inference speeds, YOLO models have emerged as an optimal choice for such tasks. Their ability to perform real-time object detection with lower computational cost is critical, particularly in UAV-based applications where processing power is constrained. The YOLO models is known for balancing accuracy with speed, making it suitable for the rapid decision-making required in drone-based monitoring scenarios. However, standard datasets like COCO and PASCAL VOC 2007, which are commonly used for training object detection models, are not ideal for UAV-based detection due to their focus on terrestrial object views. UAV images, by contrast, are captured from an aerial perspective, where objects appear smaller and often in different orientations, making detection more challenging. As a result, we constructed a specialized dataset tailored to this specific UAV monitoring scenario.

To create this dataset, we used a UAV to fly over multiple waypoints, collecting data during each flight. In total, we gathered over 3000 images from different flight points, ensuring that the dataset covered a wide range of environments and conditions encountered in real-world UAV monitoring. Each image was carefully annotated, as shown in Fig. 5, to indicate the presence of workers, equipment, and other objects of interest. The annotated images, displayed in Fig. 6, ensure that the model is trained to detect objects from a UAV's perspective, accounting for factors like varying object sizes, occlusions, and perspectives. This customized dataset accounts for the unique perspectives and conditions encountered during drone flights, ensuring the model is trained to recognize objects in a manner that reflects the aerial viewpoint.

Based on the YOLOv10-S model, we performed pretraining to develop the YOLOv10-LCA model, which is specifically optimized for drone-based monitoring in large-scale environments. The architecture of YOLOv10-S, as illustrated in Fig. 7, consists of three main components: (1) Backbone: The lightweight backbone efficiently extracts features from aerial images, enabling rapid processing under the limited computational capacity of UAVs. This is crucial for ensuring the system can handle the high volume of data in real-time while maintaining accuracy in large-scale environments. (2) Neck: The neck refines the extracted features, enhancing the model's ability to detect objects at varying scales. By balancing feature refinement and computational efficiency, it supports the detection of workers and equipment across dynamic construction sites, where targets may vary significantly in distance and appearance. (3) YOLO Head: The YOLO head is optimized for real-time performance, generating bounding boxes and class predictions rapidly. This ensures that the UAV system can detect and classify workers and equipment in real time, providing immediate feedback to support site safety and resource management.

In our system, video data captured by the drone during flight are transmitted in real-time to a remote workstation, where the trained YOLOv10-LCA model processes the video stream for object detection.

Fig. 6. Sample of annotated objects: (a) Raw images. (b) Annotated image.

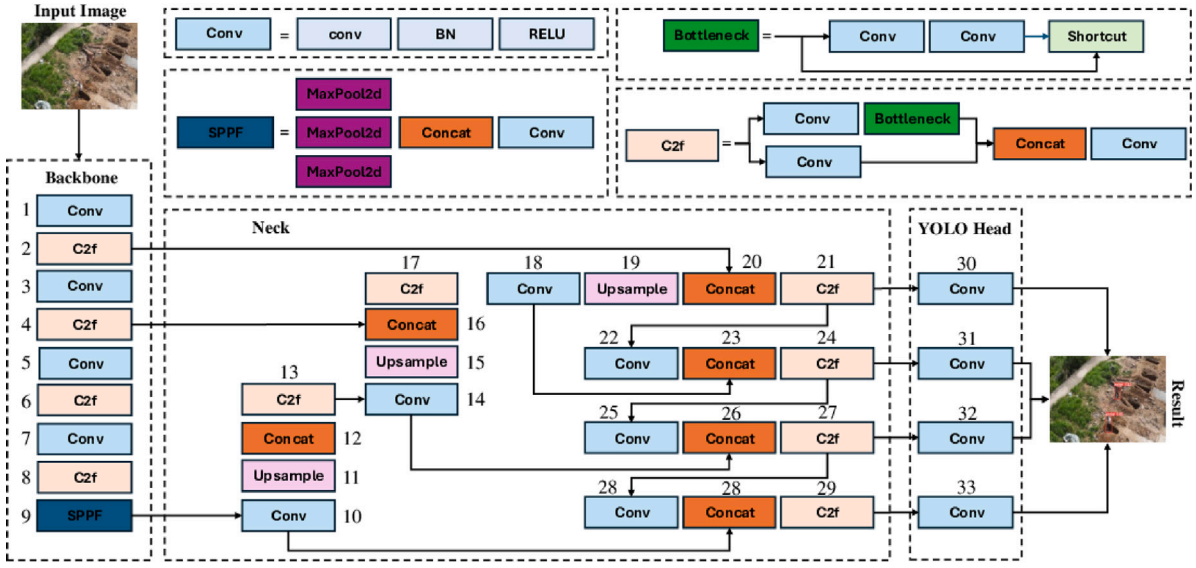

Fig. 7. Architecture of YOLOv10-LCA model based on YOLOv10s.

The model identifies workers within the video feed, drawing bounding boxes around them and providing positional information relative to the video coordinate system. This real-time detection allows for continuous monitoring of worker activities on-site, which is essential for accurate tracking and management. This positional data is subsequently used for object tracking, enabling the system to track individual workers across frames. Additionally, by combining the detected worker's location with GPS data, we can estimate the worker's latitude and longitude, providing an accurate geolocation of each worker.

3.3. Multi-object tracking and registration in virtual space

When selecting a tracking algorithm for this study, the outdoor environment presents complex scenarios, despite the relatively low density of workers to track. Therefore, ensuring stability in long-term worker tracking is essential. DeepSORT [62] is well-suited for these requirements, providing robust performance under such conditions. Additionally, DeepSORT integrates seamlessly with YOLO, utilizing the bounding boxes generated by YOLOv10-LCA for tracking purposes. This integration facilitates real-time tracking of workers in the field, making DeepSORT the optimal choice for the objectives of this study. For each frame detection in the video, DeepSORT either creates a new track $\mathbf{T}_j = \mathbf{x}_j, \mathbf{P}_j, \mathbf{f}_j, \mathrm{id}_j$ or associates the detected object with an existing track through intersection over union (IOU) matching. The IOU for bounding boxes A and B is calculated as follows:

$$\mathrm{IOU}(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{14}$$

Each track $\mathbf{T}_j$ represents a worker in the video stream, where $\mathbf{x}_j$ is the state vector, $\mathbf{P}_j$ is the covariance matrix, $\mathbf{f}_j$ is the appearance characteristic vector and id_j is the unique identifier. The Kalman filter is then employed to predict and update the state $\mathbf{x}_j$ and covariance $\mathbf{P}_j$ for each track. The Kalman filter predicts the state of the track for the next frame by using the following equations:

$$\hat{\mathbf{x}}_{j|j-1} = \mathbf{F}_j \hat{\mathbf{x}}_{j-1|j-1} + \mathbf{B}_j \mathbf{u}_j \tag{15}$$

$$\mathbf{P}_{j|j-1} = \mathbf{F}_j \mathbf{P}_{j-1|j-1} \mathbf{F}_j^T + \mathbf{Q}_j \tag{16}$$

where $\hat{\mathbf{x}}_{j|j-1}$ is the predicted state of the worker (position and velocity) at time j, based on information up to time $j-1$. $\mathbf{F}_j$ is the state transition matrix, modeling how the worker is expected to move between frames. $\mathbf{B}_j \mathbf{u} j$ represents the control input, which is often zero in tracking scenarios. $\mathbf{P} j | j-1$ is the predicted covariance, and $\mathbf{Q}_j$ is the process noise covariance, accounting for uncertainties in the worker's movement. When a new prediction is associated with an existing track, the Kalman filter updates the track's state, which can be expressed by the following equations:

$$\mathbf{K}_j = \mathbf{P}_{j|j-1} \mathbf{H}_j^T (\mathbf{H}_j \mathbf{P}_{j|j-1} \mathbf{H}_j^T + \mathbf{R}_j)^{-1} \tag{17}$$

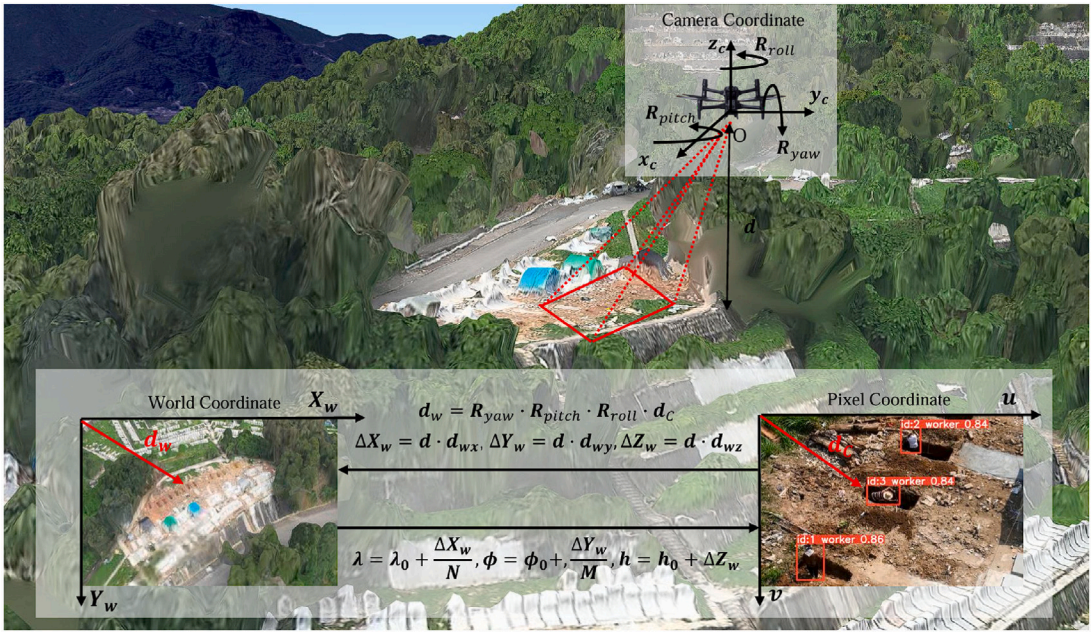

Fig. 8. Projection relationship between camera coordinate system, pixel coordinate system, and world coordinate system.

$$\hat{\mathbf{x}}_{j|j} = \hat{\mathbf{x}}_{j|j-1} + \mathbf{K}_j(\mathbf{z}_j - \mathbf{H}_j\hat{\mathbf{x}}_{j|j-1}) \tag{18}$$

$$\mathbf{P}_{j|j} = (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\mathbf{P}_{j|j-1} \tag{19}$$

where $\mathbf{K}_j$ is the Kalman gain, determining how much to trust the new measurement versus the prediction. $\mathbf{z}_j$ is the new measurement (detection) of the worker's position. $\mathbf{H}_j$ is the observation matrix, mapping the state space to the measurement space. $\mathbf{R}_j$ is the measurement noise covariance. When a predicted track cannot be associated with any existing tracks, a new track is initialized with $\hat{\mathbf{x}}_{\text{new}}$ based on the detection, and an initial covariance $\mathbf{P}_{\text{new}}$ representing uncertainty in the initial estimate. In contrast, If a track does not have a corresponding bounding box in several consecutive frames, the track will be deleted. Through the integration of DeepSORT with YOLOv10-LCA, workers can be accurately identified and tracked from the drone's perspective, enabling the further estimation of each worker's latitude and longitude information.

After confirming the specific location and shooting angles of the drone, the geographical information of the workers in the video stream can be estimated using the trained YOLOv10-LCA model. As shown in the Fig. 8, the drone's camera coordinate system is represented by (x_c, y_c, z_c), and the world coordinate system is denoted as (x_w, y_w, z_w). The pixel coordinates of the target object in the image are (u, v), where u and v represent the horizontal and vertical directions in the camera's pixel coordinate system respectively. Assuming the camera's intrinsic matrix is:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{20}$$

where f_x and f_y represent the focal lengths in the pixel coordinates, c_x and c_y are the coordinates of the optical center in the pixel coordinate system, the object's pixel coordinates in the image are defined as (x, y), which is obtained from YOLOv10-LCA. The normalized pixel coordinates u and v can then be obtained using the following equation, and the corresponding direction vector $\mathbf{d_c}$ in the camera coordinate system is derived accordingly:

$$u = \frac{x - c_x}{f_x}, \quad v = \frac{y - c_y}{f_y} \tag{21}$$

$$\mathbf{d_c} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{22}$$

The direction vector in the camera coordinate system needs to be transformed into the world coordinate system based on the drone's gimbal angles. The rotation of the gimbal is described by the angle of yaw ϕ, pitch ψ, and roll θ, then the gimbal's rotation matrix is constructed using the rotation matrices $\mathbf{R}$ corresponding to these three matrix:

$$\mathbf{R}_{\text{yaw}} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{23}$$

$$\mathbf{R}_{\text{pitch}} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{24}$$

$$\mathbf{R}_{\text{roll}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix} \tag{25}$$

$$\mathbf{R} = \mathbf{R}_{\text{yaw}} \cdot \mathbf{R}_{\text{pitch}} \cdot \mathbf{R}_{\text{roll}} \tag{26}$$

where $\mathbf{R}_{\text{yaw}}$ is the rotation matrix in the yaw direction, $\mathbf{R}_{\text{pitch}}$ is the rotation matrix in the pitch direction, and $\mathbf{R}_{\text{roll}}$ is the rotation matrix in the roll direction. By applying the combined rotation matrix $\mathbf{R}$, the direction vector $\mathbf{d}_c$ in the camera coordinate system can be transformed into the world coordinate system $\mathbf{d}_w$:

$$\mathbf{d_w} = \mathbf{R} \cdot \mathbf{d_c} \tag{27}$$

Given the distance d from the drone to the object location, the relative displacement of the object in the world coordinate system can be calculated:

$$\Delta X_w = d \cdot d_{wx}, \quad \Delta Y_w = d \cdot d_{wy}, \quad \Delta Z_w = d \cdot d_{wz} \tag{28}$$

d_{wx}, d_{wy}, and d_{wz} are the respective components of the global direction vector $\mathbf{d_w}$. Using the GPS module on the drone, its position in the world coordinate system can be known as (X_0, Y_0, Z_0), allowing the position of the target object in the world coordinate system to be calculated as:

$$X_w = X_0 + \Delta X_w, Y_w = Y_0 + \Delta Y_w, Z_w = Z_0 + \Delta Z_w \tag{29}$$

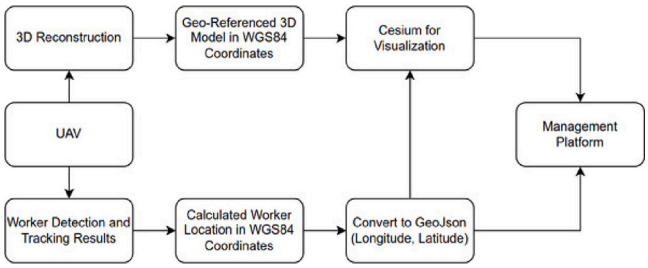

Fig. 9. Workflow for geo-referenced 3D model integration and worker visualization in the DT framework.

To convert the target object's position from the world coordinate system to geographic coordinates, the curvature of the Earth must be taken into account. The actual distance corresponding to one degree of latitude and longitude is given by the following equations:

$$M = \frac{2\pi R_{\text{earth}}}{360}, \quad N = M \cdot \cos\phi_0 \tag{30}$$

Here, $R_{earth} \approx 6\,371\,000$ m is the radius of the Earth and ϕ_0 is the initial latitude of the drone. The final longitude, latitude, and altitude of the target object can be calculated as follows:

$$\lambda = \lambda_0 + \frac{\Delta X_w}{N}, \quad \phi = \phi_0 + \frac{\Delta Y_w}{M}, \quad h = h_0 + \Delta Z_w \tag{31}$$

Through the above derivation, the real-time estimation of the geographical coordinates of all workers in the drone video stream can be achieved, with accuracy down to the centimeter level. However, this is not sufficient to verify whether workers have proper work authorization, as detecting all workers in the video stream does not confirm whether they are carrying the necessary permits. Therefore, we equip all authorized workers with a GPS locator, as shown in the figure, to distinguish between workers with and without permission in the video stream. The GPS locator provides real-time latitude and longitude information for authorized workers, also with centimeter-level accuracy. By cross-verifying the data from the GPS locator with the location information estimated from the drone's perspective, the framework can accurately identify workers without proper work permission.

Since the drone does not reside at a location for an extended period, the historical trajectories of detected workers at each work site are immediately uploaded to the visualization platform once the drone departs. Given Cesium's free and open source nature, this framework utilizes Cesium as the visualization platform [63]. As shown in Fig. 9, the workflow begins with a geo-referenced 3D model of the construction site generated from UAV-based photogrammetry, which is aligned to the WGS84 coordinate system (longitude, latitude and altitude). This model serves as the base for 3D visualization. Detected workers' locations are calculated by using our 2D–3D matching method and converted into GeoJSON format using their corresponding WGS84 coordinates. Both the 3D model and worker data are then uploaded to Cesium for real-time visualization, enabling site managers to monitor worker activities on a detailed digital twin platform.

3.4. Verification in physical space

Building upon the calculations presented in the previous section, we are able to estimate the latitude and longitude coordinates of all workers within the drone video stream. Each individual worker is tracked, and their geographic coordinates are estimated in real time based on the visual data captured by the drone. As the drone completes its flight and exits the worksite, this information is uploaded and subsequently converted into GeoJSON format, enabling its visualization on a 3D model of the entire construction site. However, relying solely on video streams is insufficient to ascertain whether workers possess valid work authorization. To address this limitation, each worker is equipped with a work permit, in the form of a GPS locator, which provides real-time positioning data with centimeter-level accuracy. By cross-referencing the GPS location with the estimated position derived from the video stream, it becomes possible to determine whether workers are engaging in unauthorized activities. If the location data from the video stream aligns closely with the GPS coordinates, it suggests that the worker is operating within the designated area, and their activities are authorized. Conversely, significant discrepancies between the video-based location estimate and the GPS data may indicate potential unauthorized actions. This verification process is then uploaded to the site management platform for further analysis and reporting. This integrated system not only enhances the security of the construction site but also provides project managers with a robust tool for efficient monitoring and inspection. By combining drone-based video tracking with precise GPS technology, this method offers an automated and highly accurate solution for worker identity verification and work authorization management. The approach minimizes human error, increases the precision of monitoring, and enables rapid detection of potential risks, thereby ensuring both safety and compliance on the construction site.

4. Experimental results

4.1. System and experiment configuration

In this study, the DJI M30T was used to complete our tasks due to its exceptional endurance, stable flight performance, precision range capabilities, and robust environmental adaptability, as shown in Fig. 10(a). The DJI M30T offers a maximum flight time of 41 min, which is crucial for large-scale continuous inspection, reducing the need for frequent battery replacements, thereby enhancing overall mission efficiency. Additionally, the drone supports real-time data synchronization, allowing real-time video streams to be accessed through its corresponding Flighthub 2 platform, facilitating subsequent object detection and geographical information estimation.

For the experimental site, tests were conducted at the location consistent in Fig. 4, situated in Fanling, North District of Hong Kong's New Territories. The site covers an area of approximately 22,500 square meters, with an elevation ranging from the base to the ridgeline, reaching a maximum height of about 565 m. Due to the complex terrain and large area, traditional UAVs and algorithms face multiple challenges in this environment. Firstly, variations in elevation and dense vegetation can lead to unstable GPS signals, affecting UAV positioning accuracy. Secondly, the diverse terrain and features increase the difficulty of image processing and object detection, potentially degrading algorithm performance. Additionally, the vast area requires UAVs with long endurance and efficient data processing capabilities to ensure complete coverage and real-time analysis of the collected data. The DJI M30T's extended flight duration and stable performance make it particularly well-suited for overcoming these challenges. During these tests, the drone maintained an altitude of 30 m above ground level across all regions, consistent with the designed flight path in virtual space. This altitude was selected to avoid most obstacles while still capturing high-resolution video streams to support worker detection and tracking.

Once the real-time video stream transmitted by the DJI M30T was accessed through Flighthub 2, it was relayed to a remote fixed workstation. This workstation was equipped with an Intel i9-14900K central processing unit (CPU) and an NVIDIA A6000 graphics processing unit (GPU). Object detection, geographical estimation, online registration, and worker authorization verification were all processed on this workstation, ensuring efficient and accurate data analysis.

To evaluate the economic feasibility of the proposed UAV-based framework, a detailed cost–benefit analysis was conducted, comparing its financial investment against traditional manual supervision and fixed CCTV systems. The total system cost includes the DJI M30T UAV ($9800), a computing workstation ($5600), software development

Fig. 10. (a) DJI M30T and its controller. (b) Sample inspection site with flight path in virtual space. (c) Drone following the designed flight path in physical space.

Table 1
ROI analysis of UAV-based system vs. manual supervision and fixed CCTV systems.

Timeframe	UAV Total Cost (USD)	Manual Supervision Cost (USD)	Fixed CCTV System Cost (USD)	ROI (%)
1 month	17,640	11,200	95,000	−36.5
3 months	17,920	33,600	95,000	87.5
6 months	18,340	67,200	95,000	266.4
12 months	19,180	134,400	95,000	600.7

($1400), training costs ($700), and monthly maintenance costs ($140). In contrast, manual supervision costs are calculated based on four supervisors patrolling the site daily, each earning a monthly salary of $2800, resulting in a total monthly labor cost of $11,200. Additionally, the patrol process requires at least three hours per session, and manual verification of worker authorization is labor-intensive and prone to errors. For fixed CCTV systems, we assume the deployment of one camera per inspection point, requiring a total of 19 cameras. Each CCTV system costs $5000, leading to an initial investment of $95,000, excluding additional expenses such as installation, wiring, and maintenance. Although CCTV systems provide continuous monitoring, they lack mobility and adaptability to large-scale dynamic environments, where UAV-based systems offer greater flexibility and efficiency.

Given these factors, we conducted a return on investment (ROI) analysis over different timeframes (1, 3, 6, and 12 months) to assess the long-term benefits of adopting UAV-based monitoring. As shown in Table 1, the UAV system initially incurs a higher upfront cost, but by the third month, it becomes more cost-effective than manual supervision, achieving an ROI of 87.5%. By the sixth month, the ROI reaches 266.4%, and by twelve months, the UAV system delivers a sixfold return on investment (ROI: 600.7%), confirming its long-term financial viability. In comparison, the fixed CCTV system requires an upfront investment of $95,000, significantly higher than the UAV system. Although CCTV provides continuous monitoring, it lacks adaptability to dynamic environments, cannot track workers outside its fixed field of view, and requires additional costs for installation and infrastructure maintenance. Unlike UAVs, which can adjust flight paths and dynamically reposition based on site conditions, CCTV cameras are static and require additional units to cover blind spots. These limitations reduce the cost-effectiveness and flexibility of CCTV systems in large-scale, evolving construction environments. Moreover, as the scale of construction increases, the cost of manual supervision and CCTV system rises proportionally, whereas the UAV system remains scalable with minimal additional cost. This highlights the UAV-based framework's potential to significantly reduce operational costs, improve efficiency, and enhance worker monitoring accuracy in large-scale, complex construction environments.

4.2. UAV-oriented worker detection and tracking results

We evaluated YOLOv5s, YOLOv8s, YOLOv10s, Faster R-CNN, DETR, and our proposed YOLOv10-LCA model on our custom dataset, as detailed in Table 2. The YOLOv10-LCA model achieved mAP and mAP50-95 scores of 96.4% and 50.4%, respectively, significantly outperforming the other models. In terms of inference speed, YOLOv10-LCA demonstrated the best performance, with an inference time of just 3 ms per image, far exceeding Faster R-CNN's 167 ms and DETR's 83 ms. Additionally, similar to other YOLO models, YOLOv10-LCA exhibits low computational resource consumption. In summary, the YOLOv10-LCA model excels in accuracy, inference speed, and computational efficiency.

For the YOLOv10-LCA training process, we used our own dataset by dividing it into a training set (2123 images, approximately 70%), a test set (304 images, approximately 10%), and a validation set (606 images, approximately 20%). The training was based on the YOLOv10-s model provided by the official YOLOv10. Without pre-training on any other datasets, we trained the model on our dataset for 2000 epochs, using stochastic gradient descent (SGD) as the optimizer, and the relevant metrics throughout the training process were also recorded per 100 epoch.

The model's bounding box prediction loss function is based on the IoU metric, as presented earlier. Since the task only requires detecting workers, the binary cross-entropy (BCE) loss was chosen as the classification loss function. For the bounding box refinement loss, the distribution focal loss (DFL) is employed to reduce the uncertainty around the boundary of the target boxes. The relevant formulas are as follow:

$$\text{BCE Loss} = -(y\log(p) + (1-y)\log(1-p)) \tag{32}$$

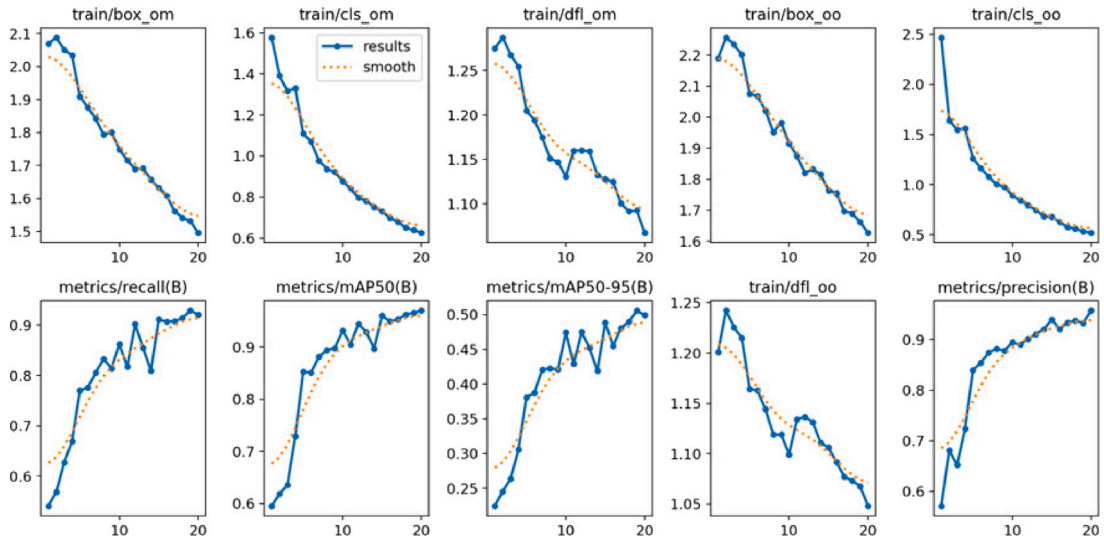

Fig. 11. Model training showing the convergence progress.

Fig. 12. Results of real-time worker detection and tracking.

Table 2
Performance comparison of different object detection models on our dataset.

Model	mAP50	mAP50-95	Inference speed (time/image)	Computing consumption
YOLOv5s	31.5%	28.4%	6.4 ms	Low
YOLOv8s	33.6%	28.7%	5.6 ms	Low
YOLOv10s	34.2%	29.1%	7 ms	Low
Faster RCNN	41.1%	34.8%	167 ms	High
DETR	45.0%	36.2%	83 ms	Medium
YOLOv10-LCA (ours)	**96.4%**	**50.4%**	**3 ms**	**Low**

$$\mathrm{DFL} = \sum_{i=1}^{N} -\left(\alpha_i \log(p_i) + (1-\alpha_i)\log(1-p_i)\right) \tag{33}$$

where y is the true label (0 or 1), p is the predicted probability, α_i is the weight assigned to each boundary prediction, and p_i is the predicted probability distribution for the target boundaries.

Recall and precision measure the model's ability to detect all existing objects and correctly identified objects respectively. mAP@50 represents the mean average precision across all categories when the IoU threshold is set as 50%. mAP50-95 evaluates the performance of model across multiple IoU thresholds from 50% to 95%. The relevant formulas are as follows:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{34}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{35}$$

$$\mathrm{mAP@50} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i \tag{36}$$

$$\mathrm{mAP@50\text{-}95} = \frac{1}{10}\sum_{t=0.5}^{0.95}\frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i(t) \tag{37}$$

The Fig. 11 illustrates the gradual decline in the model's bounding box prediction loss, classification loss, and bounding box refinement loss throughout the training process, while mAP@50, mAP@50-95, precision, and recall steadily improve, indicating that the training is converging. The Table 3 also presents the evaluation results of the model after training. For the "worker" category, a total of 1535 workers were detected across 606 test images, with a box precision of 0.931 and a box recall of 0.933. The mAP@50 and mAP@50-95 are 0.965 and 0.596 respectively, and the inference speed is 0.2 ms per frame. All these metrics indicate that the trained model fully meets the accuracy and real-time requirements of the framework under our evaluation.

Table 3
Relevant metrics for evaluation of the trained model.

Evaluation metrics	Related value
Images	606
Instances	1535
Box(P)	0.931
R	0.933
mAP50	0.964
mAP50-90	0.506
Inference speed	3 ms/image

For the task of object tracking, we integrated the DeepSORT tracking algorithm into YOLOv10-LCA, allowing for real-time tracking during detection. Since the drone hovers at each work point, and the movement of workers within the drone's field of view results in minimal pixel displacement, any work detected can be continuously tracked. The sample figures of the detection and tracking results are shown in Fig. 12. These examples demonstrate that the framework are capable of

Table 4
Impact of hover time on detection accuracy and worker count (averaged over 5 sites).

Hover time (s)	Average detection accuracy	Detected workers	True number
10	91.2%	1.4	5
30	92.5%	3.8	5
60	93.8%	4.8	5
90	93.1%	5.0	5
120	94.9%	5.0	5

Table 5
Performance of YOLOv10-LCA on public UAV datasets.

Dataset	Environment type	Precision (%)	Recall (%)	mAP50 (%)
UAVDT	Urban/Highway	43.2	45.6	47.1
VisDrone	Urban/Rural	40.7	42.1	41.5
UAV123	Mixed scenarios	41.9	44.0	43.2

accurately detecting and tracking all workers in the video stream, even in challenging conditions such as worker gathering, smoke obstruction, small target size, or when target features are less distinguishable in the context environment.

We have also validated the relationship between UAV hover time, detection accuracy, and the number of detected workers, confirming that optimizing hover duration is crucial to ensuring comprehensive worker monitoring while maintaining operational efficiency. Table 4 shows the relationship between UAV hover time, detection accuracy, and the number of detected workers. While detection accuracy does not exhibit a direct correlation with hover time, shorter hover times result in fewer workers being detected compared to the actual number of workers (True Number). This suggests that insufficient hover time may cause certain workers to be missed during the detection process, rather than significantly affecting detection accuracy. These findings further emphasize the importance of adaptive hover times in ensuring comprehensive monitoring while maintaining operational efficiency.

Furthermore, to validate the model's generalizability, we tested it on three publicly available UAV-based datasets: UAVDT [45], VisDrone [46], and UAV123 [47]. These datasets feature diverse environmental conditions and drone perspectives, providing a robust benchmark for assessing the model's adaptability. As shown in Table 5, our model achieves mAP50 values ranging from 41.5% to 47.1% across the three datasets, demonstrating moderate generalizability to diverse UAV-based scenarios. While the results indicate the model's potential to adapt to different environments, there is room for improvement, particularly in scenarios with highly complex backgrounds or low-contrast objects.

4.3. Results of worker registration and verification

For the design of our electrical work permit, we created a custom GPS locator specifically tailored to this purpose. The figure of the permit and its key parameter are shown in Fig. 13 and Table 6. This device is capable of providing real-time latitude and longitude data with centimeter-level accuracy by utilizing a single SIM card. To address the issue of endurance, the locator is equipped with an 1800 mAh battery, offering approximately 10 h of continuous operation. Additionally, the locator supports playback of historical trajectories, facilitating the registration and validation of the tracked paths.

To further validate the feasibility of the electrical work permit in this task, we conducted tests with and without the permit. Fig. 14(a) illustrates the scenario where the worker is equipped with the card, while Fig. 14(b) shows the scenario without the card. When the card is worn, the worker's real-time geographic information is accurately mapped, as shown in Fig. 14(c). In contrast, without the card, no information about the worker is displayed. After measurement, the geographic information provided in Fig. 14(c) closely matches the position in Fig. 14(a), demonstrating the feasibility and accuracy of our electronic work permit system.

Fig. 13. (a) Design diagram of electronic work permit. (b) Physical representation of the electronic work permit.

Table 6
Key specifications of the electrical work permit.

Parameter	Specification
Module	4G: A7670C, GPS: L76K
Battery	1800 mAh
Battery life	10 h
Real-time positioning	Supported
Historical playback	Supported
Alert messages	Supported
Receiving sensitivity	1 cm
First-time fix	Cold start: Average 32 s

Algorithm 1 Outdoor Inspection Task Workflow

Input: Total inspection sites $N = 19$
Output: Report of workers with and without work permits, detailed times for each inspection site
for $i = 1$ to N **do**
  Fly to inspection site P_i at T_i ▷ Fly to the site at the scheduled time T_i
  Hover at height $H = 30m$ for $T_{hover} = 1$ minute
  Detect and track all workers W_i at site P_i using the camera feed
  if detection is successful **then**
    Upload detection results $U(W_i)$ to management platform
    Estimate workers' geographical positions $G(W_i)$
    Cross-verify positions using electronic work permits $E(W_i)$
    for each worker w_j in W_i **do**
      if GPS position matches work permit information for w_j **then**
        Add worker w_j to authorized workers list $L(W_i)$
      else
        Add worker w_j to unauthorized workers list $I(W_i)$
      end if
    end for
  end if
  Fly to next inspection site P_{i+1} ▷ Continue to the next site after hovering for 1 minute
end for
Generate final report of $L(W_1), L(W_2), \ldots, L(W_N)$ and $I(W_1), I(W_2), \ldots, I(W_N)$
Include detailed times for each inspection site: $T_1, T_2, \ldots, T_N$

Throughout the inspection process, the workflow of the task is shown in Algorithm. 1. The task consists of a total of 19 inspection sites. After all inspections are completed, we collect data on all workers detected during the inspections and verify whether they are operating illegally. At each inspection site, the UAV hovers for 1 min at a fixed gimbal angle and altitude, during which workers are detected and

Fig. 14. (a) and (b) Scenarios where worker is wearing and not wearing the work permit. (c) Real-time location information when worker is wearing the work permit.

Fig. 15. (a) Real-time worker detection and tracking from UAV video feed. (b) Real-time status of the UAV. (c) Statistics of workers detected and verified illegal workers.

tracked. After the 1-min hover, the UAV flies to the next site, and during this transition, it uploads the detection and tracking results to our management platform. The platform then estimates the geographic locations of all workers and cross-validates them with the GPS information provided by the electronic work permit, integrating the results into our database. Once all inspection points have been completed, our management platform generates a report summarizing the task and produces a line chart for a more intuitive data visualization. The Fig. 15 shows our management platform, which provides a detailed display of all modules in our inspection system. The key modules include: (a) the real-time status of the drone during the inspection process, including the inspection progress and battery level; (b) the detection and tracking results of the real-time video stream transmitted by the drone; and (c) the records of all workers detected during the inspection, including those with and without authorization. Through verification via the electronic work permit, the system identified illegal workers at each site and provided their latitude and longitude information, along with the current time of the UAV. The report automatically generates as Table 7 and Fig. 16 shown, which presents a statistical excel and chart of detected workers, legal workers, and illegal workers at each inspection site, offering a clearer visualization of the inspection results.

For real-time worker registration, Fig. 17 illustrates an example at inspection site 4, where the UAV detected and tracked three workers. Their latitude and longitude information was estimated through projection, as shown in Fig. 17(a). Subsequently, all their trajectories during the UAV's one-minute hover were converted into GeoJSON format and registered onto the geo-referenced 3D model we previously established, as shown in Fig. 17(b), providing a direct visualization of the workers' movement at this inspection site. However, our management platform only received GPS data for two of the workers at this site. Through cross-verification, we identified that the worker with ID 3 was operating illegally.

Table 7
Worker statistics at each inspection site during a subsequent inspection task.

Inspection site	Detected workers	Illegal workers	Geolocation of illegal workers	Hover time (min)	Current time
Site 1	W01-1; W01-2	None	None	1	10:01:33
Site 2	W02-1; W02-2; W02-3; W02-4; W02-5	None	None	1	10:03:29
Site 3	W03-1; W03-2; W03-3	None	None	1	10:05:03
Site 4	W04-1; W04-2; W04-3	W04-1	22.4749360, 114.1448692	1	10:06:21
Site 5	W05-1	None	None	1	10:08:55
Site 6	W06-1	None	None	1	10:10:40
Site 7	None	None	None	1	10:12:25
Site 8	W08-1; W08-2; W08-3; W08-4; W08-5	W08-2	22.4718796, 114.1388772	1	10:14:03
Site 9	W09-1; W09-2; W09-3	None	None	1	10:15:24
Site 10	W10-1; W10-2; W10-3	None	None	1	10:16:59
Site 11	None	None	None	1	10:18:14
Site 12	W12-1; W12-2; W12-3; W12-4; W12-5; W12-6	None	None	1	10:19:38
Site 13	W13-1; W13-2; W13-3; W13-4	W13-1	22.4750975, 114.1376703	1	10:21:47
Site 14	W14-1	None	None	1	10:24:14
Site 15	W15-1	None	None	1	10:26:02
Site 16	W16-1; W16-2	None	None	1	10:27:52
Site 17	W17-1; W17-2; W17-3	W17-2	22.4768613, 114.1396474	1	10:30:33
Site 18	W18-1; W18-2; W18-3; W18-4	None	None	1	10:32:45
Site 19	W19-1; W19-2; W19-3	None	None	1	10:34:21

Table 8
Comparison of UAV-based monitoring systems.

System	Data integration level	Real-time performance	Visualization capability
Proposed system	High	High	Advanced (geo-referenced 3D DT)
Singh et al. [64]	Medium	Medium	Limited (focus on network design)
Zhou et al. [65]	High	High	None (focus on routing and service network design)

Fig. 16. Line chart displaying statistics of legal and illegal workers across 19 inspection sites.

We also compared our proposed framework with existing systems, noting that while these frameworks are not directly related to our specific task, their methodologies are adaptable to our research goals. We present a quantitative comparison to emphasize the strengths of our system in data integration, real-time performance, and advanced visualization. Table 8 highlights the key differences between our system and those of Singh et al. [64] and Zhou et al. [65] Singh et al.'s system offers moderate data integration and real-time monitoring but lacks advanced visualization and worker verification. Zhou et al.'s system focuses on routing and service network design with high real-time performance but does not include comprehensive worker monitoring or dynamic visualization. In contrast, our system provides superior integration, real-time updates, and advanced 3D visualization, making it a more comprehensive solution for construction site monitoring and worker management.

5. Conclusions and future work

5.1. Conclusions

This paper presented a UAV-enabled DT framework for real-time worker monitoring and authorization in large-scale construction sites, addressing the limitations of traditional surveillance. By integrating UAV-based visual detection with GPS tracking, the system automates worker verification and enhances detection accuracy and speed using the YOLOv10-LCA model. Additionally, automated 2D–3D spatial alignment ensures precise worker localization, improving site security and operational efficiency. Real-world validation confirms its effectiveness in data-driven site management, reducing the need for manual supervision while optimizing workforce allocation. Its adaptability to complex construction environments makes it a scalable solution for enhancing compliance monitoring, site security, and resource management, contributing to smart city initiatives through automated workflow optimization and reduced environmental impact.

The continuous UAV-driven data collection enables iterative model retraining, steadily improving detection robustness and verification accuracy. By centralizing workforce data and automating site surveillance, the framework minimizes manual inspections, reduces supervisory costs, and streamlines real-time workforce management. The system's ability to detect unauthorized personnel and provide real-time analytics ensures enhanced site safety and operational efficiency. These capabilities establish a scalable and intelligent construction monitoring solution, advancing the digital transformation of construction site management while enabling more autonomous and data-driven decision-making.

However, as this framework is based on a single UAV, it is inherently constrained by the UAV's battery life and flight duration, limiting the maximum length of a single inspection cycle. For larger or more complex construction environments, future implementations can consider dividing the site into sub-regions to allow sequential inspections or employing multi-UAV collaborative strategies to achieve comprehensive monitoring and management across expansive work zones. Such extensions would further enhance the scalability and adaptability of the system, making it suitable for even more demanding construction scenarios.

Fig. 17. (a) Real-time detection and tracking of workers with GPS information. (b) Historical trajectories of workers' localization on geo-referenced 3D model, with the red box indicating working region.

Moreover, the high-quality data generated provides substantial opportunities for expanded analytical applications, such as assessing workers' safety compliance and verifying protective equipment standards. This system thus represents a breakthrough in both research and operational practice, paving the way for innovative, data-driven site management. Its adaptability and comprehensive data integration establish a strong foundation for monitoring solutions within large-scale, dynamic construction environments.

5.2. Future work

The potential of this UAV-enabled DT modeling method extends beyond the current system's capabilities, and several avenues for future research and development could further enhance its robustness, adaptability, and functionality in real-world applications. One priority for future work is to improve the system's resilience under varied environmental conditions, including adverse weather, low-light environments, and densely built urban areas where GPS signal interference may be common. Integrating additional sensing technologies, such as infrared, LiDAR, or micrometer-wave imaging, will expand the framework's applicability, allowing it to perform consistently across diverse and challenging operational settings. These sensor integrations could also enable the detection of worker health conditions and environmental hazards, thus enhancing the safety monitoring functions of the platform.

Moreover, while this DT modeling method demonstrates strong potential for worker monitoring and management, one limitation lies in its generalization capabilities across diverse construction environments. The current framework is trained and validated on specific datasets, which may not fully capture the variability in construction site conditions, such as differing terrains, lighting, and worker behaviors. Addressing this limitation requires expanding the training dataset to include more diverse construction scenarios, collected from various geographic locations and project types. This will enhance the model's ability to generalize, ensuring robust performance in real-world applications. Expanding the scope of this DT modeling method to include asset tracking, equipment monitoring, and automated anomaly detection represents another essential step in transforming it into a fully comprehensive site management tool. Integrating AI-based predictive analytics could enable the system to assess potential safety risks proactively, allowing managers to take preventive measures and thereby reduce accidents on-site. Material tracking and equipment usage monitoring would further streamline logistics, improving overall resource allocation and ensuring timely project progression. Additionally, incorporating real-time anomaly detection would allow the system to identify potential safety or operational issues before they escalate, supporting faster, more proactive responses. By addressing the current limitations and continuously expanding its capabilities, this framework has the potential to evolve into a versatile and intelligent solution for modern construction site management.

In summary, this study establishes a foundation for the UAV-enabled DT technology application in large-scale urban construction management, offering a scalable and sustainable solution for smart city development. By enhancing workflow efficiencies, improving worker safety, and enabling more sustainable site management, the integration of UAV and DT technologies promises to contribute significantly to the resilience, safety, and sustainability objectives of modern urban infrastructure projects. Future developments will aim to expand the system's scope, flexibility, and intelligence, positioning it as a core component in the digital transformation of the construction industry.

CRediT authorship contribution statement

Mingqiao Han: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jihan Zhang:** Writing – original draft, Supervision, Investigation, Formal analysis, Conceptualization. **Yijun Huang:** Formal analysis, Conceptualization. **Jiwen Xu:** Investigation, Formal analysis. **Xi Chen:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Ben M. Chen:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ben M. Chen reports financial support was provided by The Chinese University of Hong Kong. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the InnoHK of the Government of the Hong Kong Special Administrative Region via the Hong Kong Centre for Logistics Robotics (HKCLR), in part by the Environment and Conservation Fund, Hong Kong SAR (Project No. 142/2023), and in part by the Research Grants Council of Hong Kong SAR under Grants 14206821, 14217922 and 14209623. The authors would like to acknowledge the significant contributions of Qingxiang Li in the conceptualization, investigation, and formal analysis of this study. Additionally, Pei Wang's valuable support in the investigation and formal analysis processes is greatly appreciated. Their hard work and insightful contributions were essential to the successful completion of this research.

Data availability

Data will be made available on request.

References

[1] C. Wang, L. Chang, L. Zhao, R. Niu, Automatic identification and dynamic monitoring of open-pit mines based on improved mask R-CNN and transfer learning, Remote. Sens. 12 (21) (2020) 3474, http://dx.doi.org/10.3390/rs12213474.

[2] M. Martinez-Luengo, A. Kolios, L. Wang, Structural health monitoring of offshore wind turbines: A review through the statistical pattern recognition paradigm, Renew. Sustain. Energy Rev. 64 (2016) 91–105, http://dx.doi.org/10.1016/j.rser.2016.05.085.

[3] R. Samsami, A. Mukherjee, C.N. Brooks, Mapping unmanned aerial system data onto building information modeling parameters for highway construction progress monitoring, Transp. Res. Rec. 2676 (4) (2022) 669–682, http://dx.doi.org/10.1177/03611981211064277.

[4] J.T. Woo, A study on analysis of construction monitoring cost and improvement measures of railway tunnel construction in seoul, J. Soc. Disaster Inf. 19 (1) (2023) 18–30, http://dx.doi.org/10.15683/kosdi.2023.3.31.018.

[5] R. Jin, H. Zhang, D. Liu, X. Yan, IoT-based. detecting, Locating and alarming of unauthorized intrusion on construction sites, Autom. Constr. 118 (2020) 103278, http://dx.doi.org/10.1016/j.autcon.2020.103278.

[6] W. Choi, S. Na, S. Heo, Integrating drone imagery and AI for improved construction site management through building information modeling, Build. 14 (4) (2024) 1106, http://dx.doi.org/10.3390/buildings14041106.

[7] T. Bamford, F. Medinac, K. Esmaeili, Continuous monitoring and improvement of the blasting process in open pit mines using unmanned aerial vehicle techniques, Remote. Sens. 12 (17) (2020) 2801, http://dx.doi.org/10.3390/rs12172801.

[8] C. Klaufus, Safeguarding the house of the dead: Configurations of risk and protection in the urban cemetery, Int. J. Urban Reg. Res. 45 (6) (2021) 1056–1063, http://dx.doi.org/10.1111/1468-2427.12890.

[9] J. Kim, Visual analytics for operation-level construction monitoring and documentation: State-of-the-art technologies, Res. Chall., Futur. Dir. Front. Built Environ. 6 (2020) http://dx.doi.org/10.3389/fbuil.2020.575738.

[10] J.G. Martinez, M. Gheisari, L.F. Alarcón, UAV integration in current construction safety planning and monitoring processes: Case study of a high-rise building construction project in Chile, J. Manag. Eng. 36 (3) (2020) 05020005, http://dx.doi.org/10.1061/(ASCE)ME.1943-5479.0000761.

[11] A. Ibrahim, M. Golparvar-Fard, K. El-Rayes, Multiobjective optimization of reality capture plans for computer vision–driven construction monitoring with camera-equipped UAVs, J. Comput. Civ. Eng. 36 (5) (2022) 04022018, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0001032.

[12] D. Han, S.B. Lee, M. Song, J.S. Cho, Change detection in unmanned aerial vehicle images for progress monitoring of road construction, Build. 11 (4) (2021) 150, http://dx.doi.org/10.3390/buildings11040150.

[13] I. Jeelani, M. Gheisari, Safety challenges of UAV integration in construction: Conceptual analysis and future research roadmap, Saf. Sci. 144 (2021) 105473, http://dx.doi.org/10.1016/j.ssci.2021.105473.

[14] Q. Ye, Y. Fang, N. Zheng, Performance evaluation of struck-by-accident alert systems for road work zone safety, Autom. Constr. 168 (2024) 105837, http://dx.doi.org/10.1016/j.autcon.2024.105837.

[15] G. D'Amico, K. Szopik-Depczyńska, I. Dembińska, G. Ioppolo, Smart and sustainable logistics of port cities: A framework for comprehending enabling factors, Domains Goals, Sustain. Cities Soc. 69 (2021) 102801, http://dx.doi.org/10.1016/j.scs.2021.102801.

[16] F. Pour Rahimian, S. Seyedzadeh, S. Oliver, S. Rodriguez, N. Dawood, On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning, Autom. Constr. 110 (2020) 103012, http://dx.doi.org/10.1016/j.autcon.2019.103012.

[17] R. Xiong, Y. Song, H. Li, Y. Wang, Onsite video mining for construction hazards identification with visual relationships, Adv. Eng. Inform. 42 (2019) 100966, http://dx.doi.org/10.1016/j.aei.2019.100966.

[18] R. Duan, H. Deng, M. Tian, Y. Deng, J. Lin, SODA: A large-scale open site object detection dataset for deep learning in construction, Autom. Constr. 142 (2022) 104499, http://dx.doi.org/10.1016/j.autcon.2022.104499.

[19] I. Cárdenas-León, M. Koeva, P. Nourian, C. Davey, Urban digital twin-based solution using geospatial information for solid waste management, Sustain. Cities Soc. 115 (2024) 105798, http://dx.doi.org/10.1016/j.scs.2024.105798.

[20] Q. Li, G. Yang, C. Bian, L. Long, X. Wang, C. Gao, C. Wong, Y. Huang, B. Zhao, X. Chen, B. Chen, Autonomous design framework for deploying building integrated photovoltaics, Appl. Energy 377 (2025) 124760, http://dx.doi.org/10.1016/j.apenergy.2024.124760.

[21] H. Liang, S.-C. Lee, W. Bae, J. Kim, S. Seo, Towards UAVs in construction: Advancements, Challenges, Futur. Dir. Monit. Insp. Drones 7 (3) (2023) 202, http://dx.doi.org/10.3390/drones7030202.

[22] K. Shaad, Y. Ninsalam, R. Padawangi, P. Burlando, Towards high resolution and cost-effective terrain mapping for urban hydrodynamic modelling in densely settled river-corridors, Sustain. Cities Soc. 20 (2016) 168–179, http://dx.doi.org/10.1016/j.scs.2015.09.005.

[23] Y. Ham, K.K. Han, J.J. Lin, et al., Visual monitoring of civil infrastructure systems via camera-equipped unmanned aerial vehicles (UAVs): a review of related works, Vis. Eng. 4 (2016) 1, http://dx.doi.org/10.1186/s40327-015-0029-z.

[24] J.-Y. Han, C.-R. Hsu, C.-J. Huang, Automated progress monitoring of land development projects using unmanned aerial vehicles and machine learning, Autom. Constr. 168 (2024) 105827, http://dx.doi.org/10.1016/j.autcon.2024.105827.

[25] N. Jacob-Loyola, F. Muñoz-La Rivera, R.F. Herrera, E. Atencio, Unmanned aerial vehicles (UAVs) for physical progress monitoring of construction, Sensors 21 (12) (2021) 4227, http://dx.doi.org/10.3390/s21124227.

[26] H. Niwa, R. Manabe, WBGT prediction with high spatial resolution using actual measurement data and data acquired using infrared sensors mounted on UAVs, Sustain. Cities Soc. 107 (2024) 105470, http://dx.doi.org/10.1016/j.scs.2024.105470.

[27] S. Zhao, F. Kang, J. Li, C. Ma, Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction, Autom. Constr. 130 (2021) 103832, http://dx.doi.org/10.1016/j.autcon.2021.103832.

[28] Z. Shang, Z. Shen, Real-time 3D reconstruction on construction site using visual SLAM and UAV, 2017, http://dx.doi.org/10.48550/arXiv.1712.07122, arXiv.

[29] Y. Tan, W. Yi, P. Chen, Y. Zou, An adaptive crack inspection method for building surface based on BIM, UAV Edge Comput. Autom. Constr. 157 (2024) 105161, http://dx.doi.org/10.1016/j.autcon.2023.105161.
[30] H. Ren, Y. Zhao, W. Xiao, Z. Hu, A review of UAV monitoring in mining areas: current status and future perspectives, i, Nternational J. Coal Sci. Technol. 6 (3) (2019) 320–333, http://dx.doi.org/10.1007/s40789-019-00264-5.
[31] F. Riyanto, J. Juliastuti, O. Setyandito, A. Pramudya, Realtime monitoring study for highway construction using unmanned aerial vehicle (UAV) technology, IOP Conf. Series: Earth Environ. Sci. 729 (2021) 012040, http://dx.doi.org/10.1088/1755-1315/729/1/012040.
[32] S. Lee, M. Song, S. Kim, J.-H. Won, Change monitoring at expressway infrastructure construction sites using drone, Sensors Mater. 32 (2020) 3923, http://dx.doi.org/10.18494/SAM.2020.2971.
[33] J. Wu, L. Peng, J. Li, X. Zhou, J. Zhong, C. Wang, J. Sun, Rapid safety monitoring and analysis of foundation pit construction using unmanned aerial vehicle images, Autom. Constr. 128 (2021) 103706, http://dx.doi.org/10.1016/j.autcon.2021.103706.
[34] G. Lindner, K. Schraml, R. Mansberger, J. Hübl, UAV monitoring and documentation of a large landslide, Appl. Geomat. 8 (1) (2016) 1–11, http://dx.doi.org/10.1007/s12518-015-0165-0.
[35] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding, YOLOv10: Real-time end-to-end object detection, 2024, arXiv, https://arxiv.org/abs/2405.14458.
[36] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: Single shot MultiBox detector, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 21–37, http://dx.doi.org/10.1007/978-3-319-46448-0_2.
[37] G. Han, X. Zhang, C. Li, Revisiting faster R-CNN: A deeper look at region proposal network, in: D. Liu, S. Xie, Y. Li, D. Zhao, E.-A. M. El-Sayed (Eds.), Neural Information Processing, Springer International Publishing, Cham, 2017, pp. 14–24, http://dx.doi.org/10.1007/978-3-319-70090-8_2.
[38] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149, http://dx.doi.org/10.1109/TPAMI.2016.2577031.
[39] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August (2020) 23–28, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2020, pp. 213–229, http://dx.doi.org/10.1007/978-3-030-58452-8_13.
[40] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, J. Chen, DETRs beat YOLOs on real-time object detection, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2024, pp. 16965–16974, http://dx.doi.org/10.1109/CVPR52733.2024.01605.
[41] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, T. Huang, UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios, Sensors 23 (16) (2023) 7190, http://dx.doi.org/10.3390/s23167190.
[42] Y. Zeng, T. Zhang, W. He, Z. Zhang, YOLOv7-UAV: An unmanned aerial vehicle image object detection algorithm based on improved YOLOv7, Electron. 12 (14) (2023) 3141, http://dx.doi.org/10.3390/electronics12143141.
[43] S. Ehsan, K.D. McDonald-Maier, On-board vision processing for small UAVs: Time to rethink strategy, in: 2009 NASA/ESA Conference on Adaptive Hardware and Systems, 2009, pp. 75–81, http://dx.doi.org/10.1109/AHS.2009.6.
[44] M. Jouhari, A.K. Al-Ali, E. Baccour, A. Mohamed, A. Erbad, M. Guizani, M. Hamdi, Distributed CNN inference on resource-constrained UAVs for surveillance systems: Design and optimization, IEEE Internet Things J. 9 (2) (2022) 1227–1242, http://dx.doi.org/10.1109/JIOT.2021.3079164.
[45] P.-Y. Chen, M.-C. Chang, J.-W. Hsieh, Y.-S. Chen, Parallel residual bi-fusion feature pyramid network for accurate single-shot object detection, IEEE Trans. Image Process. 30 (2021) 9099–9111, http://dx.doi.org/10.1109/TIP.2021.3118953.
[46] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, H. Ling, Detection and tracking meet drones challenge, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2022) 7380–7399, http://dx.doi.org/10.1109/TPAMI.2021.3119563.
[47] M. Mueller, N. Smith, B. Ghanem, A benchmark and simulator for UAV tracking, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 445–461, http://dx.doi.org/10.1007/978-3-319-46448-0_27.
[48] G.A.Q. Abdulrahman, N.A.A. Qasem, Optimizing endurance and power consumption of fixed-wing unmanned aerial vehicles powered by proton-exchange membrane fuel cells, Arab. J. Sci. Eng. 49 (2) (2024) 2605–2623, http://dx.doi.org/10.1007/s13369-023-08335-1.
[49] G. Tang, J. Ni, Y. Zhao, Y. Gu, W. Cao, A survey of object detection for UAVs based on deep learning, Remote. Sens. 16 (1) (2024) 149, http://dx.doi.org/10.3390/rs16010149.
[50] X. He, Z. Tang, Y. Deng, G. Zhou, Y. Wang, L. Li, UAV-based road crack object-detection algorithm, Autom. Constr. 154 (2023) 105014, http://dx.doi.org/10.1016/j.autcon.2023.105014.
[51] W. Ding, H. Yang, K. Yu, J. Shu, Crack detection and quantification for concrete structures using UAV and transformer, Autom. Constr. 152 (2023) 104929, http://dx.doi.org/10.1016/j.autcon.2023.104929.
[52] J.-L. Xiao, J.-S. Fan, Y.-F. Liu, B.-L. Li, J.-G. Nie, Region of interest (ROI) extraction and crack detection for UAV-based bridge inspection using point cloud segmentation and 3D-to-2D projection, Autom. Constr. 158 (2024) 105226, http://dx.doi.org/10.1016/j.autcon.2023.105226.
[53] M. Khazen, M. Nik-Bakht, O. Moselhi, Monitoring workers on indoor construction sites using data fusion of real-time worker's location, Body Orientat. Prod. State, Autom. Constr. 160 (2024) 105327, http://dx.doi.org/10.1016/j.autcon.2024.105327.
[54] X. Chen, Y. Wang, J. Wang, A. Bouferguene, M. Al-Hussein, Vision-based real-time process monitoring and problem feedback for productivity-oriented analysis in off-site construction, Autom. Constr. 162 (2024) 105389, http://dx.doi.org/10.1016/j.autcon.2024.105389.
[55] M. Grieves, J. Vickers, Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems, in: Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches, 2017, pp. 85–113, http://dx.doi.org/10.1007/978-3-319-38756-7_4.
[56] Y. Sun, H. Fesenko, V. Kharchenko, L. Zhong, I. Kliushnikov, O. Illiashenko, O. Morozova, A. Sachenko, UAV and IoT-based systems for the monitoring of industrial facilities using digital twins: Methodology, reliability models, and application, Sensors 22 (17) (2022) 6444, http://dx.doi.org/10.3390/s22176444.
[57] J. Li, S. Hong, C. Ai, An efficient method for safety monitoring of substation workers using fusion of point cloud and image data, in: 2024 5th International Conference on Electronic Communication and Artificial Intelligence, ICECAI, 2024, pp. 353–356, http://dx.doi.org/10.1109/ICECAI62591.2024.10674791.
[58] T.H. Wang, A. Pal, J.J. Lin, S.-H. Hsieh, Construction photo localization in 3D reality models for vision-based automated daily project monitoring, J. Comput. Civ. Eng. 37 (6) (2023) 04023029, http://dx.doi.org/10.1061/JCCEE5.CPENG-5353.
[59] H. Feng, Q. Chen, B. García de Soto, Application of digital twin technologies in construction: An overview of opportunities and challenges, in: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, 2021, p. 0132, http://dx.doi.org/10.22260/ISARC2021/0132.
[60] S. Wang, C. Rodgers, G. Zhai, T.N. Matiki, B. Welsh, A. Najafi, J. Wang, Y. Narazaki, V. Hoskere, B.F. Spencer, A graphics-based digital twin framework for computer vision-based post-earthquake structural inspection and evaluation using unmanned aerial vehicles, J. Infrastruct. Intell. Resil. 1 (1) (2022) 100003, http://dx.doi.org/10.1016/j.iintel.2022.100003.
[61] Y. Lu, S. Wang, S. Fan, J. Lu, P. Li, P. Tang, Image-based 3D reconstruction for multi-scale civil and infrastructure projects: A review from 2012 to 2022 with new perspective from deep learning methods, Adv. Eng. Inform. 59 (2024) 102268, http://dx.doi.org/10.1016/j.aei.2023.102268.
[62] 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, http://dx.doi.org/10.1109/ICIP.2017.8296962.
[63] Cesium: The platform for 3D geospatial visualization, 2024, Available: https://cesium.com (Accessed 20 Febuary 2025).
[64] P.K. Singh, A. Sharma, An intelligent WSN-UAV-based IoT framework for precision agriculture application, Comput. Electr. Eng. 100 (2022) 107912, http://dx.doi.org/10.1016/j.compeleceng.2022.107912.
[65] B. Zhou, W. Liu, H. Yang, Unmanned aerial vehicle service network design for urban monitoring, Transp. Res. Part C: Emerg. Technol. 157 (2023) 104406, http://dx.doi.org/10.1016/j.trc.2023.104406.