Build Simul (2025) 18: 2531–2558
https://doi.org/10.1007/s12273-025-1332-9

AI-empowered digital twin modeling for high-precision building defect management integrating UAV and GeoBIM

Jihan Zhang[1], Benyun Zhao[1] (✉), Guidong Yang[1], Xunkuai Zhou[1,2], Yijun Huang[1], Chuanxiang Gao[1], Xi Chen[1] (✉), Ben M. Chen[1]

1. Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
2. School of Electronics and Information Engineering, Tongji University, Jiading, Shanghai, China

Abstract

Recent advances in artificial intelligence (AI) and cyber-physical systems have fostered innovative approaches to performance assessment and management of existing building stock. This study presents an AI-assisted digital twin (DT) framework for the automated and high-precision detection of façade defects in large-scale buildings. Leveraging unmanned aerial vehicles (UAVs) for visual data acquisition, the proposed framework integrates building information modeling (BIM) and geographic information systems (GIS) into a GeoBIM-assisted DT environment. An end-to-end pipeline is developed for defect localization and semantic registration, in which a virtual building model and camera geometry are constructed using geographic metadata. Synthetic views are generated to simulate real image capture conditions, enabling depth-based inference of each defect's spatial location. This facilitates the projection of defect data into georeferenced DT models. A dual-verification method combining image and geographic features is employed to eliminate duplicate detection across overlapping images, and structural context is retrieved via GeoBIM for semantic enrichment of defect information. The proposed system exemplifies the fusion of DT technologies with deep learning and cyber intelligence to enhance defect detection accuracy, resilience optimization, and timely building health monitoring. Experimental validation on a high-rise building in Hong Kong demonstrates the robustness and scalability of the framework, indicating strong potential for smart building maintenance and operation.

Keywords

digital twin
AI-assisted building simulation
UAV-based visual inspection
sustainable building maintenance
deep learning for detection

Article History

Received: 13 April 2025
Revised: 23 June 2025
Accepted: 15 July 2025

© The Author(s) 2025

1 Introduction

Periodic inspection is essential for maintaining the physical and functional conditions of civil infrastructure systems such as bridges, dams, roads, and buildings. For instance, since 2012, the Hong Kong government has initiated the mandatory building inspection scheme to address safety concerns arising from over 50% of private residences that have surpassed a 30-year lifespan (De Filippo et al. 2023). Visual inspection is a common approach aims to identify and locate potential defects caused by infrastructure degradation, such as cracks, spalling, and moisture, to prevent serious safety problems (Spencer et al. 2019). Traditional visual inspection methods rely on trained engineers for manual identification, characterized by high subjectivity, low accuracy, and low efficiency (See et al. 2017; Chen et al. 2023a).

The recent trend is to combine robotic technology, such as unmanned aerial vehicles (UAVs), with computer vision technology, such as defect detection algorithms, for automatic data collection and analysis (Rakha and Gorodetsky 2018; Agnisarman et al. 2019; Abouelaziz and Jouane 2024; Wang et al. 2024). On the one hand, UAVs equipped with cameras demonstrate significant advantages in terms of safety, cost-effectiveness, and maneuverability (Rakha and Gorodetsky 2018). Duque et al. (2018) distributed a national survey and find that UAV-enabled infrastructure inspection has been extensively applied and its feasibility has been

E-mail: byzhao@mae.cuhk.edu.hk (Benyun Zhao); xichen002@cuhk.edu.hk (Xi Chen)

substantiated. On the other hand, the large amount of data derived from efficient data collection requires rapid automated processing methods, specifically artificial intelligence (AI) algorithms (Liu et al. 2025). The deep learning-based object detection algorithms have been widely adopted for different structures (Zhang et al. 2023c; Jha and Babiceanu 2023; Zheng et al. 2025), including tunnels (Li et al. 2021), buildings (Zheng et al. 2020), and bridges (McLaughlin et al. 2020). The recent novel network models achieve better performance, including faster region-based convolutional neural networks (R-CNN) (Ren et al. 2017) and the You Only Look Once (YOLO) series (Jocher 2020; Ge et al. 2021; Li et al. 2022; Wang et al. 2022).

UAV-based imagery has become an essential tool for detecting surface defects on building façades. However, the identified defects on 2D aerial image data alone offer limited insight into the overall condition of the building. While 2D detection methods can localize defects, they often fail to provide the necessary spatial context for a comprehensive assessment of complex architectural structures. Zhang et al. (2023a) emphasized that the localization of damage on a building is crucial for accurate condition assessments. Many existing approaches focus on defect detection in 2D images without integrating these findings into a global 3D model of the building, limiting the usefulness of the detected data for broader structural evaluations. Furthermore, although building information modeling (BIM) models contain detailed architectural and structural information, their potential for defect management remains underutilized, as BIM's semantic information is rarely incorporated into these detection processes.

Beyond BIM, other approaches such as terrestrial laser scanning (TLS) (Mohammadi et al. 2023) and 3D reconstruction techniques like stereo photogrammetry (Jati 2021; Chen et al. 2024; Wang and Gan 2024) have been utilized to build accurate representations of building façades. Integrating structure from motion (SfM) with learning-based multi-view stereo (MVS) has enhanced these techniques, allowing for the efficient creation of detailed and cost-effective digital twin (DT) models of structures (Hosamo and Hosamo 2022; Chen et al. 2023d; Li et al. 2024). These DT models provide high-fidelity representations of the physical building, enabling a better understanding of its current condition (Yang et al. 2022). However, despite their geometric accuracy, these methods often ignore the integration of semantic information from BIM, which limits their ability to fully capture the complexity of building structures and fails to provide effective maintenance and repair strategies.

As illustrated in Figure 1, current defect detection methodologies predominantly identify cracks and defects in 2D images and project these findings onto a 3D model. This process facilitates surface-level defect localization but fails to address the deeper integration of detection data with the semantic structure of 3D building information modeling (BIM) models. For buildings with complex geometries, such as irregular or highly detailed wall surfaces, this traditional projection approach is insufficient for comprehensive structural evaluations. Existing limitations include the reliance on orthogonal images for accurate projection and the inability to incorporate overlapping images or irregular perspectives into the analysis. These gaps hinder the potential for a holistic assessment of architectural integrity, as crucial spatial and semantic relationships between defects and structural components are often disregarded. To overcome these challenges, recent studies have explored the integration of GIS with BIM, termed GeoBIM (Hajji and Oulidi 2022). GeoBIM combines the extensive spatial analytical capabilities of GIS with the rich semantic information provided by BIM, enabling more precise and context-aware management of infrastructure and built environment conditions (Liu et al. 2017; Moretti et al. 2021).

Our proposed system introduces a novel GeoBIM-assisted registration method, leveraging GIS and BIM to construct a DT environment. This approach allows for the precise registration of UAV-captured 2D defect data onto a 3D model enriched with semantic information. By incorporating depth calculations and semantic retrieval, our method achieves accurate defect localization even under challenging conditions, such as non-flat surfaces or overlapping images. This comprehensive solution advances the state of defect detection and registration, enabling component-level evaluations that guide targeted maintenance and repair decisions.

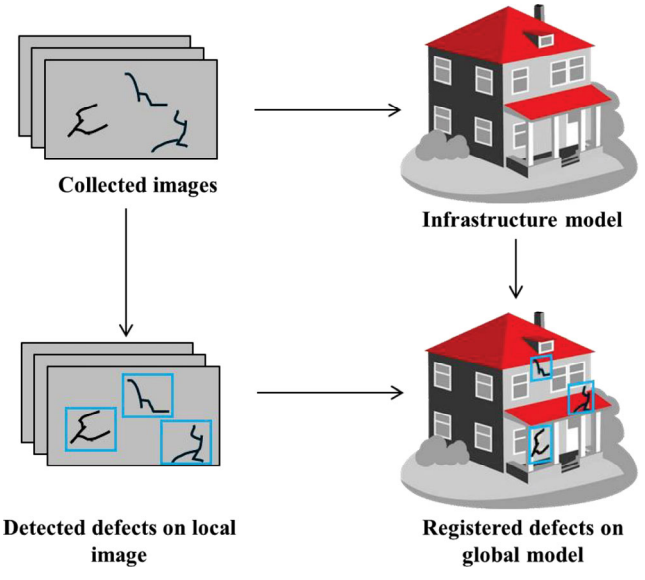


Fig. 1 The detected defects on local image and the corresponding defects registered on model with global reference

2 Related works

2.1 UAV-based visual inspection

Visual inspection serves as a valuable tool in detecting preliminary indicators of structural deficiencies in building façade, thereby averting potential severe risks (Alencastro et al. 2018). Moreover, visual inspection imposes minimal equipment requirement, thereby mitigating the costs associated with large-scale evaluations, such as those performed on high-rise buildings and bridges. However, traditional visual inspection requires substantial human resources and faces challenges such as data collection in hard-to-access areas and automated image data processing.

The progression of robotic technology, particularly the advent of UAVs, has significantly facilitated the visual inspection process for large structures. Incorporating advanced sensing technology, lightweight sensors, such as laser scanners, RGB or infrared cameras, can be equipped on UAVs for the capture of video or image data (Alencastro et al. 2018). It amplify the volume of data that can be collected within a brief time span, thereby integrating visual inspection efficiently into routine building maintenance tasks (Ruiz et al. 2022). However, the data collected by UAVs in visual inspection exhibits certain characteristics that presents considerable challenges for processing, including the vast quantity, the high degree of overlap (over 35%), multi-perspective views, and close-range to the surface (Chen et al. 2021).

The task of conducting precise and efficient defect detection on such image data has surpassed the capabilities of traditional manual identification methods. Recently, computer vision techniques have been widely applied in structural inspection tasks, offering significant advancements in automated defect detection. For instance, Li et al. (2021) successfully employed R-CNN for large-scale surface defect detection in tunnels, demonstrating the potential of deep learning models in infrastructure assessment. Given the increasing complexity and scale of inspection environments, real-time processing has become a key focus, leading to the frequent application of frameworks such as YOLO, known for its rapid object detection capabilities. Despite these advances, current methods primarily deliver detection results in the form of annotations on 2D images. While this approach is effective for recognizing defects in individual images, it lacks the ability to provide a global spatial context, which is particularly problematic in large-scale environments.

2.2 Digital twin for inspection and defect management

The task of constructing accurate digital models of buildings is critical in UAV-enabled inspections, especially when incorporating DT technology for infrastructure management. Several methods have emerged in recent years to facilitate model construction, including: (1) photogrammetry; (2) image-to-BIM projection models; (3) level of detail (LOD) models. Each of these methods presents distinct advantages and limitations in terms of accuracy, efficiency, and semantic richness.

Photogrammetry involves capturing the geometry and topology of buildings using aerial imagery. It has become widely adopted in UAV-based inspection workflows due to its relatively low cost, high efficiency, and ease of integration into automated processes. As demonstrated in recent studies, photogrammetry enables UAVs to gather high-resolution images, which are then processed into 3D models (Yu et al. 2022). However, while photogrammetry provides a fast and scalable solution, its derived models often lack detailed structural information, limiting their utility in more complex analyses. For example, the models generated may capture external geometries but fail to offer insights into the internal structure or material composition of the building. Chen et al. (2023d) emphasized that photogrammetry-based reconstructions face significant challenges due to environmental factors like lighting and surface texture, which often lead to inaccuracies in the generated point clouds. However, they have leveraged advancements in deep learning and MVS techniques to enhance both the completeness and accuracy of these models, making photogrammetry more reliable and applicable in a variety of DT applications.

Image-to-BIM modeling takes a different approach by leveraging existing BIM and integrating UAV-captured images into the BIM framework. This method involves aligning captured images with the BIM structure and superimposing additional data, such as optical or infrared imagery, onto the BIM for enhanced analysis (Hu et al. 2022; Zhang et al. 2023b). The primary advantage of this method lies in its ability to provide semantic information and structural insights, as BIM typically contains data about the materials, construction methods, and lifecycle of a building. However, one of the major limitations of current image-to-BIM methods is their reliance on flat surfaces. As Zhang et al. (2023a) pointed out, existing image-to-BIM approaches cannot accommodate non-flat surfaces with complex architecture. Nevertheless, image-to-BIM remains a valuable tool for inspections requiring high structural precision and semantic detail, especially in scenarios where the UAV is deployed in a controlled environment with predictable building geometries.

LOD models provide another means of digital representation by simplifying building geometries. LOD models prioritize real-time visualization and processing speed

by reducing the complexity of the geometric representation (Huang et al. 2020). These models are often used in applications where computational resources are limited, or when a high degree of accuracy is not required. For example, LOD models are frequently employed in city-scale simulations, where a simplified geometric representation is sufficient for planning or monitoring purposes (Hensel et al. 2019). However, LOD models lack the semantic and structural details needed for comprehensive inspections, making them unsuitable for more complex or large-scale buildings (Pantoja-Rosero et al. 2023). Furthermore, the simplification process can result in the loss of key features, such as façade details or intricate architectural elements, limiting their usefulness in precision-based tasks like defect detection or structural health monitoring.

Above methods can all construct DTs of the as-is building condition, but most work only focuses on geometric structure, such as the point clouds generated by photogrammetry. The data generated by different methods are incompatible and only viable within their own frameworks. The model built can seldom provide subsequent value for continuous or periodic inspection and management.

2.3 Defect registration

Defect registration refers to the process of mapping 2D defect information identified in images onto a 3D spatial model, providing a global reference framework to support decision-making (Artus et al. 2021). By integrating localized 2D defect data with the 3D global distribution of building structures, this technology enables comprehensive assessments that guide maintenance and repair strategies (Zhang et al. 2023a). As shown in Table 1, depending on the type of 3D model used as the carrier, existing methods can be categorized into three main types: (1) point cloud-based registration, (2) BIM-based registration, and (3) GIS-based registration. However, each has notable limitations in achieving high-precision registration with semantic fidelity, particularly in real-world architectural scenarios.

Point cloud-based registration uses dense spatial representations generated through UAV photogrammetry or laser scanning. These models effectively capture external geometry at high resolution (Valero et al. 2018). Zhang et al. (2022) employed finite-element methods to assess damage, and Chen et al. (2023b) annotated defect locations in point clouds. However, point clouds inherently lack semantic information, making it difficult to relate defects to specific structural elements or contextual attributes, thereby limiting their utility for semantic-level defect management.

BIM-based registration incorporates building information modeling (BIM) to integrate visual inspection data with semantically rich architectural models. These methods have demonstrated advantages in associating detected defects with structural metadata, such as component IDs, material types, and design specifications. Prior studies (Chen et al. 2019; Musella et al. 2021) have linked images to BIM elements through manual and automated region-of-interest extraction. More recent efforts (Tan et al. 2022; Zhang et al. 2023a) use coordinate transformations to project defect locations from UAV imagery into BIM space. However, such methods often assume flat, orthogonal façades and require cumbersome or imprecise coordinate conversion workflows, which reduce applicability to buildings with complex geometries.

GIS-based registration aligns images with real-world coordinates using geographic information systems. GIS excels in handling large-scale spatial data and integrating multi-source information (Xia et al. 2022). For instance, Chen et al. (2021, 2023c) used 2D façade unfolding techniques to register UAV images within GIS frameworks, simplifying visualization and documentation. Nevertheless, GIS approaches frequently reduce 3D geometry into flattened 2D maps, which undermines fidelity when mapping defects to non-planar surfaces or associating them with specific building components.

In summary, existing methods exhibit trade-offs among geometric accuracy, semantic integration, and practical applicability. Point cloud-based methods provide high

Table 1 Comparison between existing registration methods and our approach

Publication	Application	Method	Associated technology	Limitations
Chen et al. 2021	Geo-register 2D GIS spatial model of building façades	Geo-registration	GIS	Applicable only to flat surfaces with rich features; limited to 2D image mosaics
Tan et al. 2022	Project segmented defect image to BIM based on UAV GPS localization	Defect registration	BIM	Significant registration errors; dependent on path planning; limited to orthogonal, flat facades
Zhang et al. 2023a	Project image to BIM using improved generalised Hough transform	Image-to-BIM registration	BIM	Effective only for flat walls with unique surface features; limited robustness in complex scenes
Mohammadi et al. 2023	Integrate BIM with decision support system to provide high-fidelity model	Asset management	BIM	Did not address the problem of precise defect localization or semantic matching

spatial resolution but lack component-level interpretability. BIM-based approaches support semantic enrichment but are hindered by complex and error-prone registration procedures. GIS-based methods afford scalability but oversimplify 3D structures, limiting their descriptive power.

To address these methodological shortcomings, we propose a GeoBIM-based registration framework that unifies the spatial accuracy of GIS with the semantic richness of BIM. The novelty of our approach lies in the algorithmic design of a high-precision pose estimation module and a semantic matching mechanism that automatically associates detected defects with corresponding BIM components. By leveraging the georeferenced BIM model for camera pose correction, we achieve centimeter-level accuracy in defect localization, and by incorporating semantic metadata, we facilitate structured defect management and query. This enables precise, component-level mapping of defects directly onto the DT with both spatial and contextual fidelity—a capability that, to our knowledge, has not been comprehensively realized in prior frameworks. Our methodology not only integrates but also enhances existing processes to meet the rigorous demands of automated, scalable, and semantically enriched defect management in complex urban structures.

2.4 Research gaps and contributions

Building on the preceding literature review and addressing the key methodological concerns raised, we identify the following critical research gaps:

(1) Inadequate support for high-precision 2D-to-3D defect localization: Existing registration approaches, especially those based on BIM, rely on complex and error-prone coordinate transformations (e.g., from WGS-84 to local BIM coordinates) (Liu et al. 2019). These multi-stage projections introduce accumulated uncertainties, often compromising localization accuracy. Moreover, these methods do not sufficiently account for image perspective distortions or UAV-acquired oblique imagery, which are common in real-world inspections.

(2) Limitations in handling unordered and overlapping visual inputs: Current methods frequently depend on orthographic and non-overlapping images (Tan et al. 2022), which restricts their practical applicability. Overlap, although useful for structural completeness, leads to redundancy and ambiguity in defect interpretation. There remains a lack of robust solutions that can effectively utilize overlapping, non-perpendicular images to achieve precise and unambiguous defect localization.

(3) Lack of semantic integration in spatial registration frameworks: GIS-based registration methods often reduce complex 3D geometries to flat 2D surfaces (Chen et al. 2021, 2023c), making them inadequate for evaluating structures with irregular geometries. Moreover, traditional point cloud or SfM-based approaches lack semantic depth, hindering intelligent retrieval, component-level interpretation, and downstream decision support.

To address these gaps, we introduce a novel UAV-based GeoBIM-integrated DT framework for large-scale defect inspection. The major contributions of this study are summarized as follows:

(1) We design a high-precision defect registration framework that integrates UAV-based image acquisition, real-time defect detection, and GeoBIM-based semantic depth rendering. The registration pipeline introduces a calibrated pose estimation module to resolve positional inaccuracies, ensuring accurate 2D-to-3D localization.

(2) We develop a GeoBIM-based depth mapping strategy that provides sub-centimeter spatial accuracy by leveraging virtual camera rendering and physically aligned coordinate systems. Unlike conventional SfM-generated models, our approach benefits from high-quality BIM references and avoids distortions commonly found in point cloud reconstructions.

(3) We implement a semantic defect matching mechanism, aligning detected defects not only geometrically but also with the associated BIM structural components. This enhances traceability, enriches defect characterization, and supports intelligent maintenance planning.

(4) We construct an interactive, web-based visualization and data management platform using WebGIS, enabling intuitive interaction with defect data and structural semantics. The proposed system has been quantitatively validated across various real-world urban scenarios, demonstrating its effectiveness, scalability, and practical applicability for large-scale architectural maintenance.

3 Methodology

The basic flow chart of this registration method is shown in Figure 2, which starts with the aerial images collected by the UAV and ends up with a DT of the as-is building condition that integrates defects and a geometric model. The UAV-collected images have additional sensing data in two aspects: optical attributes such as POV, image size, etc., and the geographical state of the UAV, including attitude (from Inertial Measurement Unit (IMU)) and position (from Global Positioning System (GPS)).

With the aerial images, the 3D model (either photogrammetry model or BIM) is constructed by a 3D reconstruction algorithm or manual modeling method. Then, the depth texture is generated with the geo-referenced model and the corresponding original aerial image. The

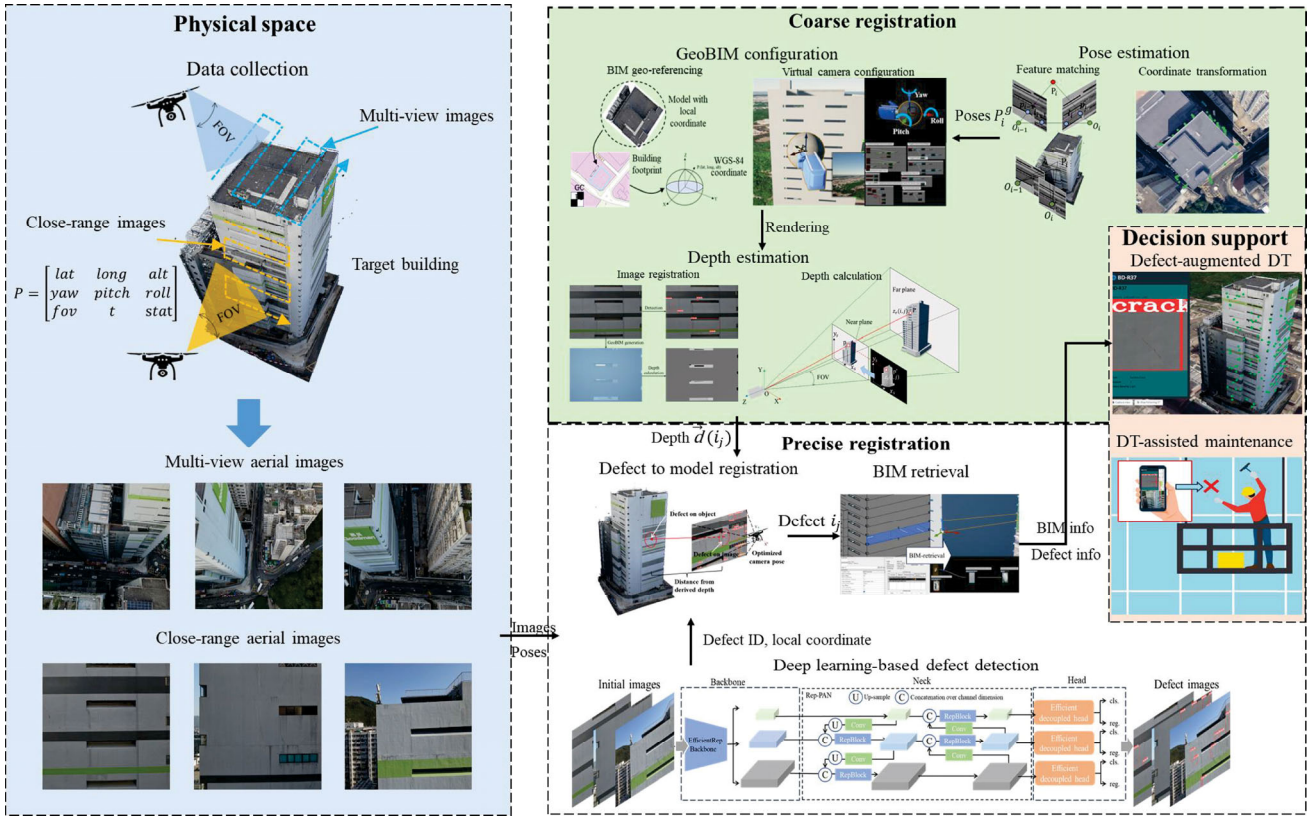


Fig. 2 A workflow of autonomous building façade defect detection and management under the DT framework

detected defects on the images are aligned with the distance information on the depth texture to calculate the global coordinate localization. Additionally, the alignment of BIM's semantic information with the geographic data ensures that each defect is matched to its corresponding structural information. Finally, the 2D defects are registered to the 3D model, creating a DT of the target building with its as-is conditions.

3.1 Coarse registration

This section introduces and outlines the GeoBIM approach based on the integration of GIS and BIM, which aligns physical objects with their virtual counterparts in the virtual space, ultimately resulting in the construction of a coarse DT environment. The coarse registration involves configuring the GeoBIM environment and processing the input images and UAV pose data. Through virtual camera registration, pixel-level depth values for the original images are calculated.

Although BIM is widely used as the dominant paradigm in architectural projects, BIM platforms are usually configured for the specific design, construction, and data management of unique buildings. As a result, their assistance for regional inspections, especially those that utilise unmanned systems and comprehensive spatio-temporal data, is significantly restricted. Through the process of geo-referencing, the integration of GIS with BIM establishes a global reference framework that enables BIM models to precisely correspond with real-world geographic data. Utilising ground control points (GCPs) with established geographic coordinates allows for geo-referencing of the BIM model to accurately represent the real spatial environment. By integrating the spatial accuracy of GIS with the comprehensive semantic information of BIM, a unified environment is established. This methodology not only surpasses the constraints of conventional BIM by facilitating wider geographic compatibility but also expands the use of BIM data to environmental analysis and infrastructure management, therefore augmenting its total usefulness and decision-making capacities.

Innovatively, we are the first to propose the application of the BIM+GIS approach in the realm of large-scale UAV-based visual inspections, aiming to construct the GeoBIM as the as-is DT representation of architectural defects. The proposed method for the primary DT configuration entails two fundamental steps: (1) GIS-based georeferencing, encompassing both the model and aerial photography POV, and (2) the acquisition of corresponding depth data to facilitate subsequent defect localization. GeoBIM integrates

the detailed structural and semantic information of BIM with the geospatial capabilities of GIS, providing a comprehensive framework for defect management and spatial analysis in building inspections.

3.1.1 *GeoBIM environment configuration*

This subsection proposes the processes of GeoBIM environment configuration with BIM, GIS and service attributes as in Figure 3. BIM attribute provides both the geometry model and the semantic metadata of the structures, while the GIS attribute mainly ensures that elements such as terrain, 3D assets, and world base maps in the environment are consistent with physical scenes. The GeoBIM platform is integrated based on Unreal Engine 4.27 (Sanders 2016), which is compatible with Cesium's GIS and BIM portals for the unreal API (Cesium 2023). In addition, the UAVs are modeling with the physical and optical settings.

3.1.2 *Pose estimation*

In practical UAV-based inspections, two types of aerial images are utilized: mult-iview aerial images and close-range images. Each serves distinct purposes and plays a complementary role in ensuring accurate defect detection and localization. For multi-view aerial images, they are captured from multiple angles and distances around the structure to provide a comprehensive view of the building. They are primarily used for 3D reconstruction through an incremental SfM framework, which generates a dense 3D point cloud model of the building. This process enables the derivation of global geometric features that serve as a contextual reference for aligning close-range images. The multi-view data compensates for the limited perspectives of close-range images and enhances the global pose estimation accuracy. For close-range images, they are collected at close proximity to the building façade, offering high-resolution detail critical for detecting surface defects. However, due to the constrained field of view and limited spatial context, close-range images may suffer from inaccuracies in pose estimation when used in isolation.

To overcome the limitations of close-range images, we integrate multi-view aerial images to provide global contextual features that refine pose estimation for close-range images. This involves matching key features between the two image types, aligning their respective pose data, and applying transformation matrices to ensure consistency within the 3D reconstruction model. The combined data approach leverages the global geometric accuracy of multi-view images and the detailed defect detection capability of close-range images, resulting in enhanced defect localization and registration accuracy. The data collection strategy is illustrated in physical space in Figure 2, showing the UAV's flight path for capturing both multi-view and close-range images. Multi-view images are captured through oblique photogrammetry along the UAV's broader flight path, while close-range images are taken along a façade-parallel path for detailed inspection. This coordinated flight plan ensures comprehensive data acquisition for both global and local tasks, laying the foundation for robust integration into the GeoBIM framework.

Using an incremental SfM framework, we calibrate the UAV's pose data, incorporating constraints derived from both multi-view and close-range images as in Figure 4. This optimization corrects GPS errors, mitigates misalignment issues, and establishes accurate camera poses. The optimized pose data is subsequently used to transform the close-range images into the global coordinate system, enabling precise defect localization and integration into the GeoBIM framework.

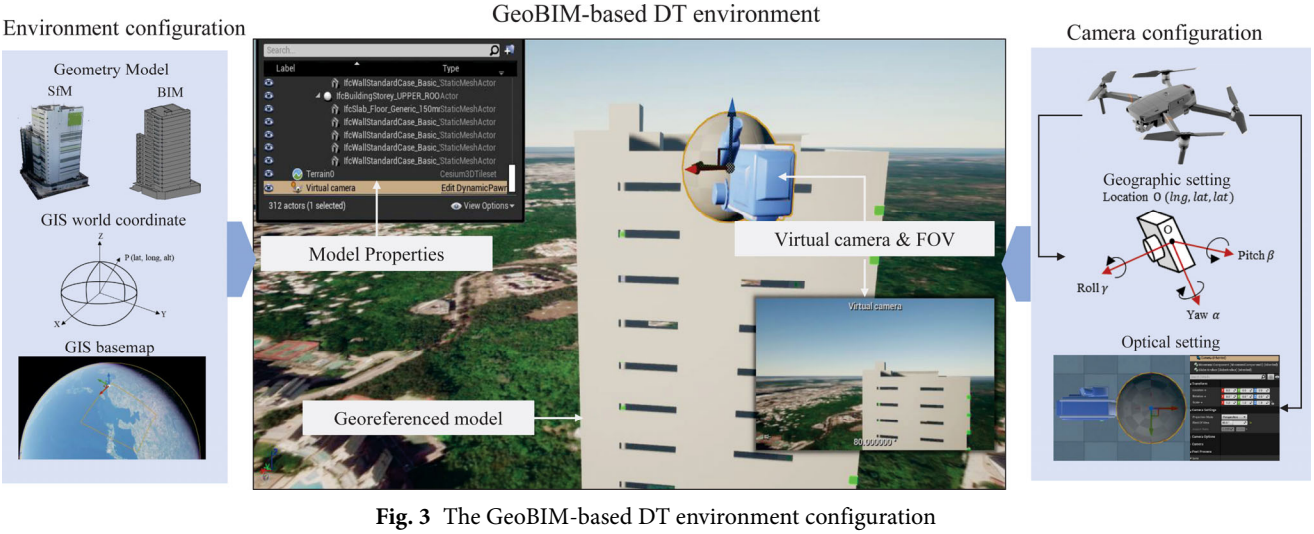


Fig. 3 The GeoBIM-based DT environment configuration

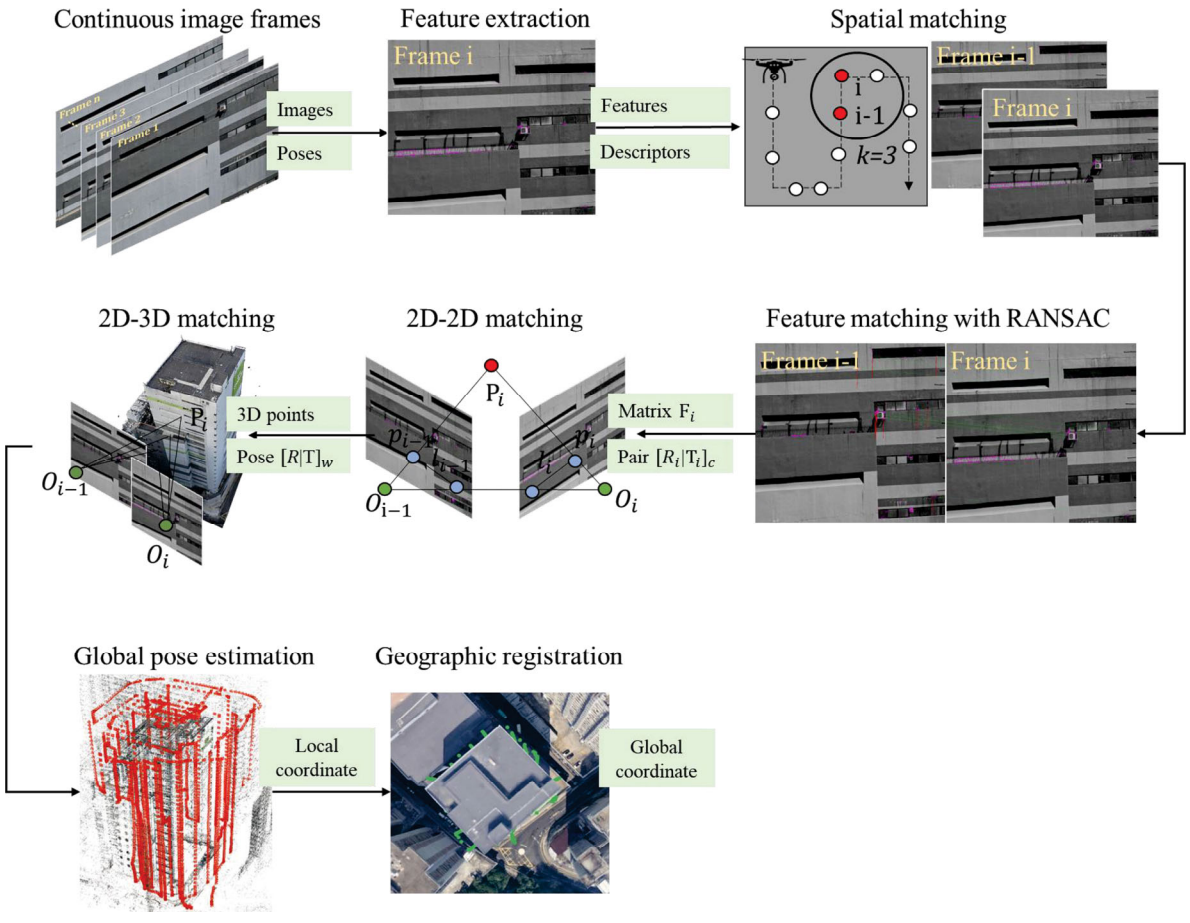


Fig. 4 Pose estimation

Firstly, for a large batch of consecutive frame images, the ORB algorithm provides fast and robust image feature extraction. Subsequently, spatial matching selects two consecutive frames, $i-1$ and i, for initialization. The matching of consecutive frames involves fundamental matrix calculation using the random sample consensus (RANSAC) 8-point method to eliminate outliers. The resulting image pairs and matching relationships are used for 2D-2D matching based on epipolar geometry, obtaining the relative poses of image pairs. Then, triangulation is employed to generate 3D points, establishing 2D-3D matching relationships, and utilizing the PnP algorithm to solve the camera's poses.

In practice, it is prone to outliers in the matching process, i.e., incorrect matches of feature points. To address this issue, the normalized 8-point algorithm with RANSAC is introduced. It comprises the following steps: randomly select eight pairs from all matched keypoints in the image pair, compute the corresponding fundamental matrix $\boldsymbol{F}_i$ (since there is no need to consider the case where the camera centers of matched frames are the same, there is no need to involve the homography matrix $\boldsymbol{H}$); then, calculate the Sampson distance errors for all pairs of matching points with $\boldsymbol{F}_i$, and if the error is less than a threshold, it is considered an inlier; repeat the above steps until the maximum number of iterations is reached to obtain the optimal match with the maximum number of inliners.

The calculation formula from epipolar constraint for matrix $\boldsymbol{F}$ is as follows:

$$\begin{aligned} &F' = \arg\min_{\det F'=0} \left\| \boldsymbol{F} - \boldsymbol{F}' \right\| \\ &\begin{cases} \boldsymbol{p}_i^{\top} \boldsymbol{F}_i \boldsymbol{p}_{i-1} = 0 \\ \boldsymbol{F}_i = \boldsymbol{K}^{-\top} \cdot \boldsymbol{t}^{\wedge} \cdot \boldsymbol{R} \cdot \boldsymbol{K}^{-1} \end{cases} \end{aligned} \tag{1}$$

where, $\boldsymbol{p}_i$ and $\boldsymbol{p}_{i-1}$ are the corresponding points of the physical point $\boldsymbol{P}$ in the image pair; $\boldsymbol{K}$ refers to the intrinsic matrix of camera; $\boldsymbol{t}^{\wedge}$ is the skew-symmetric of translation matrix from frame $i-1$ to i; $\boldsymbol{R}$ is the rotation matrix from frame $i-1$ to i.

The calculation formula for Sampson distance to filer the outliers is as follows:

$$\begin{cases} d(\boldsymbol{p}_{i-1}, \boldsymbol{p}_i) = \dfrac{\boldsymbol{p}_i^{\top} \boldsymbol{F}_i \boldsymbol{p}_{i-1}}{(\boldsymbol{F}_i \boldsymbol{p}_{i-1})_x^2 + (\boldsymbol{F}_i \boldsymbol{p}_{i-1})_y^2 + (\boldsymbol{p}_i^{\top} \boldsymbol{F}_i)_x^2 + (\boldsymbol{p}_i^{\top} \boldsymbol{F}_i)_y^2} \\ d(\boldsymbol{p}_{i-1}, \boldsymbol{p}_i) < \tau \end{cases} \tag{2}$$

where, τ is the maximum error for inliners as an approximate estimate from $\boldsymbol{p}_i$ to $\boldsymbol{F}_i\boldsymbol{p}_{i-1}$. Then the number of the corresponding inliers is recorded in this iteration. Given the ratio of inlier is 0.5, the iteration time is set as 1000 to achieve the matching rate of 0.99. The transformation matrix between the image pairs can be decomposed from matrix $\boldsymbol{F}$. Then the poses of following frames are solved by Perspective-n-Point (PnP) algorithm. The corresponding formula is as following:

$$[\boldsymbol{R}|\boldsymbol{t}] = \underset{R\in SO(3),t\in\mathbb{R}}{\arg\min} \sum_{i=1}^{n} \min\left(\left\| \boldsymbol{p}_i - \pi(\boldsymbol{R}\boldsymbol{P}_i^w - \boldsymbol{t}) \right\|^2, d^2 \right) \tag{3}$$

where, function π is the projection function to project the 3D point $\boldsymbol{P}_i$ to pixel coordinate and the threshold d is to filter the outliers. Hence, the optimized poses for all the current frames are updated to the camera property in GeoBIM library.

Through the preceding steps, we have filtered and calibrated the poses of each frame; however, these poses are defined within a local coordinate system. To be more precise, the current 3D spatial coordinate system is a Cartesian coordinate system represented by the matrix $[\boldsymbol{R}|\boldsymbol{t}]$. Recognizing that transformation matrix calculations cannot be directly performed in the WGS84 coordinate system, we establish an initial earth-centered, earth-fixed (ECEF) coordinate system through a 3D similarity estimation with more than 3 settled frames. Subsequently, we calculate the poses of other frames in the ECEF coordinate system as $\boldsymbol{P}_i^{\mathrm{e}} = \{X_i, Y_i, Z_i\}$. Finally, these poses are transformed into WGS84 coordinates $P_i^{\mathrm{g}} = \{lng_i, lat_i, alt_i\}$, resulting in optimized global pose localization information. The transformation formula from ECEF coordinate to WGS84 coordinate is as follows:

$$\begin{cases} p = \sqrt{X_i^2 + Y_i^2} \\ N = \dfrac{a}{\sqrt{1 - f(2-f)\sin^2(lat_i)}} \\ lng_i = \arctan\dfrac{Y_i}{X_i} \\ alt_i = \dfrac{p}{\cos(lat) - N} \\ lat_i = \arctan\left| \dfrac{z}{p}\left(1 - \mathrm{e}^2 \dfrac{N}{N + alt_i}\right)^{-1} \right| \end{cases} \tag{4}$$

In our framework, 3D model construction is based on SfM's pose estimation, which enhances UAV positioning accuracy and serves as the basis for model construction. This is followed by a learning-based MVS method to generate a dense point cloud (Yang et al. 2023). Additionally, the framework is flexible, allowing alternative 3D reconstruction algorithms, provided the models maintain necessary accuracy and completeness. The integration with GeoBIM ensures that this flexibility does not affect the system's overall functionality, preserving its effectiveness for tasks such as defect localization.

3.1.3 *Image registration and depth calculation*

The fundamental principle of 3D engine visualization rendering is to transform a scene represented in 3D into a 2D form. The coordinate processing involves various stages such as modeling transformation, viewing transformation, projection transformation, perspective transformation, and viewport mapping (De Vries 2015). Through these transformations, object coordinates progress through multiple spaces: local space, world space, eye space, clip space, normalized device coordinate (NDC) space, and finally screen space. Each space serves a specific role in the rendering pipeline. Local space defines an object's geometry relative to itself, while world space positions it in a global scene for interaction. Camera space aligns the global scene relative to the camera's perspective, determining what is visible. Clipping space projects the scene onto a 2D plane while removing objects outside the camera's view. NDC space normalizes these coordinates into a standard range, and screen space maps them to pixel positions for visualization. These transformations ensure accurate visualization of 3D scenes and facilitate the integration of UAV-collected data into the GeoBIM framework. The entire process is illustrated in Figure 5.

To obtain the distance from the camera to an object, i.e., the depth value, it is necessary to perform a reverse calculation based on the depth texture acquired in screen space. Given that the transformation from local space to eye space (camera space) does not involve scale changes, the depth transformation process can be focused solely on the transition from screen space to view space. As in Figure 8(e), it's required to calculate the depth in eye coordinate $z_{\mathrm{e}}(i, j)$ from the screen coordinate $z_{\mathrm{s}}(i, j)$, where (i, j) is the pixel coordinate of the depth image.

The depth transformation between the NDC space $z_{\mathrm{n}}(i, j)$ and the screen space $z_{\mathrm{s}}(i, j)$ is as follows:

$$z_{\mathrm{n}}(i,j) = \frac{2z_{\mathrm{s}}(i,j) - (f_{\mathrm{s}} + n_{\mathrm{s}})}{f_{\mathrm{s}} - n_{\mathrm{s}}} \tag{5}$$

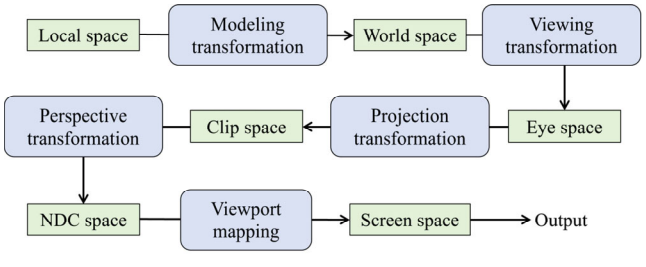


Fig. 5 The coordinate transformation pipeline for 3D to 2D rendering

where f_s is the far plane and n_s is the near plane on screen space with the default value $f_s = 1$ and $n_s = 0$. Accordingly, the depth transformation between the NDC space and the eye space is as follows:

$$z_e(i,j) = \frac{2fn}{(f-n)z_n - (f+n)} \tag{6}$$

As a result, from screen space to view space, the depth value are calculated as the formula:

$$z_e(i,j) = \frac{n}{z_s(i,j)(f-n) + f} \tag{7}$$

3.2 Precise registration

Coarse registration establishes the GeoBIM environment, allowing the creation of virtual BIM images and physical depth maps that accurately align with the POV of the physical photographs. Achieving global defect registration and semantic annotation requires further refinement through precise registration. This process encompasses the following objectives: (1) the detection and local positioning of defects; (2) the registration of individual defects onto the model; (3) the employment of semantic retrieval for structure alignment.

The core of this approach is the automation of defect detection and registration, achieved through the analysis of UAV-captured images. Each detected defect is cataloged, linked to a unique identifier within the architectural component, and integrated into the system. This method overcomes traditional limitations related to the UAV's POV, the characteristics of the detection subject, and the photographic technique. It allows flexibility in UAV positioning relative to irregular wall surfaces, eliminating the need for perpendicular angles, and does not impose strict requirements on image overlap, offering an efficient solution for managing architectural defects.

3.2.1 *Deep learning-based defect detection*

To ensure the generalizability and reproducibility of the proposed GeoBIM-based defect registration framework, we adopt a modular approach to defect detection that allows seamless integration with existing or future deep learning models. While defect detection is not the core innovation of our work, we recognize its pivotal role in the entire workflow and thus provide a comprehensive presentation of the dataset, algorithm selection, and evaluation procedures. Our aim is to minimize ambiguity and support flexible deployment of our pipeline across diverse use cases and detection models.

1) Dataset construction

The effectiveness of current learning-based methods in defect detection for large-scale infrastructures is significantly hindered by the lack of a high-quality open-source dataset. To bridge this gap, we present CUBIT-Det, the first high-resolution dataset specifically designed for detecting various defects in extensive infrastructures (Zhao et al. 2024). This dataset includes 5527 images captured by unmanned systems, with a remarkable maximum resolution of 8000 × 6000. The dataset's defect images are taken from multiple angles and distances under various lighting conditions, offering a comprehensive array of structural details. This variety ensures the robustness of models in practical inspection scenarios. The dataset covers the three most common types of infrastructure: buildings (65%), pavements (29%), and bridges (6%), focusing on the inspection of three primary defect types: cracks (82%), spalling (12%), and moisture (6%), as shown in Figure 6.

2) Real-time detection and localization

Based on the self-established dataset, we conduct evaluations on a multitude of state-of-the-art (SOTA) learning-based real-time object detection algorithms to ascertain the optimal solution of the task of defect detection in terms of both speed and accuracy. We train and test 12 SOTA series algorithms (nearly 30 models): YOLOv5 (Jocher 2020), YOLOv6 (Li et al. 2022), YOLOv7 (Wang et al. 2022), YOLOv8 (Jocher et al. 2023), YOLOX (Ge et al. 2021), PP-YOLO (Long et al. 2020), PP-YOLOv2 (Huang et al. 2021), PP-YOLOE (Xu et al. 2022), PP-YOLOE+ (Xu et al. 2022), MobileViT (Mehta and Rastegari 2021), RT-DETR (Zhao et al. 2023) and Faster R-CNN (Ren et al. 2017).

3) Evaluation metrics

Precision (P), Recall (R), and Average Precision (AP) are the three most commonly used metrics in object detection for infrastructure defect detection. Precision measures the accuracy of detected defects, denoting the ratio of correctly identified defects to all detections made by the model. Recall, on the other hand, assesses the rate of missed detections, indicating the proportion of correctly identified defects among all actual defects. Precision and Recall are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

The AP metric represents the weighted mean of precision scores at each threshold on the precision–recall (PR) curve, using the increase in recall from the previous threshold as

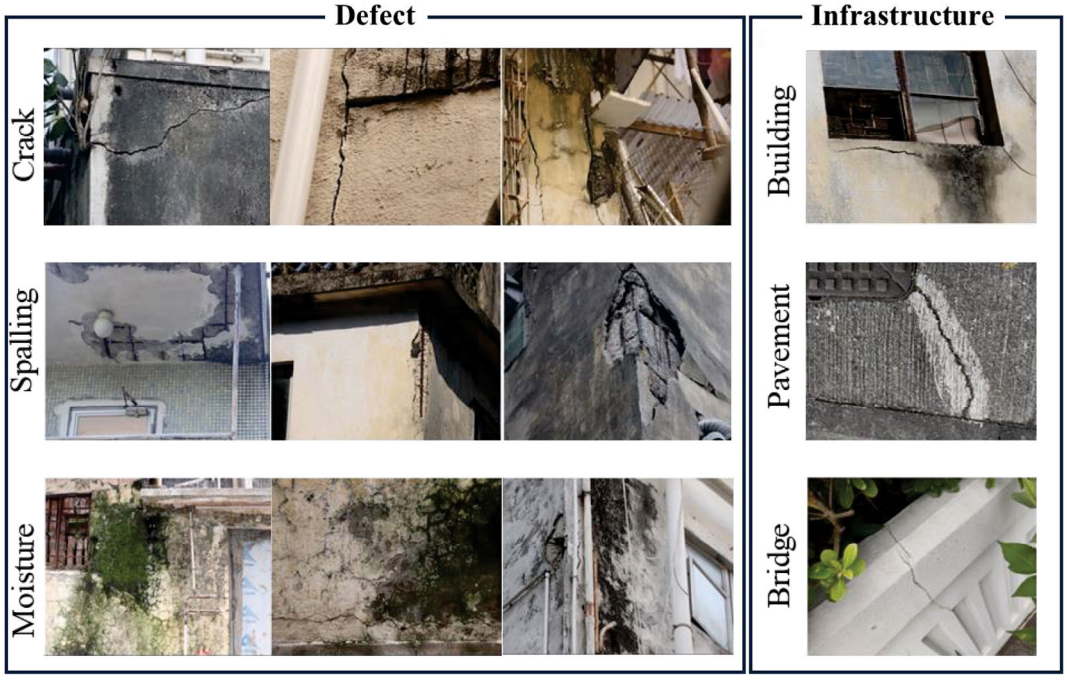


Fig. 6 Self-established dataset for infrastructure defect detection. The defect category includes crack, spalling, and moisture, and the infrastructure category includes building, pavement and bridge

the weight. Given the multi-category nature of our detection task, we calculate the AP for each category and then compute the mean Average Precision (mAP) across all categories. The equations for AP and mAP are provided below. Here, AP_i represents the AP for class i, and m is the total number of classes.

$$\begin{aligned} \mathrm{AP} &= \int_0^1 p(r)\mathrm{d}r \\ \mathrm{mAP} &= \frac{1}{m}\sum_{i=1}^{m}\mathrm{AP}_i \end{aligned} \tag{10}$$

Unlike traditional integral methods used to calculate AP, the computation of AP in MS COCO involves a discretization process: the PR curve is defined as the average of precision values at a set of 101 evenly spaced recall levels [0, 0.01, ..., 1] (from 0 to 1, with the increments of 0.01). The equations for mAP in MS COCO is shown below:

$$\mathrm{mAP}_{\mathrm{COCO}} = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{1}{101}\sum_{r\in(0,0.01,\ldots,1)} p\times interp(r)\right] \tag{11}$$

4) Model selection

We visualize the inference time versus $AP_{0.5:0.95}$ of the selected SOTA models in Figure 7. For all series of algorithms, as the model size increases, the inference speed will decrease while the detection capability will improve. However, there is a bottleneck in detection capability, which means that simply enlarging the model to realize the enhancement of detection performance cannot always be effective. From the top-left corner of Figure 7, it becomes clearer that YOLOv8 (Jocher et al. 2023) (red star) networks demonstrate a fabulous trade-off between accuracy and latency on large-scale infrastructure defect detection task.

YOLOv8 (Jocher et al. 2023) builds upon the YOLO series with a fully anchor-free design, predicting object bounding boxes directly from feature maps without relying on predefined anchor templates. It introduces a decoupled head architecture that separates classification and localization tasks, improving convergence and accuracy. YOLOv8 also integrates advanced techniques such as dynamic label assignment and a simplified backbone, enabling more accurate and efficient detection of defects like cracks, spalling, and moisture, even under varied lighting and surface conditions. For each defect, YOLOv8 (Jocher et al. 2023) outputs a bounding box with four parameters: x and y coordinates of the center, and the box's width and height. These parameters, normalized to the image's dimensions, offer a scalable object localization method. Alongside these spatial parameters, the model also outputs a confidence score reflecting the model's certainty in the detection, as well as class probabilities indicating the type of defect detected. The result is a set of bounding boxes, each associated with a defect type and its relative location within the image, providing critical data for subsequent analysis and rectification in architectural maintenance and restoration efforts.

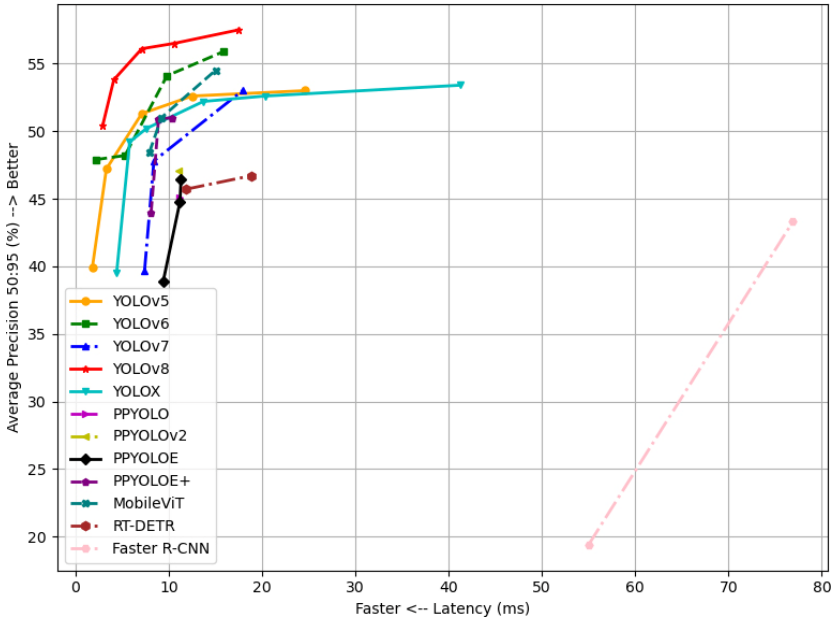


Fig. 7 Trade-off performance of different models about inference time versus $AP_{0.5:0.95}$ trained on CUBIT-Det dataset. The further the point is toward the top-left corner, the stronger the detection capability and the shorter the inference time

3.2.2 *Individual defect to model registration*

A key step in integrating and managing defect detection outcomes—such as defect images, local positioning data, and classification—is the registration of these results. This process involves determining global positions and eliminating redundant defects. Registration ensures that each detected defect is accurately mapped within the GeoBIM framework, enabling precise localization across the entire structure. By aligning and consolidating detection results, this step lays the foundation for building a comprehensive DT of architectural defects, crucial for effective maintenance and remediation strategies.

We follow the geographic transformation paradigm shown in Figure 9 to project the detected defects from the image coordinate system onto the geo-referenced model. In previous sections, we derived the corresponding detected images, GeoBIM images, and depth images within the GeoBIM environment (see Figures 8(a)–(d)). The defect images contain the origin images with defects marked by red bounding boxes, while the local coordinates and defect size on the image plane are recorded in the data library. For the j^{th} defect in the i^{th} image (denoted as defect i_j), we first capture the geographic coordinate of the image center O based on the pose estimation results, denoted as P_i^g. Instead of incorrectly describing the distance $\vec{d}(i_j)$ as the projection from O to O'', this vector actually represents the depth from the defect i_j to the model surface, as inferred from the depth map generated in the GeoBIM environment. The vector $\vec{d}(i_j)$ is aligned with the depth axis and provides the distance from the defect to the wallsurface. The direction is consistent with $\overrightarrow{OO''}$, but the magnitude reflects the depth from the defect to the 3D model. Subsequently, we compute the relative distance between O'' and the center of the defect's bounding box, converting it into a metric distance along the tangential vector $\vec{l}(i_j)$ from O'' to the defect i_j. This allows us to determine the global position of the defect in the 3D space. The corrected calculation process is reflected in Algorithm 1.

Algorithm 1 Individual defect registration

Require: Geographical coordinates $P_i^g(lon_i, lat_i, alt_i)$; defect distance $z_e(i,j)$; projection vector $\vec{n}$

Ensure: Global location $g(i,j) = (lng_{i,j}, lat_{i,j}, alt_{i,j})$

for $i = 1$ to i_{max} **do**
  Compute projection point: $P_i^{g*} \leftarrow P_i^g + z_e(i,0) \cdot \vec{n}$
  for $j = 1$ to j_{max} **do**
    $L_D(i) \leftarrow 2z_e(i,0) \cdot \tan\left(\frac{FOV}{2}\right)$
    $\vec{L}(i,j) \leftarrow L_D(i) \cdot \frac{l(i,j)}{l_D(i)}$
    $g(i,j) \leftarrow P_i^{g*} + \vec{L}(i,j)$
    $T \leftarrow \sqrt{(0.5 / \text{unit}_{lng})^2 + (0.5 / \text{unit}_{lat})^2}$
    for $k = 1$ to $i - 1$ **do**
      for $t = 1$ to $j_{max} - 1$ **do**
        if $\|g(k,t) - g(i,j)\| > T$ **then**
          Mark $g(i,j)$ as a new defect
        end if
      end for
    end for
  end for
end for
return $g(i,j)$

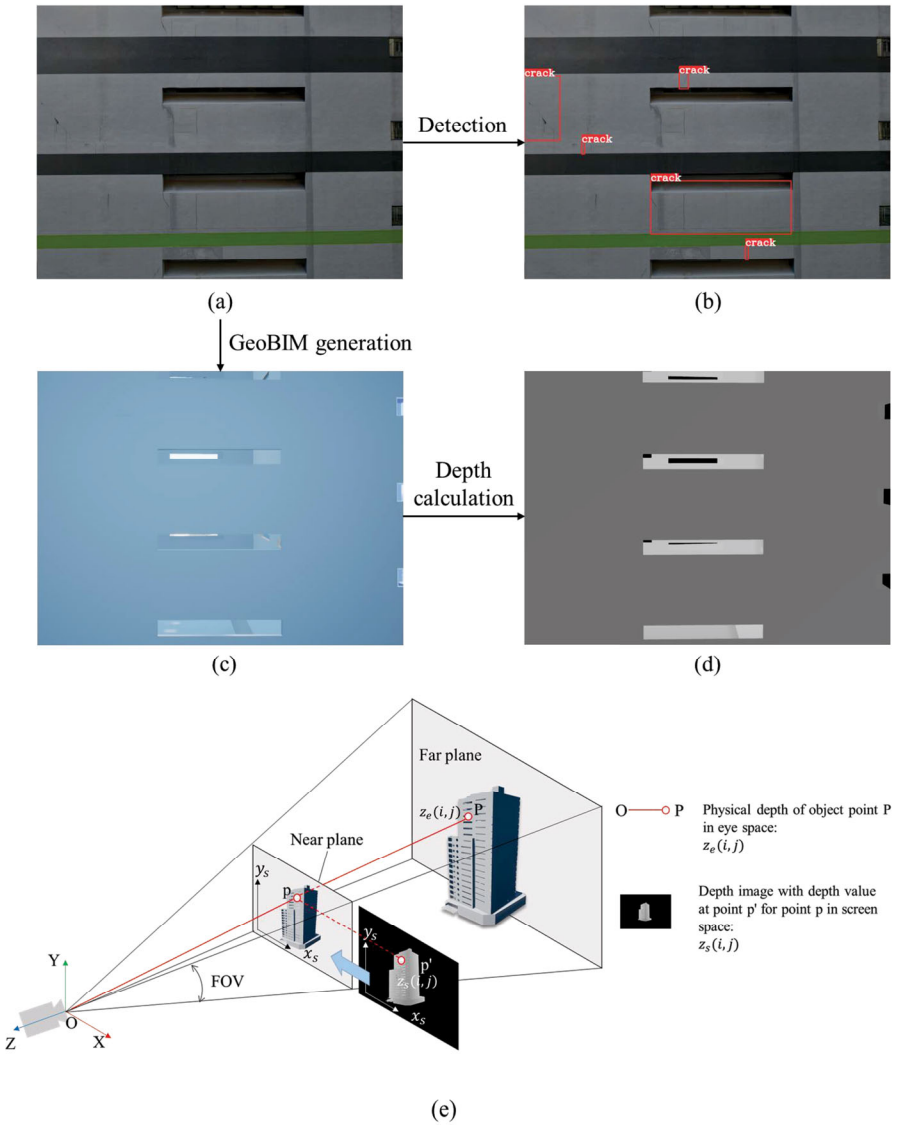


Fig. 8 Corresponding images from different sources: (a) original aerial image; (b) detected image; (c) GeoBIM generated image; (d) GeoBIM derived depth image. (e) The transformation process to compute the physical depth $\overrightarrow{OP}$ from image to the building surface

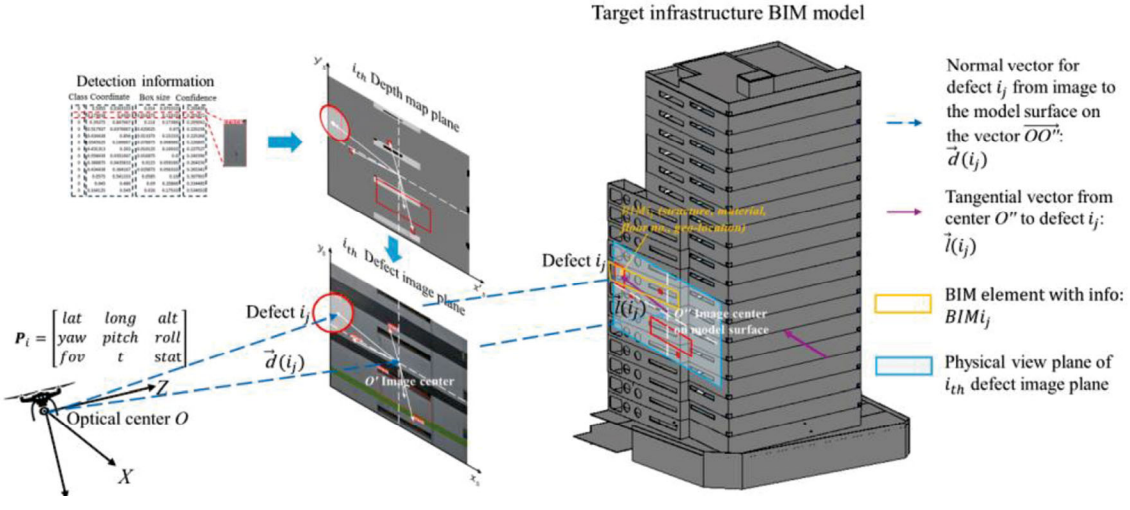


Fig. 9 The projection of defects from image to model

3.2.3 *Semantic retrieval for semantic matching*

Global registration furnishes the geographical coordinates of defects, facilitating the establishment of a one-to-one correspondence between these defects and the geospatially coupled architectural structures within the GeoBIM environment. In this section, we will expound on the methodology for conducting structural retrieval correlated with identified defects through GeoBIM, aiming to construct an assessment of building defects at the structural level. Given that GeoBIM incorporates metadata from BIM and introduces GIS data as well as inspection data from UAVs, as illustrated in Figure 8. Utilizing collected images along with their fully corresponding GeoBIM-derived images allows for the retrieval of BIM semantic information corresponding to elements present in the images.

This approach not only enhances the precision of defect assessments but also contributes to the strategic allocation of resources for infrastructure maintenance, underscoring the critical role of GeoBIM in advancing the state-of-the-art in architectural defect management.

As depicted in Figure 10, while BIM provides essential structural prior knowledge for constructing a building's DT, it does not automatically link to the defects detected within the structure. To bridge this gap, we have developed an automated workflow, which systematically facilitates the integration of BIM with detected defects and other relevant data. This process is entirely automated, as explained below.

(1) BIM registration: This initial step involves aligning and geo-referencing the BIM data with real-world physical data gathered from the site. Using Unreal Engine's

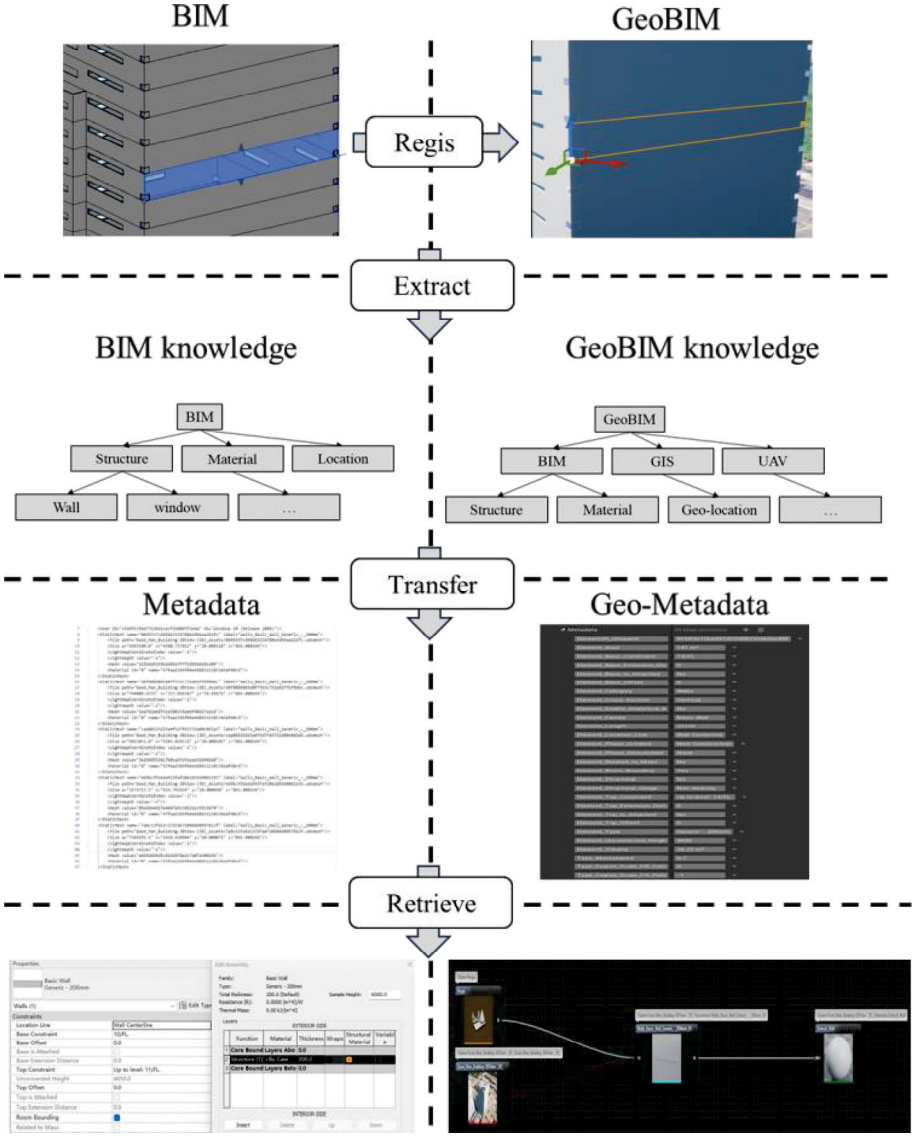


Fig. 10 The hierarchy of corresponding BIM and GeoBIM

Datasmith API, we automate the registration process by transferring BIM metadata into the GeoBIM environment. The geo-referencing process, as discussed in Section 3.1.3, assigns geographic coordinates to the BIM model, ensuring that its semantic structures are aligned with the actual geographic location. This allows for precise overlay of detected defects on the GeoBIM model, which forms the foundation for further analysis.

(2) Knowledge extraction: Once the BIM registration is completed, we automatically extract semantic information from the BIM model using the same Datasmith API. This extraction process parses data such as material properties, structural components, and spatial hierarchies (e.g., floor levels, orientations) to construct a knowledge base. This information is critical for understanding the context of the building's structure and linking it to the detected defects.

(3) Metadata transference: Through Unreal Engine's metadata transfer capabilities, both static BIM data and dynamic UAV data (such as flight paths, timestamps, and aerial defect logs) are transferred and stored in a consolidated metadata system. By automating this transference process, the system ensures that each defect is linked to relevant geographic and semantic information, forming a traceable record of the building's condition. All defect locations and their corresponding high-precision geographic information, as calculated and discussed earlier in the manuscript, are archived in the database.

(4) Retrieval mechanism: With metadata from both the static BIM model and UAV-acquired dynamic data properly aligned and registered, the system automatically allows semantic retrieval. Defects can be queried based on their geographic location and semantic context, such as the material type or structural component affected. By leveraging the metadata's geo-referenced alignment with BIM, the system provides a fully automated mechanism for defect localization and retrieval based on semantic and geographic information. Users can retrieve defect details, such as the defect's severity and material specifications, enabling comprehensive analysis and decision-making. This automated retrieval mechanism ensures that defects are correctly localized within the broader context of the structure, providing actionable insights for maintenance and monitoring.

Through these interlinked processes, the automated workflow enhances the building DT with detailed defect-related information and transforms it into a dynamic tool for ongoing structural health monitoring. The entire process, from metadata extraction to defect localization, is fully automated, ensuring precision and efficiency.

4 Implementation

We deploy our proposed inspection framework on various large-scale scenarios to verify its effectiveness and efficiency. Here, we take a large-scale high-rise warehouse (36 m × 27 m × 100 m) as a representative instance.

To verify the effectiveness and efficiency on real large-scale scenarios, we have deployed the method on a large-scale high-rise warehouse. Figure 11 illustrates the application of our methodology to a commercial building which rises to a height of 100 meters and spans an area of approximately 27 m × 36 m, located in the Shatin district of Hong Kong. This 18-story structure was extensively surveyed using three DJI Mavic 2 drones, each equipped with a camera capable of capturing images at a resolution of 8000 pixel × 6000 pixel. These UAVs were deployed to

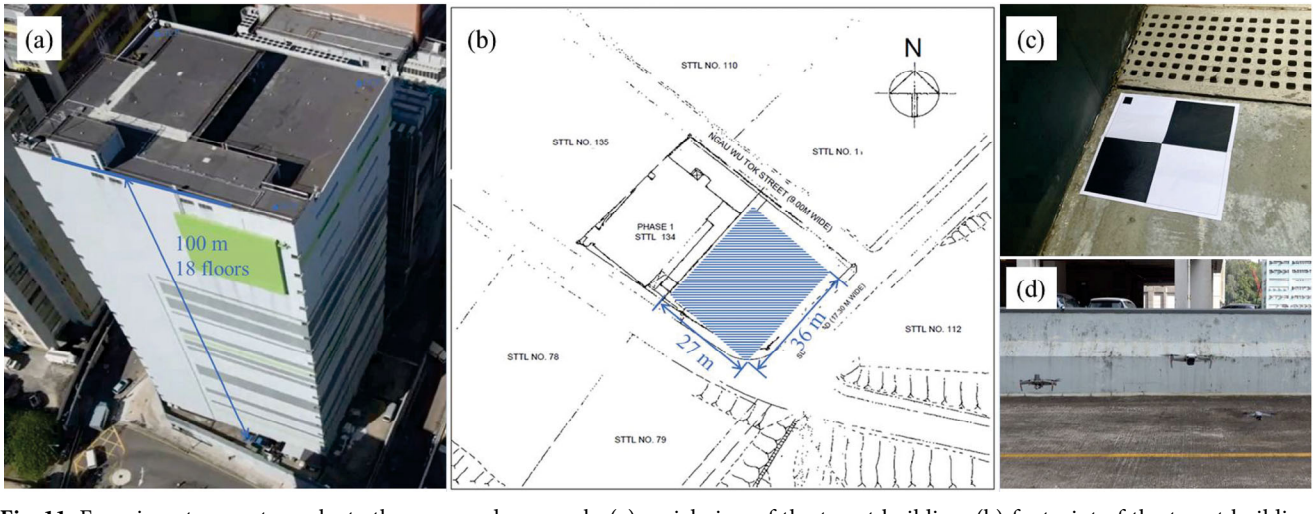


Fig. 11 Experiment scene to evaluate the proposed approach: (a) aerial view of the target building; (b) footprint of the target building; (c) GCP on the building; (d) multiple UAVs used in data collection

collect over 1000 aerial photographs to facilitate a detailed analysis of the building's features and conditions. The specific locations from which these images were acquired are detailed in Figure 11(b).

To process this substantial amount of high-resolution data, the study utilized advanced computing hardware, comprising an Intel(R) Core(TM) i9-10920X CPU and an NVIDIA GeForce RTX 3090Ti GPU. This setup was chosen to ensure robust and efficient handling of the data, enabling precise and timely analysis of the structural integrity of the building.

4.1 Field experiment results

In the field experiments, the primary focus was on the collection of data via UAVs, with specific attention to ensuring image quality, data completeness, and flight safety. To facilitate this, flight paths were meticulously planned based on the GeoBIM surface model of the structure. Typically, this planning necessitates maintaining the UAVs in a perpendicular orientation to the building's walls while keeping an approximate distance of 10 meters to optimize image capture and data accuracy.

Following these guidelines, the UAVs executed their flights along predetermined trajectories, effectively adhering to the designed flight paths. To enhance the efficiency of data collection and mitigate potential issues such as battery depletion, three drones were deployed simultaneously. This strategy allowed for comprehensive aerial coverage of the target building's surface, achieving complete data acquisition within a span of thirty minutes. This coordinated approach not only maximized the productivity of the data collection phase but also ensured the safety and reliability of the operational process.

4.1.1 Results of individual defect registration

As proposed in Section 3, the registration process involves projecting individual defect images from their original GPS locations to geo-referenced 3D model. Specifically, this is achieved by utilizing GeoBIM to generate corresponding coarse registration images and employing depth maps to ascertain physical distances, thereby accurately localizing each defect onto the model.

We collected 1,016 aerial images at a resolution of 8000 × 6000, covering the entire exterior surface of the building. These images were then processed for detection and registered to the corresponding SfM model for visualization and evaluation. Figure 12 presents the sequence of intermediate results generated throughout this process. According to evaluations by building inspection experts, the detection

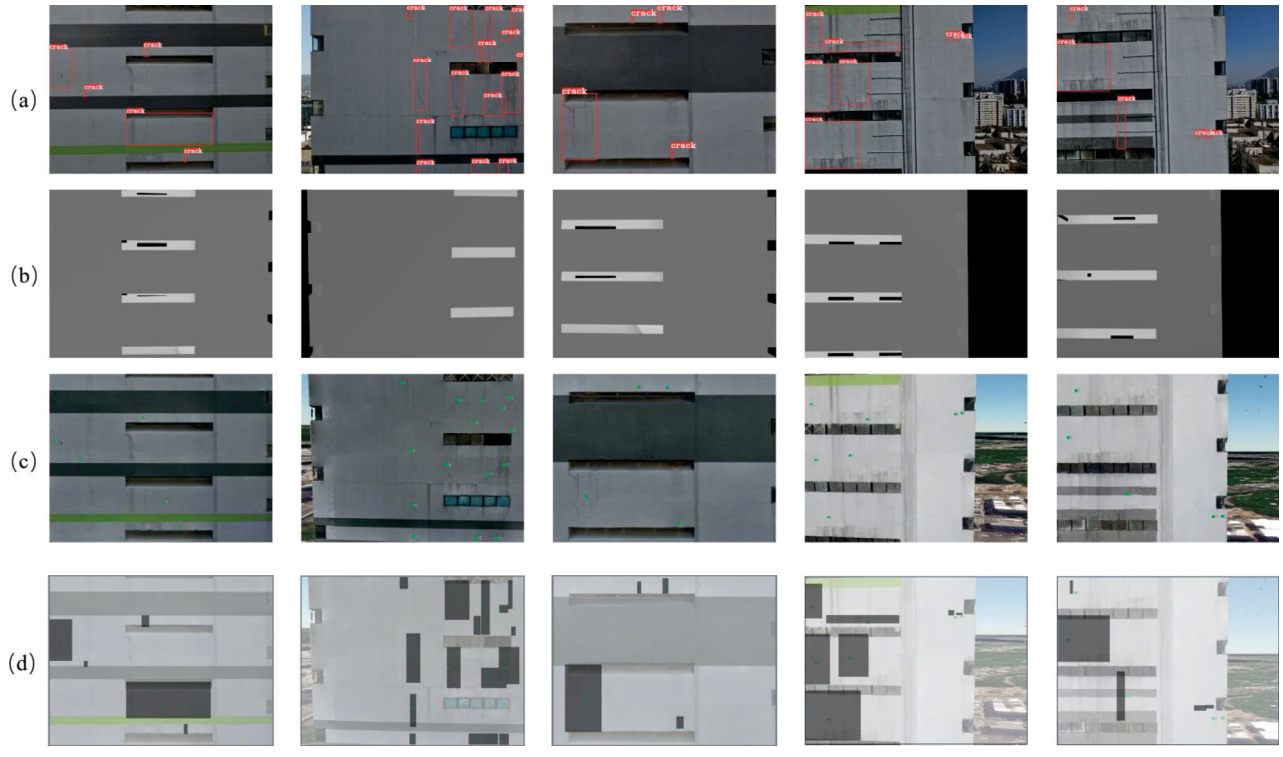


Fig. 12 The process of quantitative evaluation of the results of detection defect registration: (a) detection images with red bounding boxes to illustrate the defects; (b) GeoBIM derived depth image; (c) registration results of individual defects and each defect is illustrated by a green mark; (d) mask of defects for registration evaluation, while the masks are from (a)'s bounding box and defects are from (c)'s green marks

results achieved over 80% $mAP_{0.5}$ accuracy (at 30 FPS), demonstrating strong consistency within the dataset and the effectiveness of the chosen model for accurate detection. Figure 12(a) shows the defect detection outcomes from the aerial photographs, with red bounding boxes highlighting the defects. Figure 12(b) illustrates the depth maps derived from GeoBIM, providing essential spatial information for defect localization. Finally, panel (c) visualizes the registered defects on a WebGIS platform powered by Cesium (2023), where the central positions of the defects are marked by green points. From Figure 12(c), it is clear that the SfM model more accurately represents the as-is condition of the building compared to BIM. The SfM serves as the primary geometrical and visual representation within the DT framework, with the registered defects accurately reflecting their true geographic locations. These results confirm the effectiveness of our methodology. However, errors are inevitable during data processing, and we have developed a validation method to quantify the precision of our approach.

Considering that the defects have been located in global geographic coordinates, we calculate the defect localization error as the offset between the registered defect positions on GIS-derived images and the centers of the detected defect bounding boxes on original images. To achieve this, we have reconstructed each POV of the camera within the GIS virtual space, which precisely mirrors the actual world settings and incorporates identical geographical and optical features, as shown in Figure 12(c). Following this setting, we overlay the defect images onto these virtual images to accurately determine the defect localization errors. Here, the gray square masks represent the original defect bounding boxes, and the registered defects are marked with relative green points, as illustrated in Figure 12(d). These discrepancies are then converted from pixel measurements to physical units measured in centimeters. The accuracy of our registration method is evaluated by calculating the mean absolute error (MAE), the root mean square error (RMSE), and the interquartile range (IQR)—the latter being the difference between the first quartile (Q1) and the third quartile (Q3). The results, as tabulated in Table 2, confirm the centimeter-level accuracy of our approach. These statistical measures provide a robust assessment of our method's precision, ensuring that our registration technique is both reliable and suitable for practical applications in defect detection and localization.

To validate the performance of our proposed registration framework, we conducted a comparative evaluation against three representative classes of state-of-the-art methods: GPS-based projection, image-feature-based BIM registration, and GIS-based 2 D image alignment. As summarized in Table 3, these conventional approaches suffer from distinct limitations that hinder their effectiveness in practical settings. GPS-based projection methods, such as Tan et al. (2022), rely on UAV GPS data, which is prone to drift and environmental interference, often resulting in localization errors of 1–3 meters. This level of inaccuracy significantly compromises the spatial reliability of defect registration, particularly in dense urban environments. Image-feature-based approaches (Zhang et al. 2023a) attempt to match visual features between UAV imagery and BIM models but require distinctive surface patterns or geometrical features to function effectively. In practice, however, façades often lack such distinguishing characteristics, leading to failed or unstable registration. GIS-based 2D alignment methods (Chen et al. 2021) mitigate some spatial alignment issues by unfolding building façades into planar representations, but they disregard the 3D structure of the target surfaces. As a result, they cannot accommodate curved or complex façades and fail to provide accurate spatial and semantic correspondence. In contrast, our GeoBIM-based registration framework achieves centimeter-level accuracy by combining virtual camera rendering, pose correction, and semantic

Table 2 Defect registration error for large-scale infrastructure (computed over 1016 close-range façade images)

Registration error (cm)	Mean	MAE	RMSE	IQR
Horizontal	0.490	2.350	4.746	0
Vertical	0.592	1.037	2.385	0
Diagonal	1.360	4.056	7.149	3.747

Table 3 Benchmark comparison of representative defect registration methods

Reference	Approach type	Error level	Limitations
Tan et al. 2022	BIM+GPS pose	~1–3 m	High error due to GPS drift; flat façades
Zhang et al. 2023a	BIM+image-based registration	Failed	Fails on textureless surfaces
Chen et al. 2021	GIS+2D façade unfolding	Failed	Ignores 3D geometry and curved façades
Our method	GeoBIM+pose correction + semantic matching	**~1–5 cm**	Complex geometries; semantically enriched

matching. It is robust to visual ambiguity, overlapping images, and irregular geometries, demonstrating superior fidelity and applicability across diverse architectural conditions.

Our evaluation methodology is designed to rigorously assess the accuracy of defect localization and the fidelity of the geometric representations in our DT models. By systematically comparing the derived positions and conditions of structural defects against empirical measurements and DT-derived data, we can not only validate the effectiveness of our process but also identify areas for further refinement and enhancement. In fact, both theoretically and in practice, we have demonstrated that this method possesses commendable robustness and scalability, effectively addressing the limitations inherent in existing methodologies. Our approach facilitates the registration of irregular images, the exclusion of non-target areas, the merging of redundant defects, and the verification of model integrity. Details are listed as below.

1) Irregular defect image registration

In practical applications, UAV flight paths rarely align perfectly with the planned trajectories due to factors like localization errors, planning inaccuracies, and wind forces. This issue is especially prominent during manual flights, where aerial photographs are often captured at skewed angles relative to the building façades. Discarding these images would compromise the completeness of the data. Unlike existing methods that require the camera to be perpendicular to flat wall surfaces, our approach performs robustly even with skewed angles and on irregular wall surfaces. As shown in Figure 13, the defect registration on the corresponding GIS platform aligns well with the original detection images, accurately pinpointing defects at skewed and irregular positions.

To further evaluate the robustness of our method, we conducted additional experiments on structures with complex surface geometries, including both modern and historical buildings. Specifically, we tested on the China Resources Logistics Kader Centre in Hong Kong, a high-rise building featuring curved glass and metal façades, and the Fujian Tulou, a traditional circular earth building with prominent curved and inclined surfaces. These structures present significant challenges for defect registration due to their non-planar walls and low-texture surfaces. Our method successfully registered high-resolution defect images on both structures, even under non-orthogonal viewing conditions. Sample results from these experiments are shown in Figure 14, where green markers indicate the registered defect positions and red bounding boxes highlight the affected areas. The registration accuracy remained within a few centimeters, consistent with our previous evaluations on planar surfaces.

This confirms that our registration framework maintains high precision across diverse architectural forms and

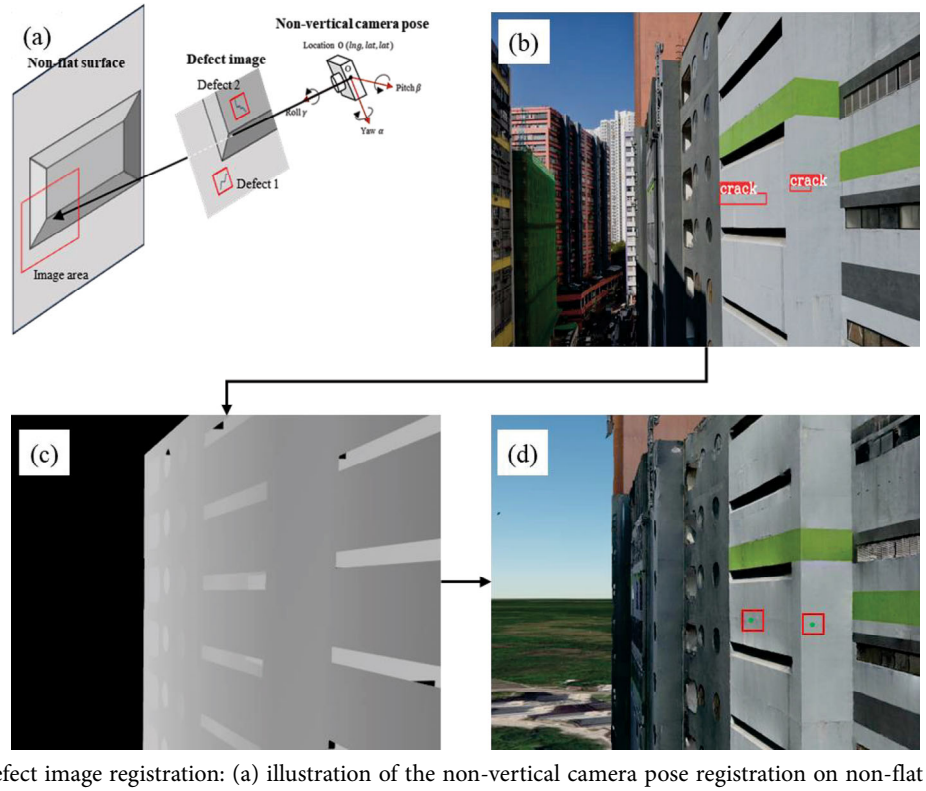


Fig. 13 Irregular defect image registration: (a) illustration of the non-vertical camera pose registration on non-flat surface; (b) original defect image; (c) GeoBIM-derived depth map; (d) registered defect image (defects are marked as green spots)

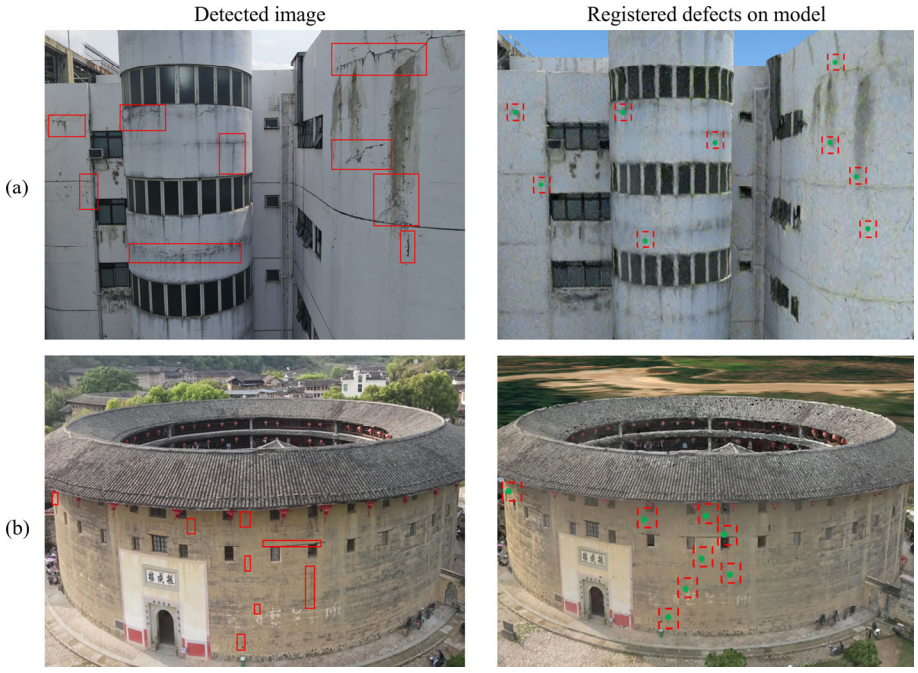


Fig. 14 Extended validation result on two scenarios: (a) modern high-rise building with curved façades; (b) complex historic architectural structure. The left column is the detected image with defects' bounding box and the right column is the corresponding registered defects on model

demonstrates its applicability to real-world scenarios with complex structural features. The ability to robustly align defect images on irregular surfaces significantly expands the potential use cases of our method in both modern infrastructure inspection and heritage conservation.

2) Exclusion of non-target areas

Due to the reasons mentioned above, aerial photographs captured by UAVs not only encompass the target building but may also include extraneous elements such as adjacent structures and trees. Utilizing GeoBIM-derived depth maps, which focus exclusively on the architectural structure itself, aids in the elimination of non-target areas within the images. As depicted in Figure 15(a), the presence of neighboring buildings can interfere with detection algorithms, leading to erroneous results. Depth maps, as shown in Figure 15(b), facilitate the direct generation of masked images (Figure 15(c)), where the target areas are denoted in white (value 1 in image) and non-target areas in black (value 0 in image). Consequently, the final image output (Figure 15(d)) is purged of non-target areas, thereby also removing incorrect detection outcomes and enhancing the precision of the data. This methodology significantly improves the quality of the analysis by focusing solely on relevant architectural details, thus optimizing the effectiveness of the detection process in urban and complex environments.

3) Redundant defect merging

Previous research indicates that UAV-based defect detection tasks often require a necessary overlap rate to ensure the completeness of data collection, which can result in the same defect appearing in multiple images, as illustrated in Figure 16(a). Our method uses high-precision global geographic registration to uniquely localize each defect, enabling the merging of duplicate detection results. This approach ensures the uniqueness of each defect by effectively consolidating overlapping detection results from adjacent images, as demonstrated in Figure 16(b). By implementing this strategy, we not only streamline the data but also enhance the accuracy of our defect mapping, ensuring that each defect is represented just once in the analysis. This reduction in redundancy significantly reduces data clutter and improves the efficiency of subsequent processing and analysis, leading to more reliable and actionable insights.

4) Verification of model integrity

The results of 3D reconstruction, specifically the 3D models of target buildings, often suffer from issues such as voids and distortions due to insufficient data completeness during collection. Typically, the evaluation of reconstruction methods is conducted on datasets, but such datasets for large-scale architectural scenes are exceedingly rare. Furthermore, generating ground truth for each target building (e.g., through comprehensive laser scanning) is cost-prohibitive and impractical in real-world applications. Therefore, developing effective evaluation methods for assessing the quality of model constructions is a critical need in

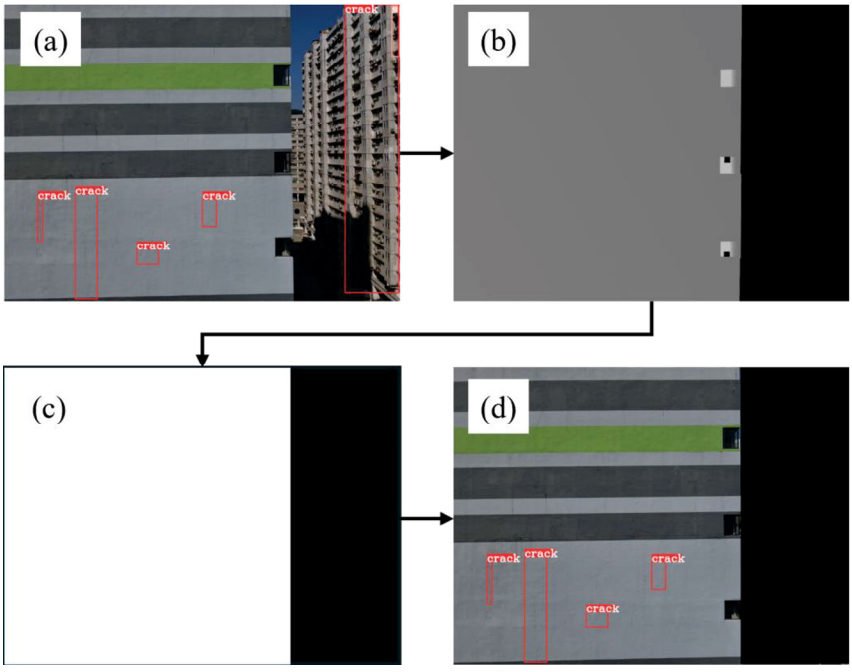


Fig. 15 Exclusion of non-target areas: (a) original defect image with non-target area defect (FP result); (b) GeoBIM-derived depth map; (c) mask image from depth map; (d) defect image with target area

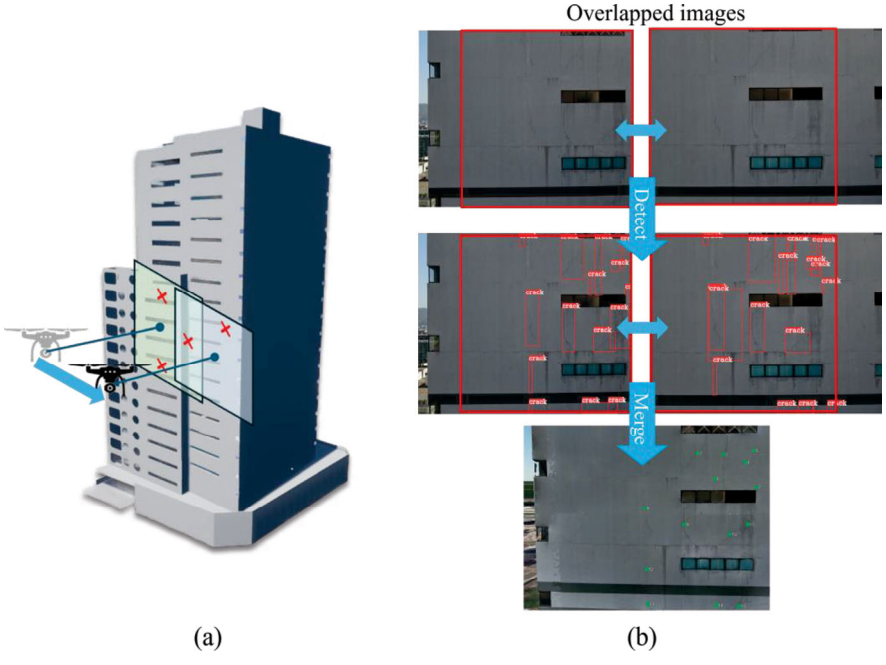


Fig. 16 Registration to merge redundant defects on overlapped images: (a) aerial photography with overlap ratio; (b) merging of redundant defects (overlapped area is marked by bounding box)

DT modeling. Our approach offers a feasible quantitative perspective to address this challenge. By comparing the reconstructed models with BIM-derived depth maps from identical POV, we assess the structural integrity of the corresponding constructions. As illustrated in Figure 17, by comparing depth images from GeoBIM with corresponding SfM model image, structure defects in the modeling process can be precisely localized. Structural differences between images are quantified using the structural similarity index measure (SSIM), which is 81.9% for the given sample. The formula of SSIM for image i and j is shown below:

$$\mathrm{SSIM}(i,j) = \frac{(2\mu_i\mu_j + C_1) + (2\sigma_{ij} + C_2)}{(\mu_i^2 + \mu_j^2 + C_1)(\sigma_i^2 + \sigma_j^2 + C_2)} \tag{12}$$

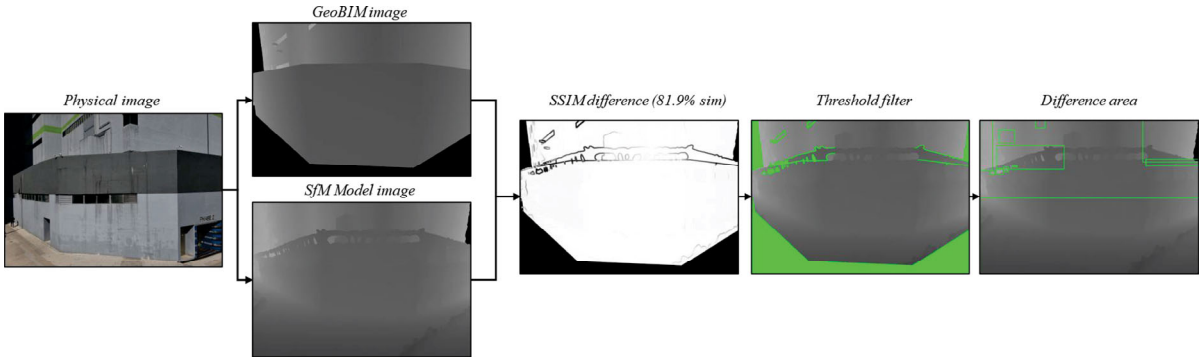


Fig. 17 SSIM compare for verification of model integrity

where μ_i and μ_j are the pixel sample mean; σ_{ij} is the covariance of i and j; σ_i^2 and σ_j^2 are respectively the covariance of each image; C_1 and C_2 are variables to stabilize the division with weak denominator.

Subsequent differences that exceed a predefined threshold are filtered to delineate predictive bounding boxes, thereby identifying specific defect areas. This comparison allows for the quantitative identification of structural anomalies, such as voids or boundary distortions, at specific locations. Such assessments are instrumental in guiding further data capture and model updates, thereby enhancing the accuracy and utility of the 3D reconstruction process. This method not only improves the fidelity of architectural models but also supports the iterative refinement and updating of DTs, ensuring their applicability and reliability in practical scenarios.

4.1.2 *Results of GeoBIM retrieval*

Using the aforementioned approach, GeoBIM has successfully extracted all structural semantic information from the BIM system and completed geographic registration. As illustrated in Figure 18(a), detailed information about each architectural element is accessible. To automate the retrieval of structural information corresponding to each defect, we utilized the script in Figure 18(b) to acquire the geo-position and geometric boundary data of all structures, comparing these with the locations of defects. Figure 18(c) presents an image of a specific defect, while Figure 18(d) shows the structural information corresponding to that defect retrieved via GeoBIM. This method efficiently links each defect with its respective structural location, thereby providing a detailed depiction of the defect distribution within the building structure.

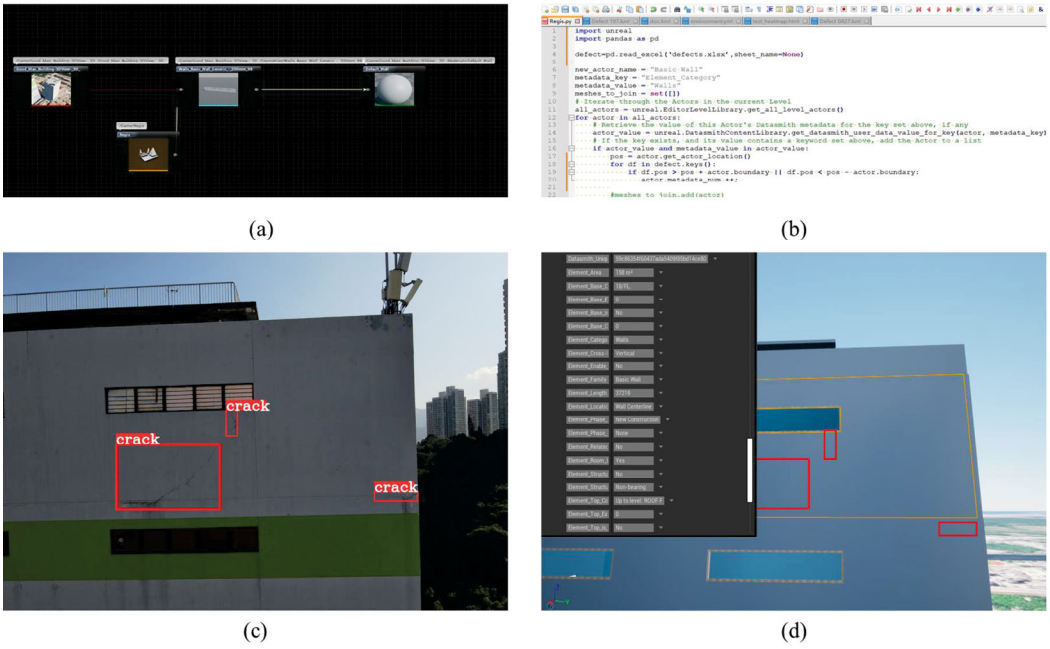


Fig. 18 Results of GeoBIM retrieval: (a) GeoBIM structure element as scene actor; (b) scripts for GeoBIM retrieval; (c) defect images with registered defects; (d) GeoBIM retrieval result for structure element matching

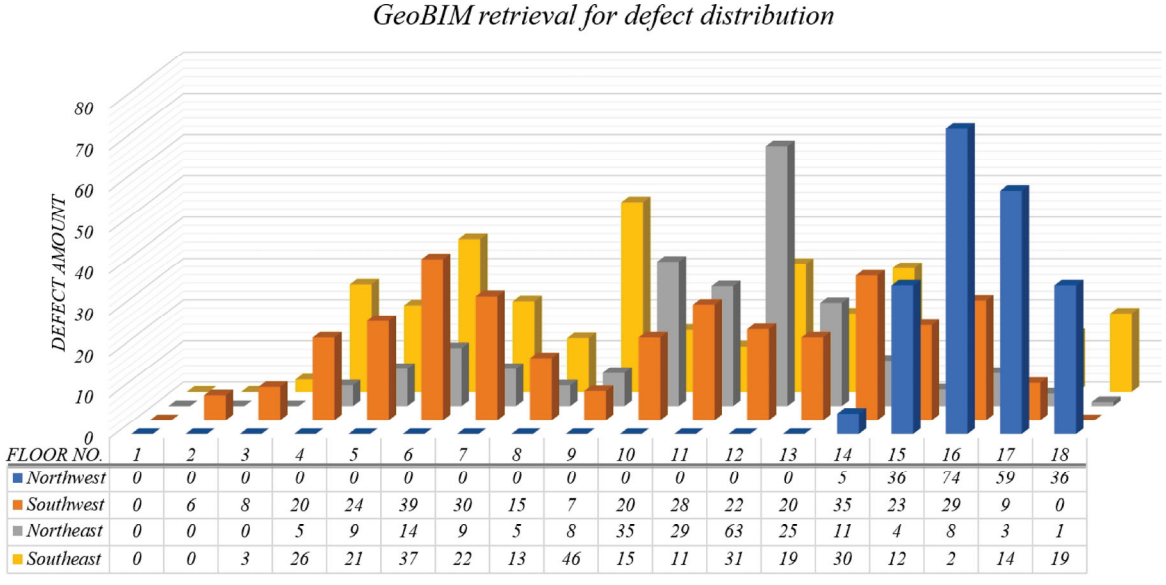
Particularly, since defect detection primarily focuses on the façade of concrete structures, the final data display the distribution of defects across each floor as in Figure 19. Each direction represents a different wall surface; for instance, the northwest-facing wall, due to visual obstructions, only includes defect data from the upper floors, with no lower floor defects included. This structure-oriented distribution of defects supports systematic assessments of structural damage in buildings, guiding efficient and targeted maintenance strategies. Moreover, the precision of defect localization is crucial for ensuring that the detected defects are accurately registered onto the building façade in the GeoBIM environment. As shown in Table 2, the pose estimation and depth alignment processes achieve centimeter-level accuracy, which ensures that the defect positions are geo-referenced with a high degree of precision. The robust global geographic registration methodology, incorporating GCPs, further minimizes errors in the overall spatial alignment of the building model. Given the high accuracy achieved in defect localization, the distribution of defects across floors and surfaces offers a reliable and precise representation of the building's condition. This precision allows for clear visualization of defect patterns and concentrations, facilitating the development of systematic, floor-by-floor maintenance strategies. The centimeter-level accuracy ensures that defects are mapped accurately to their real-world locations, enabling maintenance teams to address the most critical areas efficiently and without ambiguity. Consequently, the precision of the system is sufficiently high to support both macro-level damage assessments and micro-level defect management, ensuring that the data can effectively inform targeted repair strategies.

4.1.3 *Decision support through GeoBIM*

In the realm of building maintenance, the use of building maintenance units (BMUs) plays a pivotal role in executing repair operations. To enhance operational efficiency, we have introduced an algorithm that optimizes maintenance trajectories based on the building defect DT model illustrated in Figure 20(a). This approach not only streamlines repair activities but also facilitates strategic planning through precise defect localization and distribution analysis.

The user interface (UI) depicted in Figure 20(b) provides clear and actionable guidance for both on-site engineers and management staff. This tool enables efficient planning and oversight of maintenance operations, ensuring that defect rectification is both systematic and targeted. The BMU in operation is shown in Figure 20(c), while engineers use GPS-enabled mobile devices (Figure 20(d)) to acquire field data and monitor the progress of maintenance tasks in real time.

The UI depicted in Figure 20(b) provides clear and actionable guidance for both on-site engineers and management staff. This tool enables efficient planning and oversight of maintenance operations, ensuring that defect rectification is both systematic and targeted. The BMU in operation is shown in Figure 20(c), while engineers use GPS-enabled mobile devices (Figure 20(d)) to acquire field data and monitor the progress of maintenance tasks in real time. To enhance spatial awareness and facilitate intuitive interaction, the DT interface is developed on the Cesium WebGIS platform (Cesium 2023), enabling real-time 3D visualization of the entire structure and its registered defects. Each green marker on the 3D model represents an individual defect location, serving not only as a visual



FLOOR NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Northwest	0	0	0	0	0	0	0	0	0	0	0	0	0	5	36	74	59	36
Southwest	0	6	8	20	24	39	30	15	7	20	28	22	20	35	23	29	9	0
Northeast	0	0	0	5	9	14	9	5	8	35	29	63	25	11	4	8	3	1
Southeast	0	0	3	26	21	37	22	13	46	15	11	31	19	30	12	2	14	19

Fig. 19 Defect distribution across four façades (indicated by direction) on 18 floors from the GeoBIM retrieval

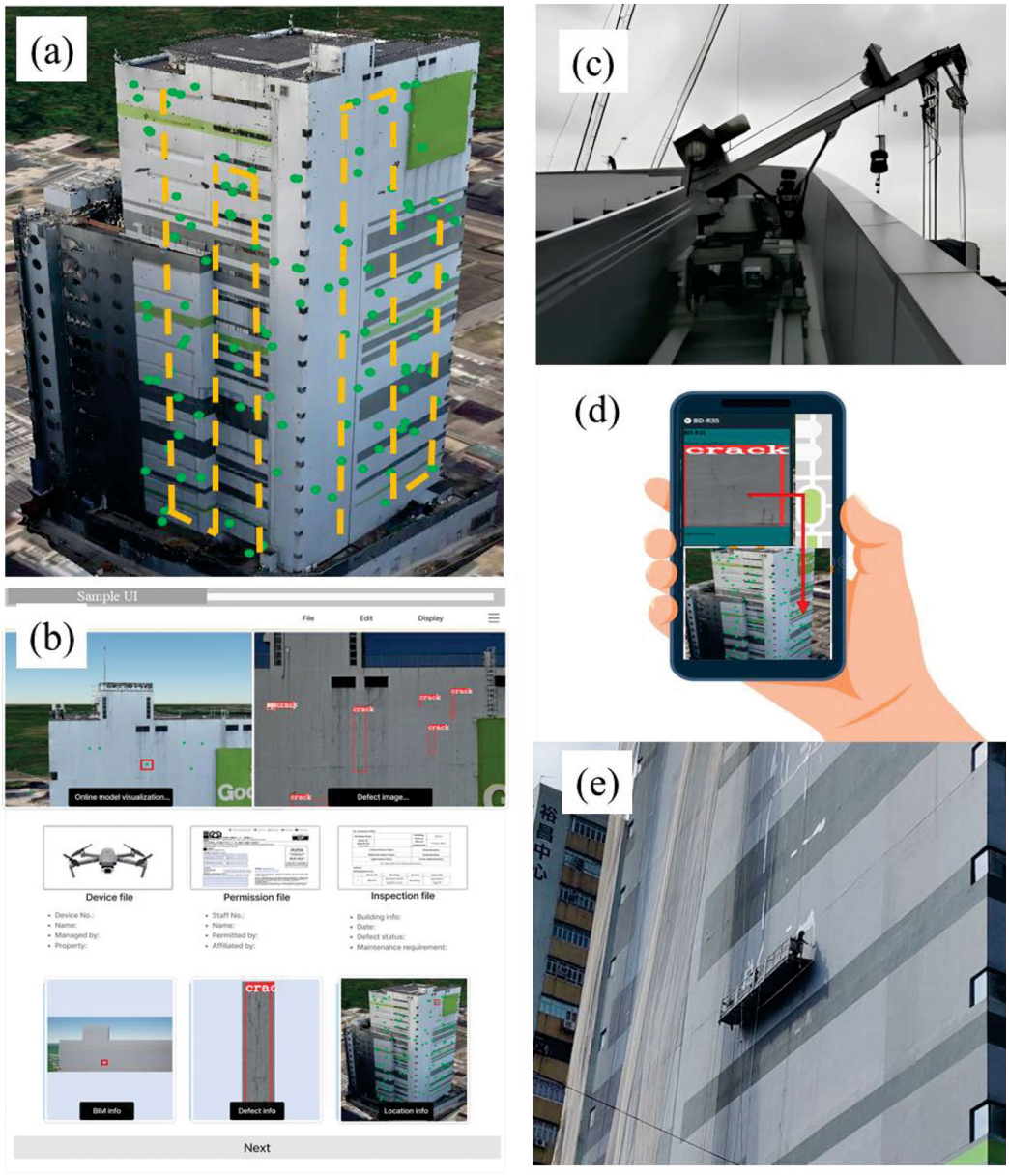


Fig. 20 Target building maintenance activities guided by this method: (a) efficient maintenance path; (b) data assignment into the web UI; (c) Equipped building maintenance unit; (d) GPS-supported mobile device for defect information access; (e) onsite maintenance activity

annotation but also as an interactive gateway. When clicked, these markers dynamically trigger the display of associated defect imagery, the semantic information of the underlying BIM component, and relevant metadata such as detection timestamp and severity assessment. This bidirectional linkage between the digital model and defect records supports both top-down review by management and bottom-up reporting by field staff. For example, engineers can upload updated inspection images or confirm repair completions directly through the interface.

Leveraging the GeoBIM framework, structural defects are accurately mapped to their corresponding building elements with high spatial precision. This mapping enables a detailed visualization of defect distributions, providing actionable insights for prioritizing maintenance activities. Figure 19 illustrates the systematic localization of defects across façades, categorized by floor, allowing maintenance teams to focus on high-priority areas with greater defect density.

The proposed method further supports the optimization of maintenance paths. By analyzing the defect distribution data, our algorithm determines the most efficient trajectories for BMUs, minimizing resource use and downtime while addressing critical defects. The integration of defect data into the interactive UI ensures that real-time guidance is available to maintenance teams, enhancing their responsiveness and operational efficiency.

5 Discussion

In this section, we discuss the advantages of the proposed method, examine its current limitations, and suggest future improvements and extensions for broader application.

5.1 Method advantages

The proposed UAV–GeoBIM inspection framework offers several notable advantages over conventional approaches. First, it achieves high precision in defect localization. Our experiments demonstrated that defects can be georeferenced with an accuracy on the order of centimeters, which is a significant improvement compared to traditional manual inspections. This level of accuracy ensures that the mapped defect positions correspond very closely to their real-world locations, facilitating reliable condition assessment and repair planning. Second, the method is robust in handling various practical challenges in data collection. It can accommodate UAV imagery captured at oblique angles or of irregular façade geometries, whereas many existing techniques require strictly perpendicular camera views or flat surfaces. In our case, even images taken from suboptimal angles were used effectively without loss of localization performance. In addition, by using depth-derived masks, the framework automatically filters out irrelevant regions (such as neighboring buildings or sky in the background), thereby avoiding false positives on non-target surfaces. The pipeline also identifies and merges duplicate detections of the same defect from overlapping images, ensuring that each unique defect is recorded only once. These features collectively improve the reliability of the defect data: the results are cleaner (focused only on the actual building) and more concise (with no duplicated entries), which simplifies subsequent analysis. Third, our approach integrates multi-source data including photogrammetric imagery, GIS, and BIM, in an automated end-to-end workflow, leading to a high degree of automation and information richness. The GeoBIM-based registration efficiently unifies the coordinate systems of the SfM model, BIM, and global GIS reference, streamlining what is often a complicated transformation process in other workflows. This integration yields a DT that not only contains the geometric representation of the building but also the semantic context for each defect through the BIM metadata. Every defect is linked to a specific building component and floor level, providing context that purely image-based methods lack. The system is scalable and was shown to work on a large 18-story building using over a thousand images, indicating its potential applicability to even larger structures or campus-scale deployments. The use of multiple UAVs and a high-performance computing setup demonstrates that the framework can handle extensive data collection and processing in a time-efficient manner. Moreover, the outcome of our pipeline is directly usable for maintenance decision-making: by visualizing defects in a WebGIS environment and retrieving their structural information, facility managers can immediately interpret the results in terms of actionable tasks.

Finally, the framework bridges the gap between static inspections and active maintenance management. Unlike conventional defect detection studies that end at reporting the locations of defects, our approach goes a step further by incorporating a decision-support mechanism. The integration of defect data with maintenance planning tools with the BMU path optimization and UI shows how the DT can guide real-world interventions. This synergy between accurate digital models and practical maintenance workflows is a key advantage for asset management: it enables data-driven scheduling of repairs, efficient allocation of resources, and continuous updating of the building's condition in the DT. In summary, the method enhances accuracy, robustness, and automation in defect detection, and it translates those improvements into tangible benefits for building maintenance operations.

5.2 Limitations

While our proposed method demonstrates strong performance and generalizability, we acknowledge several areas where further improvements or refinements may be warranted.

Firstly, although the method has been validated on a variety of building types—including high-rise curtain wall systems, curved façades, and historic masonry—it is important to recognise that real-world inspection scenarios can be even more diverse and complex. Architectural configurations with extreme occlusions, highly reflective surfaces, or dense ornamental detail may challenge the robustness of both the defect detection algorithm and the GeoBIM registration. In such scenarios, localization accuracy may decrease and processing time may increase due to greater algorithmic complexity. These effects are not intrinsic limitations of the method, but rather expected performance variations under non-ideal conditions. Addressing them may require additional data pre-processing, model retraining, or adaptive flight strategies in future deployments.

Secondly, the current system operates in a near real-time mode. UAV images are collected, processed, and visualised within a short period following data acquisition, which suffices for periodic condition assessment and maintenance planning. However, the system does not yet incorporate a continuous real-time streaming capability. This is not a

technical shortcoming but a design choice, as the current application scenarios do not demand persistent monitoring. Should use cases arise—such as disaster response or critical infrastructure surveillance—that require real-time data feeds, the system architecture is sufficiently modular to support such integration with minimal modification.

Finally, while the use of SfM/MVS-based reconstruction and BIM-GIS data fusion delivers rich spatial and semantic context, it does involve significant data volume and computational resources. In our implementation, these demands were met using modern consumer-grade hardware. However, in extremely large-scale projects or resource-constrained environments, computational efficiency and memory management may become more prominent concerns. Additionally, although data format interoperability between BIM and GIS platforms has been effectively handled using standardised coordinate transformations and conversion pipelines, residual inconsistencies in geometry resolution or metadata mapping may occasionally require manual refinement.

In summary, these issues are better characterised as manageable trade-offs or conditions for optimal performance, rather than intrinsic weaknesses of the framework. They highlight practical considerations that can inform targeted enhancements in future iterations of the system.

5.3 Future directions and application extensions

We outline several directions for future work to address the above limitations and extend the applicability of the proposed framework.

(1) Broadening applicability to diverse structures: Ensuring that the GeoBIM integration works with different BIM standards and practices (for example, varying levels of detail in BIM models or different coordinate reference systems) will make the system more universally applicable. In the long term, scaling the methodology to manage multiple buildings or an entire portfolio of assets (such as all buildings in a campus or all bridges in a city) could enable a more holistic infrastructure maintenance platform. This might involve integrating our building-level DTs into larger urban DT or smart city systems, allowing city officials and stakeholders to monitor and prioritize maintenance across many assets in a unified environment.

(2) Integration of real-time data and IoT: Another key improvement is incorporating real-time monitoring capabilities into the DT. By integrating Internet of Things (IoT) sensors and devices with the GeoBIM framework, the digital model could be kept up-to-date with live data. Real-time data integration would enable timely alerts for new or worsening defects and could support predictive maintenance – identifying areas of concern before visible defects even emerge. Additionally, future work could explore automated UAV deployments or permanent camera installations for more frequent data capture.

(3) Enhanced semantic analysis and usability: Increasing the semantic understanding and user-friendliness of the DT is another future direction. On the one hand, this involves improving how defects are characterized and reported. Such enriched information would make the DT more valuable to engineers and decision-makers. On the other hand, integrating the framework with existing facility management or maintenance scheduling software could streamline the workflow from detection to repair. This might include exporting defect data in standardized formats or developing dashboards that allow users to interact with the defect information intuitively.

6 Conclusion

In conclusion, the method introduced in this study demonstrates remarkable scalability and holds substantial practical implications for the automated construction of architectural defect DTs and for guiding real-world maintenance endeavors. By utilizing high-precision, 3D global defect localization through GeoBIM registration and incorporating automated structural adaptation, this approach effectively resolves prevalent issues encountered in existing methods. Such issues include the limited scope of UAV data collection and the challenges in applying conventional techniques to all facets of a building's exterior. Our methodology significantly improves upon these limitations by facilitating precise defect mapping across the entire structure.

This novel end-to-end solution leverages the integration of BIM+GIS, not only to enhance the accuracy of defect localization but also to enable the solution's application on a urban scale for holistic management. By adopting this comprehensive approach, the methodology is capable of executing global control over extensive urban infrastructure, thereby paving the way for smart city management.

The implementation of this method allows for a sophisticated synergy between virtual models and their physical counterparts. This synergy is pivotal in enriching the DT with detailed semantic information, which in turn, refines the maintenance strategies and actions taken on the ground. In essence, it transcends the digital-physical division and aligns the DT paradigm with operational reality.

Validated in the dense urban environment of Hong Kong on a high-rise civil structure, our solution has proven

its feasibility, effectiveness, and efficiency. It stands as a testament to the potential of similar large-scale assets, ushering in a new era for architectural maintenance and asset management. By fostering an environment where defects are not merely identified but are contextually understood and addressed, the proposed solution offers a significant leap forward from current practices. It marks a pivotal step towards more resilient and maintainable urban architectural landscapes.

Acknowledgements

This work was supported in part by the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government via the Hong Kong Centre for Logistics Robotics and in part by the Research Grants Council of Hong Kong SAR (Grant Nos: 14209020, 14206821, 14209424, 14200524).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval

This study does not contain any studies with human or animal subjects performed by any of the authors.

Author contribution statement

Jihan Zhang, Benyun Zhao, Guidong Yang, Xunkuai Zhou, Yijun Huang, and Chuanxiang Gao contributed to the investigation, formal analysis, and original draft preparation. Jihan Zhang and Xi Chen contributed to the conceptualisation of the study. Xi Chen and Ben M. Chen provided resources, supervision, and were responsible for project administration. Xi Chen and Ben M. Chen also contributed to reviewing and editing the manuscript. Ben M. Chen acquired funding for the study. All authors read and approved the final manuscript.

Open Access: This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

References

Abouelaziz I, Jouane Y (2024). Photogrammetry and deep learning for energy production prediction and building-integrated photovoltaics decarbonization. *Building Simulation*, 17: 189–205.

Agnisarman S, Lopes S, Chalil Madathil K, et al. (2019). A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection. *Automation in Construction*, 97: 52–76.

Alencastro J, Fuertes A, de Wilde P (2018). The relationship between quality defects and the thermal performance of buildings. *Renewable and Sustainable Energy Reviews*, 81: 883–894.

Artus M, Alabassy M, Koch C (2021). IFC based framework for generating, modeling and visualizing spalling defect geometries. In: Proceedings of 28th International Workshop on Intelligent Computing in Engineering (EG-ICE 2021), Berlin, Germany.

Cesium (2023). The Platform for 3D Geospatial. Available at https://cesium.com/

Chen J, Liu D, Li S, et al. (2019). Registering georeferenced photos to a building information model to extract structures of interest. *Advanced Engineering Informatics*, 42: 100937.

Chen K, Reichard G, Akanmu A, et al. (2021). Geo-registering UAV-captured close-range images to GIS-based spatial model for building façade inspections. *Automation in Construction*, 122: 103503.

Chen J, Lu W, Fu Y, et al. (2023a). Automated facility inspection using robotics and BIM: a knowledge-driven approach. *Advanced Engineering Informatics*, 55: 101838.

Chen J, Lu W, Lou J (2023b). Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model. *Computer-Aided Civil and Infrastructure Engineering*, 38: 1079–1098.

Chen K, Reichard G, Xu X, et al. (2023c). GIS-based information system for automated building façade assessment based on unmanned aerial vehicles and artificial intelligence. *Journal of Architectural Engineering*, 29: 04023032.

Chen S, Fan G, Li J (2023d). Improving completeness and accuracy of 3D point clouds by using deep learning for applications of digital twins to civil structures. *Advanced Engineering Informatics*, 58: 102196.

Chen X, Pan Y, Gan VJL, et al. (2024). 3D reconstruction of semantic-rich digital twins for ACMV monitoring and anomaly detection *via* scan-to-BIM and time-series data integration. *Developments in the Built Environment*, 19: 100503.

De Filippo M, Asadiabadi S, Kuang JS, et al. (2023). AI-powered inspections of facades in reinforced concrete buildings. *HKIE Transactions*, 30: 1–14.

De Vries J (2015) Learn OpenGL.

Duque L, Seo J, Wacker J (2018). Synthesis of unmanned aerial vehicle applications for infrastructures. *Journal of Performance of Constructed Facilities*, 32: 04018046.

Ge Z, Liu S, Wang F, et al. (2021). Yolox: Exceeding yolo series in 2021. arXiv:210708430.

Hajji R, Oulidi HJ (2022). Building Information Modeling for a Smart And Sustainable Urban Space. New York: John Wiley & Sons

Hensel S, Goebbels S, Kada M (2019). Facade reconstruction for textured LOD2 citygml models based on deep learning and mixed integer linear programming. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42W5: 37–44.

Hosamo HH, Hosamo MH (2022). Digital twin technology for bridge maintenance using 3D laser scanning: A review. *Advances in Civil Engineering*, 2022: 2194949.

Hu X, Zhou Y, Vanhullebusch S, et al. (2022). Smart building demolition and waste management frame with image-to-BIM. *Journal of Building Engineering*, 49: 104058.

Huang H, Michelini M, Schmitz M, et al. (2020). LOD3 building reconstruction from multi-source images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43B2: 427–434.

Huang X, Wang X, Lv W, et al. (2021). Pp-yolov2: A practical object detector. arXiv:210410419.

Jati DGP (2021). Uav-based photogrammetry data transformation as a building inspection tool: applicability in mid-high-rise building. *Jurnal Teknik Sipil*, 16: 113–121.

Jocher G (2020). YOLOv5 by Ultralytics. https://doi.org/10.5281/zenodo.3908559

Jocher G, Chaurasia A, Qiu J (2023). Ultralytics yolov8.

Li D, Xie Q, Gong X, et al. (2021). Automatic defect detection of metro tunnel surfaces using a vision-based inspection system. *Advanced Engineering Informatics*, 47: 101206.

Li C, Li L, Jiang H, et al. (2022). Yolov6: A single-stage object detection framework for industrial applications. arXiv:220902976.

Li Q, Yang G, Gao C, et al. (2024). Single drone-based 3D reconstruction approach to improve public engagement in conservation of heritage buildings: A case of Hakka Tulou. *Journal of Building Engineering*, 87: 108954.

Liu X, Wang X, Wright G, et al. (2017). A state-of-the-art review on the integration of building information modeling (BIM) and geographic information system (GIS). *ISPRS International Journal of Geo-Information*, 6: 53.

Liu D, Chen J, Hu D, et al. (2019). Dynamic BIM-augmented UAV safety inspection for water diversion project. *Computers in Industry*, 108: 163–177.

Liu D, Wang G, Feng B, et al. (2025). Research on non-uniform heat transfer testing of prefabricated walls based on infrared images. *Building Simulation*, 18: 499–513.

Long X, Deng K, Wang G, et al. (2020). Pp-yolo: An effective and efficient implementation of object detector. arXiv:200712099.

McLaughlin E, Charron N, Narasimhan S (2020). Automated defect quantification in concrete bridges using robotics and deep learning. *Journal of Computing in Civil Engineering*, 34: 04020029.

Mehta S, Rastegari M (2021). Mobilevit: Light-weight, general-purpose, and mobilefriendly vision transformer. arXiv:211002178.

Mohammadi M, Rashidi M, Yu Y, et al. (2023). Integration of TLS-derived Bridge Information Modeling (BrIM) with a Decision Support System (DSS) for digital twinning and asset management of bridge infrastructures. *Computers in Industry*, 147: 103881.

Moretti N, Ellul C, Re Cecconi F, et al. (2021). GeoBIM for built environment condition assessment supporting asset management decision making. *Automation in Construction*, 130: 103859.

Musella C, Serra M, Menna C, et al. (2021). Building information modeling and artificial intelligence: Advanced technologies for the digitalisation of seismic damage in existing buildings. *Structural Concrete*, 22: 2761–2774.

Pantoja-Rosero BG, Achanta R, Beyer K (2023). Damage-augmented digital twins towards the automated inspection of buildings. *Automation in Construction*, 150: 104842.

Rakha T, Gorodetsky A (2018). Review of Unmanned Aerial System (UAS) applications in the built environment: Towards automated building inspection procedures using drones. *Automation in Construction*, 93: 252–264.

Ren S, He K, Girshick R, et al. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39: 1137–1149.

Ruiz RDB, Lordsleem AC Jr, Rocha JHA, et al. (2022). Unmanned aerial vehicles (UAV) as a tool for visual inspection of building facades in AEC+FM industry. *Construction Innovation*, 22: 1155–1170.

Sanders A (2016). *An Introduction to Unreal Engine 4*. A K Peters/CRC Press.

See JE, Drury CG, Speed A, et al. (2017). The role of visual inspection in the 21st century. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61: 262–266.

Spencer BF, Hoskere V, Narazaki Y (2019). Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5: 199–222.

Tan Y, Li G, Cai R, et al. (2022). Mapping and modelling defect data from UAV captured images to BIM for building external wall inspection. *Automation in Construction*, 139: 104284.

Valero E, Bosché F, Forster A (2018). Automatic segmentation of 3D point clouds of rubble masonry walls, and its application to building surveying, repair and maintenance. *Automation in Construction*, 96: 29–39.

Wang C, Bochkovskiy A, Liao H (2022) Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv:220702696

Wang M, Xu W, Cao G, et al. (2024). Identification of rural courtyards' utilization status using deep learning and machine learning methods on unmanned aerial vehicle images in North China. *Building Simulation*, 17: 799–818.

Wang T, Gan VJL (2024). Multi-view stereo for weakly textured indoor 3D reconstruction. *Computer-Aided Civil and Infrastructure Engineering*, 39: 1469–1489.

Xia H, Liu Z, Efremochkina M, et al. (2022). Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and

building information modeling integration. *Sustainable Cities and Society*, 84: 104009.

Xu S, Wang X, Lv W, et al. (2022). Pp-yoloe: An evolved version of yolo. arXiv:220316250

Yang B, Lv Z, Wang F (2022). Digital twins for intelligent green buildings. *Buildings*, 12: 856.

Yang G, Zhou X, Gao C, et al. (2023). Multi-view stereo with learnable cost metric. In: Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA.

Yu R, Li P, Shan J, et al. (2022). Structural state estimation of earthquake-damaged building structures by using UAV photogrammetry and point cloud segmentation. *Measurement*, 202: 111858.

Zhang C, Shu J, Shao Y, et al. (2022). Automated generation of FE models of cracked RC beams based on 3D point clouds and 2D images. *Journal of Civil Structural Health Monitoring*, 12: 29–46.

Zhang C, Wang F, Zou Y, et al. (2023a). Automated UAV image-to-BIM registration for building façade inspection using improved generalised Hough transform. *Automation in Construction*, 153: 104957.

Zhang C, Zou Y, Dimyadi J, et al. (2023b). Thermal-textured BIM generation for building energy audit with UAV image fusion and histogram-based enhancement. *Energy and Buildings*, 301: 113710.

Zhang D, Yu X, Yang L, et al. (2023c). Data-augmented deep learning models for abnormal road manhole cover detection. *Sensors*, 23: 2676.

Zhao Y, Lv W, Xu S, et al. (2023). Detrs beat yolos on real-time object detection. arXiv:2304.08069.

Zhao B, Zhou X, Yang G, et al. (2024). High-resolution infrastructure defect detection dataset sourced by unmanned systems and validated with deep learning. *Automation in Construction*, 163: 105405.

Zheng M, Lei Z, Zhang K (2020). Intelligent detection of building cracks based on deep learning. *Image and Vision Computing*, 103: 103987.

Zheng S, Hao F, Lu Y, et al. (2025). A method for quantitatively evaluating the impact of defects on wall U-value using infrared thermal imaging. *Building Simulation*, 18: 281–293.