Drone-Based Façade Defect Detection via Decision-Level Fusion of Thermal Infrared and Visible Images

Jiaqi Li[1], Qingrui Yue[1,2], Zezhi Ding[1], Haofeng Yan[3], Nan Jin[2], Xincong Yang[1,*]

[1]School of Intelligent Civil and Ocean Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen, 518055, China
[2]Safe Urban Development Institute of Science and Technology (Shenzhen), Shenzhen, 518024, China
[3]School of Civil Engineering, Harbin Institute of Technology, Harbin, 150090, China

Abstract -

Urban façade deterioration, especially tile delamination and related surface damage, poses significant safety risks in dense urban environments. Conventional inspections based on manual visual assessment and acoustic sounding are labor-intensive, hazardous, and subjective. This paper presents an autonomous UAV-based façade inspection framework that combines oblique-photogrammetry-driven path planning with decision-level fusion of infrared and visible imagery for defect identification. A 3D model of the target building is reconstructed to extract geometric features for generating safe flight trajectories at a fixed standoff distance of 10 m, supporting near-complete and spatially consistent dual-modal data acquisition under the planned inspection route. Modality-specific defects are then segmented from RGB and thermal data using deep-learning models, and the fused defect map is produced through decision-level fusion of surface appearance cues and subsurface thermal signatures. A field deployment on a building demonstrates the feasibility and efficiency of the framework, completing the inspection of approximately 4,200 m^2 of façade area using 208 waypoints within 50 minutes of total on-site operation. The results show that the proposed method provides a practical and safer solution for façade condition assessment and maintenance prioritization.

Keywords -

Façade inspection, Drone path planning, Infrared thermography, Decision-level fusion, Subsurface delamination

1 Introduction

Building façades function as the first line of defense for infrastructure, enduring continuous environmental stress throughout their lifecycle. Synergistic effects of material deterioration, construction imperfections, and severe weather can compromise façade integrity, serviceability, and public safety. In China, statistical analysis of façade failures reveals that tile detachment is a predominant hazard, frequently resulting in casualties and economic losses [1]. The risk is amplified in dense urban environments where buildings directly face crowded streets. Consequently, implementing regular and rigorous inspection protocols for aging tiled façades is imperative for public safety maintenance.

Conventionally, façade health monitoring relies on manual visual inspection and acoustic methods. Ground-based visual surveys, often aided by optical telescopes or handheld thermography, attempt to identify external anomalies. However, these methods suffer from low efficiency and are prone to human error due to the inspectors' subjective judgment [2]. Alternatively, hammer sounding—which detects subsurface delaminations via acoustic response—is widely practiced but inherently inefficient. It requires direct physical access via scaffolds or gondolas, exposing personnel to significant safety hazards associated with working at heights.

Recent strides in unmanned aerial vehicle (UAV) technology have significantly improved infrastructure monitoring [3,4]. Equipped with high-resolution sensors, UAVs offer a promising alternative for accessing hard-to-reach vertical surfaces. Despite this potential, a robust methodology for reliably detecting subsurface delaminations through automated means remains underdeveloped. Addressing this gap, this study proposes a novel drone-based inspection framework that integrates infrared thermography and RGB imagery to enhance the precision and efficiency of façade defect detection.

To address these challenges, this paper presents an autonomous inspection framework that integrates adaptive flight path planning with decision-level fusion of thermal infrared and visible images for façade assessment. The proposed method first leverages oblique photogrammetry to reconstruct a 3D model of the target building, which supports the generation of safe and feasible flight trajectories for data acquisition. Subsequently, synchronized visible light images and thermal infrared data are analyzed using a deep learning-based segmentation pipeline. A decision-level fusion strategy is then applied to combine surface appearance cues with subsurface thermal anomalies, thereby reducing false detections. The feasibility of the framework is validated through a field deployment at a large commercial building, demonstrating efficient on-site operation and reliable defect detection compared with conventional inspection practices.

2 Methodology

2.1 Overview of the Proposed Framework

Figure 1 presents the schematic workflow of the proposed drone-based inspection system, which integrates autonomous path planning with dual-modal defect recognition. The methodology follows a coarse-to-fine strategy. Initially, oblique aerial imagery is acquired to construct a photogrammetric 3D model of the building. This geometric model serves as the environmental baseline for planning a flight path tailored for close-range inspection. During the inspection phase, the drone autonomously captures paired thermal infrared (IR) and visible light (RGB) imagery along the planned route. These images are processed by independent deep learning models to detect surface anomalies (e.g., peeling, cracks) and subsurface delamination. To maximize detection reliability, a decision-level fusion strategy is applied, synthesizing the outputs from both sensors into a unified and more reliable façade health assessment.

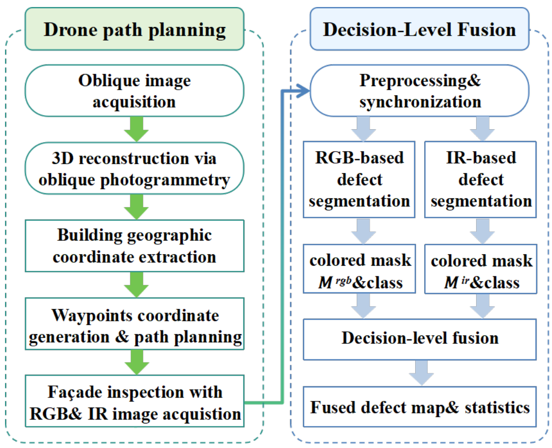


Figure 1. The overall framework of the proposed drone-based façade inspection system

2.2 Drone Path Planning Method

Drones equipped with flight control systems can execute complex, predefined trajectories [5]. However, manual path planning for complex building façades is inefficient and prone to collision risks. This study proposes an automated method for generating optimized inspection routes based on the extraction of architectural envelope key points from a preliminary 3D model.

2.2.1 Oblique 3D Modeling and Geographic Feature Extraction

The prerequisite for precise path planning is an accurate representation of the building geometry. Initially, a UAV imaging system captures the target building from multiple perspectives. This multi-view acquisition strategy reduces visual occlusions and provides sufficient image coverage for high-quality photogrammetric reconstruction.

Subsequently, the collected imagery is processed through a photogrammetric workflow, including image distortion correction, feature matching, bundle adjustment, dense point cloud generation, and mesh reconstruction. This process, often implemented using Multi-View Stereo algorithms in software such as Agisoft Metashape or ContextCapture [6,7], produces a high-precision oblique photogrammetric 3D model of the target building, as illustrated in Figure 2.

Based on the georeferenced 3D model, the geographic coordinates, including longitude, latitude, and altitude, of the building ’ s key corner points are extracted. These coordinates provide the geometric constraints for generating the subsequent inspection waypoints.

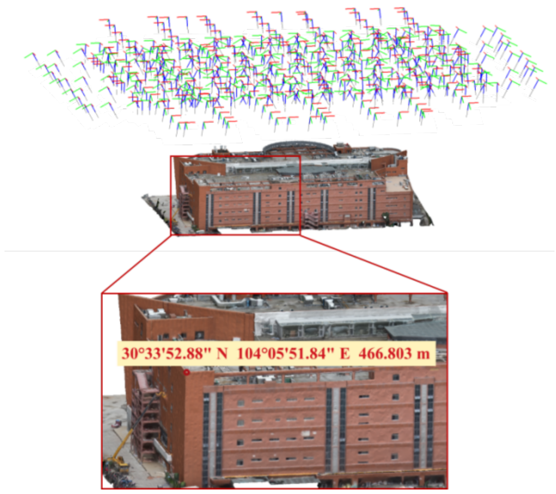


Figure 2. Process of oblique photogrammetry 3D reconstruction and key point extraction

2.2.2 Waypoint Generation Mechanism

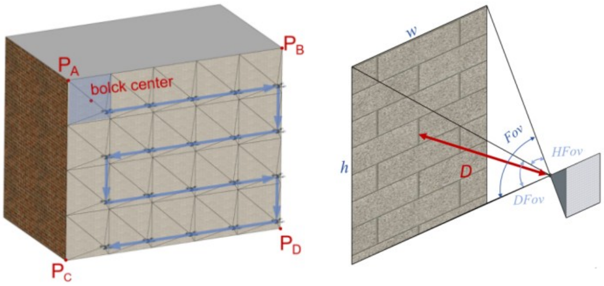
To ensure complete coverage of the façade, the continuous wall surface must be discretized into a set of discrete inspection waypoints.

First, the physical dimensions of the single-image footprint are determined. Let the four corner vertices of the target façade be denoted as $P_A(\lambda_A,\phi_A,h_A), P_B(\lambda_B,\phi_B,h_B), P_C(\lambda_C,\phi_C,h_C), P_D(\lambda_D,\phi_D,h_D)$ (Figure 3(a)). Modeling the camera as a pinhole system, the physical coverage dimensions, namely the width *w* and height *h*, at a given inspection distance *D* can be determined from the camera's horizontal and vertical fields of view, respectively, as illustrated in Figure 3(b):

$$w = 2D \cdot \tan\left(\frac{\theta_h}{2}\right) \tag{1}$$

$$h = 2D \cdot \tan\left(\frac{\theta_v}{2}\right) \tag{2}$$

where θ_h and θ_v represent the horizontal and vertical field of view (FOV) angles, respectively.



(a) Flight trajectory positioning (b) Field of View (FOV) calculation model

Figure 3. Geometric parameters for path planning

Since the extracted building key points utilize the Geographic Coordinate System (GCS, i.e., WGS84), whereas the geometric subdivision necessitates metric computation, a coordinate transformation is required. The geographic coordinates (λ,ϕ,h) are converted to the Earth-Centered, Earth-Fixed (ECEF) Cartesian coordinates (X,Y,Z) using the following standard transformation equations:

$$X = (N+h)\cos(\phi)\cos(\lambda) \tag{3}$$

$$Y = (N+h)\cos(\phi)\sin(\lambda) \tag{4}$$

$$Z = \left[\left(1-e^2\right)N + h\right]\sin(\phi) \tag{5}$$

with the radius of curvature in the prime vertical N defined as:

$$N = \frac{a}{\sqrt{1-e^2\sin^2(\phi)}} \tag{6}$$

where *a* denotes the Earth's semi-major axis $(\approx 6{,}378{,}137\text{m})$ and *e* represents the Earth's eccentricity (≈ 0.081819).

In the ECEF frame, the Euclidean distances between the corner points (e.g., P_A to P_B and P_C) are calculated to determine the total width and height of the façade:

$$d_{AB} = \sqrt{(X_B - X_A)^2 + (Y_B - Y_A)^2 + (Z_B - Z_A)^2} \tag{7}$$

$$d_{AC} = \sqrt{(X_C - X_A)^2 + (Y_C - Y_A)^2 + (Z_C - Z_A)^2} \tag{8}$$

Based on the single-image coverage (*w*, *h*) and the total façade dimensions, the façade is discretized into an $m \times n$ grid of inspection areas. To ensure sufficient image overlap between adjacent viewpoints, the horizontal and vertical step sizes are defined as $w(1-O_w)$ and $h(1-O_h)$, respectively, where O_w and O_h denote the prescribed horizontal and vertical overlap ratios. Accordingly, the required numbers of inspection columns and rows are calculated as:

$$m = 1 + \left\lceil \frac{d_{AB} - w}{w(1-O_w)} \right\rceil \tag{9}$$

$$n = 1 + \left\lceil \frac{d_{AC} - h}{h(1-O_h)} \right\rceil \tag{10}$$

The center point $(X_{\text{center}}, Y_{\text{center}}, Z_{\text{center}})$ of each grid cell is then calculated and translated outward along the normal vector $\vec{N} = (N_x, N_y, N_z)$ of the plane $P_AP_BP_C$ by the standoff distance *D*. This operation yields the precise 3D coordinates of each camera waypoint in the Cartesian frame:

$$X_{\text{shifted}} = X_{\text{center}} + D \cdot N_x \tag{11}$$

$$Y_{\text{shifted}} = Y_{\text{center}} + D \cdot N_y \tag{12}$$

$$Z_{\text{shifted}} = Z_{\text{center}} + D \cdot N_z \tag{13}$$

Finally, an inverse coordinate transformation is applied to convert these Cartesian waypoints back to the Geographic Coordinate System, allowing them to be uploaded to the drone's flight controller. The longitude λ is calculated directly, while the latitude ϕ and height h are derived using Bowring's closed-form approximation [8]:

$$\lambda = \arctan\left(\frac{Y}{X}\right) \tag{14}$$

$$\phi = \arctan\left(\frac{Z + e'^2 \cdot b \cdot \sin^3(\theta)}{p - e^2 \cdot a \cdot \cos^3(\theta)}\right) \tag{15}$$

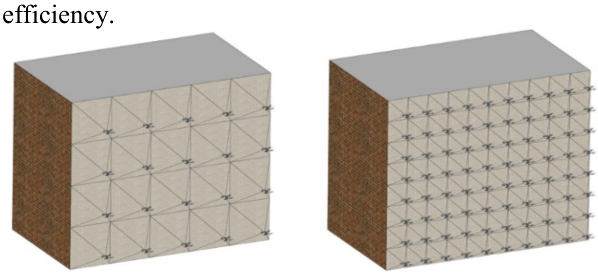
$$h = \frac{p}{\cos(\phi)} - N \tag{16}$$

To facilitate this calculation, the auxiliary parameters are defined as follows:

$$e'^2 = e^2 / (1 - e^2)\,,\; p = \sqrt{X^2 + Y^2}\,,\; b = a\sqrt{1 - e^2} \tag{17}$$

$$\theta = \arctan\left(\frac{Z \cdot a}{p \cdot b}\right) \tag{18}$$

The standoff distance D is a critical trade-off parameter. As illustrated in Figure 4, a smaller D yields smaller (higher spatial resolution) Ground Sampling Distance (GSD) and richer detail but limits the single-image field of view. This necessitates a denser grid of waypoints, resulting in longer flight times. Conversely, a larger D improves efficiency but may compromise the detection of fine cracks. Accordingly, D=10 m was selected for the field deployment as a practical compromise between image resolution and inspection efficiency.



(a) Sparse sampling at a larger distance (b)Dense sampling at a closer distance

Figure 4. Impact of standoff distance on waypoint density

2.2.3 Executable Flight Path Generation

Once the discrete waypoints are generated, the final step is to generate an executable inspection trajectory by determining an appropriate traversal sequence. The objective is to visit all waypoints while minimizing total flight time and energy consumption, subject to practical constraints such as battery endurance and a predefined safety standoff from the façade.

This problem can be addressed using coverage patterns and, when needed, graph-based routing. Geometric coverage strategies such as the Boustrophedon (back-and-forth) or spiral patterns provide simple and consistent coverage with controllable overlap [9]. When obstacle-aware local routing is required, shortest-path solvers (e.g., Dijkstra or A*) can be used to connect adjacent waypoint segments under additional constraints. Figure 5 compares the trajectories produced by different strategies. For façade inspection, the Boustrophedon pattern is adopted in this study due to its implementation simplicity and stable overlap management.

Finally, the optimized trajectory is converted into a DJI Waypoint Mission by exporting the ordered waypoint list together with mission parameters (e.g., altitude, speed, heading/yaw, gimbal angle, and camera triggering). This enables the planned route to be directly uploaded to the DJI flight controller for autonomous execution.

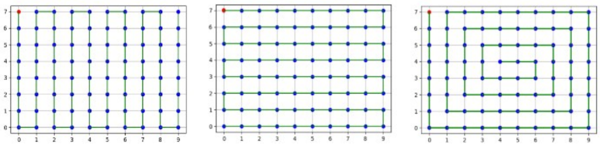


Figure 5. Flight paths generated by different methods

2.3 Façade Defect Detection and Fusion

This section details the proposed dual-modal defect detection framework, which employs deep learning models to process RGB and thermal imagery independently, followed by a decision-level fusion strategy to synthesize the results (as shown in Figure 6).

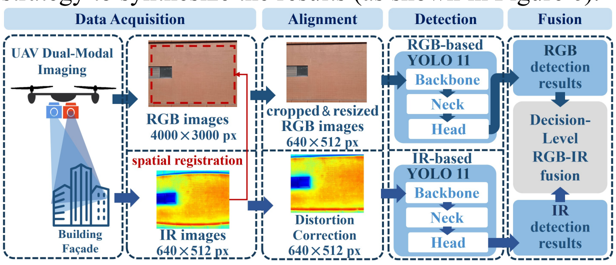


Figure 6. The proposed dual-modal defect detection framework

2.3.1 Surface Anomaly Segmentation via RGB Imagery

RGB imagery is effective in capturing high-frequency texture information and surface chromatic features. However, latent subsurface delaminations typically lack distinct visual characteristics in the early stages. They become observable in the visible spectrum only when the defect evolves into external manifestations, such as peeling or cracking. Consequently, RGB data is primarily utilized in this study to detect surface-level damage (e.g., tile cracking and peeling) associated with advanced deterioration.

To automate defect extraction, YOLO11-seg was adopted from the Ultralytics YOLO family [10,11] because it provides a strong balance between segmentation accuracy and inference efficiency. In this study, it was used as the modality-specific segmentation backbone for both the RGB and thermal branches. As a one-stage segmentation model, YOLO11-seg performs feature extraction, multi-scale feature aggregation, and pixel-level mask prediction in an end-to-end manner, enabling robust defect segmentation across different scales.

2.3.2 Subsurface Delamination Detection via Infrared Thermography

Infrared thermography is employed to reveal nonvisible subsurface defects based on differences in thermal inertia and heat-transfer behavior. In delaminated areas, an air gap between the tile and the substrate increases thermal resistance and disrupts heat conduction. Under solar excitation or during the cooling phase, the detached façade surface exhibits a different heating/cooling rate compared with sound areas, producing a measurable temperature contrast ΔT on the surface that can be captured by the thermal sensor [12].

In this study, the thermal images were converted into pseudo-color representations and used as inputs to a dedicated model. The model was fine-tuned on the infrared dataset to segment thermally anomalous regions associated with subsurface delamination. These thermal anomalies, typically manifested as localized hot or cold regions, are often not visible in RGB imagery.

2.3.3 Decision-Level Fusion of Detection Results

To obtain a holistic and physically interpretable assessment of façade condition, we adopt a decision-level fusion strategy [13] that operates on the independent segmentation outputs of the RGB and IR branches. Unlike pixel-level or feature-level fusion, the proposed strategy integrates detection results after modality-specific inference. This ensures that the distinct physical semantics of each sensor are preserved: visible imagery captures superficial appearance anomalies (M_{surf}), such as cracks and peeling, whereas infrared thermography highlights subsurface structural defects (M_{sub}) through thermal variations.

A key objective of the fusion is to identify compound defects, i.e., surface damage spatially correlated with subsurface delamination, since such regions often indicate elevated structural risk. To account for minor registration errors and boundary uncertainties between the dual-modal images, a morphological dilation is first applied to the subsurface defect mask:

$$\widetilde{M}_{sub} = \text{Dilation}(M_{sub}, r) \tag{19}$$

where r denotes the radius of the structuring element and serves as a small spatial tolerance for overlap matching, rather than a substantial geometric expansion of the defect region. In the present dataset, the residual RGB-IR misalignment after registration was empirically found to remain within approximately 5 infrared pixels. Accordingly, r was selected as a small value within this tolerance range, so that true cross-modal overlap would not be missed due to slight spatial deviation.

Based on the spatial relationship between the surface mask M_{surf} and the dilated subsurface mask $\widetilde{M}_{sub}$, the final fused result map M_{final} is constructed. For any given pixel p on the façade, the defect category is determined according to the following hierarchical logic:

$$M_{final}(p) = \begin{cases} \text{Compound Defect,} & \text{if } p \in (M_{surf} \cap \widetilde{M}_{sub}) \\ \text{Surface Defect,} & \text{if } p \in (M_{surf} \setminus \widetilde{M}_{sub}) \\ \text{Subsurface Defect,} & \text{if } p \in (M_{sub} \setminus M_{surf}) \\ \text{Background,} & \text{otherwise} \end{cases} \tag{20}$$

This hierarchical rule prioritizes compound defects by assigning the label Compound Defect to regions where visible surface anomalies overlap with subsurface delamination, while regions detected by only one modality are retained as Surface Defect or Subsurface Defect. In this way, the fusion result links superficial manifestations to potential internal risks and provides a more interpretable representation of façade condition. From an engineering perspective, cracks or peeling that spatially coincide with thermal anomalies may indicate an elevated likelihood of underlying delamination. In addition, because the RGB and IR models are trained independently, the proposed decision-level fusion decouples optimization across modalities, which simplifies implementation and facilitates the integration of heterogeneous sensing results.

3 Field Experiments and Analysis

3.1 Experimental Setup and Data Acquisition

To validate the feasibility and robustness of the proposed framework, a field experiment was conducted on a commercial building. The DJI Mavic 3T drone was employed as the flight platform. This UAV is equipped with a synchronized wide-angle RGB camera and a thermal infrared camera, enabling the simultaneous acquisition of dual-modal imagery.

The flight mission was executed following the path planning method described in Section 2.2. Figure 7 illustrates the planned trajectory. The captured RGB-IR image pairs are presented in Figure 8. It can be observed that the images collected autonomously along the planned route exhibit high spatial consistency and sufficient overlap, which establishes a high-quality dataset for subsequent algorithmic analysis.

The field deployment was conducted in November under clear-sky conditions, with stable daylight and relatively low outdoor wind speed, which were favorable for both RGB and thermal data acquisition. The thermal campaign was carried out in two rounds corresponding to the façade heating and cooling stages on November 29, 2023, in Chengdu, China, for a façade oriented about 40° south of east. At around 10:30 a.m., the solar azimuth was close to the façade normal direction, providing favorable direct solar excitation. At that time, the incident solar irradiance was approximately 750 W/m², the ambient temperature was about 14°C, and the convective heat-transfer coefficient was estimated to be 8-10 W/(m²·K). By contrast, at around 5:00 p.m., the façade had entered its self-shadowed condition and direct solar heating was no longer present. The ambient temperature was about 15°C, and the convective heat-transfer coefficient was approximately 10-12 W/(m²·K). As a result, the acquired thermal infrared images covered both the heating and cooling phases of the façade, during which delamination-related thermal signatures were relatively distinct. These conditions provided thermal images with favorable signal-to-noise characteristics for subsequent model training and defect recognition. It should be noted, however, that the quality of defect-related visual and thermal cues may vary under different environmental conditions: RGB imagery is more sensitive to illumination changes, whereas thermal infrared imagery is more strongly influenced by solar loading, wind, and the thermal properties of façade materials.

Figure 7. Visualization of the executed flight path in the 3D model

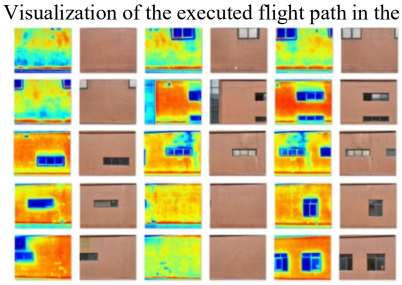


Figure 8. Samples of synchronized RGB and IR image pairs collected during the flight

Since the decision-level fusion in this study depends on the spatial correspondence between visible and thermal defect regions, an explicit cross-modal alignment procedure was applied before defect analysis. The visible and thermal cameras mounted on the UAV gimbal have a fixed relative pose, which provides the geometric basis for stable registration. Both sensors were first calibrated using a calibration board, and all images were corrected for lens distortion prior to alignment. Because the UAV acquired façade images at a fixed standoff distance along the planned route and the imaging parameters were kept unchanged throughout the mission, the relative projection relationship between the two sensors remained approximately constant across the dataset. Therefore, after distortion correction, a single affine transformation matrix was adopted for all RGB-IR image pairs.

The affine transformation parameters were estimated from five representative RGB-IR image pairs with clearly identifiable shared features, and the final matrix was obtained by averaging the estimated parameters. To evaluate the registration quality, 20 additional image pairs were randomly selected for verification. The results showed that the alignment achieved an average registration error of approximately 2 pixels, with all observed errors remaining within 5 pixels. In addition, because the RGB and thermal sensors have different native resolutions, the visible images were resampled to the thermal image size (639×509 pixels) before registration and model inference. This resolution unification improved the consistency of cross-modal spatial matching and also simplified the subsequent defect-detection pipeline.

3.2 Model Implementation and Training Details

The dataset combined the current field campaign with previously accumulated inspection data from the same façade. The waypoint route and two-round acquisition process contributed 416 paired images, yielding a final raw dataset of 766 RGB and 766 original IR images. For the thermal branch, each original infrared image was converted into pseudo-color images using three different palettes. To avoid data leakage, the train/validation/test split was performed at the level of the original RGB-IR pairs before pseudo-color expansion, using a ratio of 7:2:1. The RGB branch focused on visible surface defects, whereas the thermal branch focused on thermally anomalous regions associated with subsurface delamination.

The experimental workflow consisted of data annotation, model training, and performance evaluation. In the data annotation step, the collected images were labeled using LabelMe. Annotation masks were created by experienced civil engineers, who manually annotated tile cracks and peeling in RGB images, and thermally anomalous regions associated with subsurface delamination in infrared images.

For modality-specific defect extraction, two dedicated segmentation models were fine-tuned using the Ultralytics YOLO11 framework, one for the RGB branch and the other for the thermal branch. Training was conducted with an input resolution of 640×640 pixels, a batch size of 16, the AdamW optimizer, and early stopping with a patience of 100 epochs. To improve robustness, data augmentation including color perturbation, rotation, translation, scaling, and flipping was applied to the training set. The best-performing checkpoint was automatically selected based on validation performance. The evolution of the validation metrics during training is shown in Figure 9.

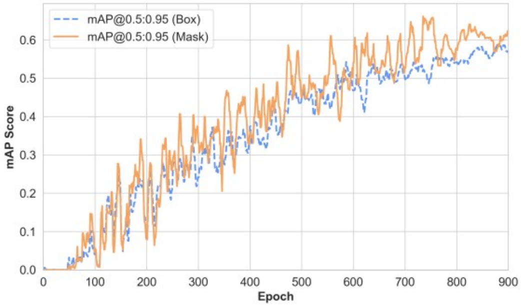


Figure 9. Evolution of validation box and mask mAP@0.5:0.95 over training epochs

3.3 Quantitative and Qualitative Analysis

Quantitative evaluation on the test set demonstrates the robust segmentation capability of the trained YOLO11n-seg model. For the segmentation masks, the overall precision and recall reached 0.795 and 0.869, respectively, with an mAP@0.5 of 0.829 and an mAP@0.5:0.95 of 0.619. Class-wise results indicate that the framework maintained effective performance across subsurface delamination, peeling, and crack categories, although subsurface delamination and crack remained more challenging than peeling. These results confirm that the framework can effectively segment both surface-visible defects and thermally expressed subsurface anomalies. The adopted YOLO11n-seg backbone also showed favorable inference efficiency, with an average per-image processing time of 0.8ms for preprocessing, 0.5ms for inference, and 1.2ms for postprocessing, supporting rapid batch processing and timely on-site review after UAV data acquisition.

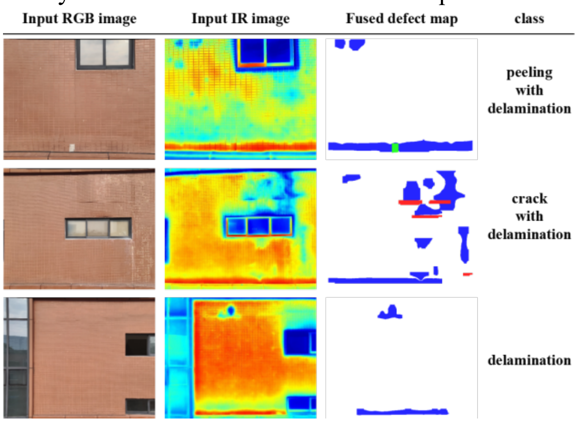


Figure 10. Visualization of defect segmentation results

Qualitative identification results are presented in Figure 10. The proposed framework effectively characterizes damage types in both modalities, identifying tile cracks and peeling in RGB imagery while simultaneously detecting thermally anomalous regions associated with subsurface delamination in infrared imagery. Following the decision-level fusion strategy, the modality-specific masks are combined to generate a comprehensive façade condition map.

In addition to qualitative visualization, the practical value of the decision-level fusion was assessed at the façade level. RGB-only results identified visible surface defects such as cracks and peeling, whereas IR-only results identified thermally anomalous regions associated with subsurface delamination. Based on the spatial-overlap rules, the fusion process further identified 18 compound-defect regions, including crack-with-delamination and peeling-with-delamination cases, which could not be explicitly characterized by either modality alone and therefore provide more actionable information for maintenance prioritization.

A brief sensitivity check, performed by varying the dilation radius from 2 to 8 pixels, showed that the compound-defect results remained largely stable, indicating limited sensitivity to this tolerance parameter under the present registration accuracy.

Field verification was conducted according to defect visibility and accessibility. Visible defects such as cracks and peeling were manually reviewed, accessible low-rise suspected delamination areas were examined by hammer sounding, and higher or less accessible regions were checked indirectly using a telephoto camera, particularly where bulging or outward deformation was visible. The majority of detected defects were practically confirmed, although some inaccessible and visually flat regions could not be fully verified. Therefore, the current validation should be regarded as a practical engineering verification rather than a fully exhaustive ground-truth survey.

4 Discussion

The proposed framework represents a practical improvement over conventional façade inspection and maintenance practices. The field experiments demonstrated that digital 3D façade models can be effectively translated into executable UAV inspection missions, reducing dependence on pilot experience and improving the consistency of image acquisition. In addition, the combination of visible and thermal infrared imagery helps overcome the limitations of single-sensor inspection: visible imagery captures surface defects such as cracks and peeling, whereas thermal infrared imagery provides complementary information related to subsurface delamination. The experimental results show that the proposed dual-modal framework can exploit this complementarity effectively, while the decision-level fusion strategy further improves the interpretability of façade condition by linking visible deterioration with potential internal risks.

The transferability of the framework should be understood at both the detector and fusion levels.

Detector performance may vary with façade material, surface texture, and environmental conditions, whereas the rule-based fusion logic is expected to remain broadly applicable as long as relevant visual and thermal cues are observable. Nevertheless, the current study was validated on a tiled planar façade, and broader validation on other façade types and more complex geometries is still needed. The current path-planning algorithm is mainly suited to planar façades, although irregular envelopes may be approximated as multiple locally planar patches for practical inspection. In addition, thermal-anomaly visibility depends on solar-loading and wind conditions. Future work will focus on broader validation, adaptive planning for complex 3D geometries, and deeper cross-modal fusion.

5 Conclusions

This study presents a façade inspection framework that integrates efficient UAV-based data acquisition with deep-learning-based defect detection for façade safety assessment. The proposed workflow was developed and validated through a field deployment on a commercial building, including oblique photogrammetry-based 3D modeling, route generation, and defect recognition based on modality-specific segmentation and decision-level fusion. The results demonstrate that the proposed approach enables efficient and safe inspections while providing a more informative assessment of both surface-visible defects and subsurface delamination risks.

The main contributions of this research are summarized as follows:

(1) Automated flight route generation: A UAV inspection route generation method was developed based on architectural key points extracted from oblique photogrammetry models, enabling consistent and automated façade data acquisition through waypoint-based missions.

(2) Dual-modal defect detection and interpretable fusion. A defect-identification framework was established using dedicated YOLO11-seg models for visible and thermal infrared imagery, followed by decision-level fusion of the modality-specific outputs. By associating surface appearance anomalies with subsurface thermal signatures, the framework provides a more comprehensive and practically interpretable assessment of façade condition, offering useful support for intelligent infrastructure inspection and maintenance.

6 Acknowledgement

This research is supported by Key Technologies R&D Program (Grant No. 2024YFC3808602), Shenzhen Science and Technology Programs (Grant No. KJZD20230923114310021, Grant No. JCYJ20241202123722029).

References

[1] P. Wang, J. Xiao, Z. Duan, et al. Intelligent development trend of building enclosure damage detection. J. Archit. Civ. Eng., 39(4):24-37, 2022.

[2] J. D. Silvestre and J. De Brito. Ceramic tiling in building façades: Inspection and pathological characterization using an expert system. Constr. Build. Mater., 25(4):1560-1571, 2011.

[3] X. Wang, C. Demartino, Y. Narazaki, et al. Rapid seismic risk assessment of bridges using UAV aerial photogrammetry. Eng. Struct., 279:115589, 2023.

[4] S. Jordan, J. Moore, S. Hovet, et al. State-of-the-art technologies for UAV inspections. IET Radar Sonar Navig., 12(2):151-164, 2018.

[5] S. Ivić, B. Crnković, L. Grbčić, et al. Multi-UAV trajectory planning for 3D visual inspection of complex structures. Autom. Constr., 147:104709, 2023.

[6] A. Y. Yiğit and M. Uysal. Automatic crack detection and structural inspection of cultural heritage buildings using UAV photogrammetry and digital twin technology. J. Build. Eng., 109952, 2024.

[7] M. Štroner, R. Urban, J. Seidl, T. Reindl, and J. Brouček. Photogrammetry using UAV-mounted GNSS RTK: Georeferencing strategies without GCPs. Remote Sens., 13:1336, 2021.

[8] B. R. Bowring. Transformation from spatial to geographical coordinates. Survey Review, 23(181):323-327, 1976.

[9] T. M. Cabreira, L. B. Brisolara, and R. F. J. Paulo. Survey on coverage path planning with unmanned aerial vehicles. Drones, 3(1):4, 2019.

[10] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr., 5(4):1680-1716, 2023.

[11] Ultralytics. "Ultralytics YOLO11," 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

[12] I. Garrido, E. Barreira, R. M. S. F. Almeida, et al. Introduction of active thermography and automatic defect segmentation in the thermographic inspection of specimens of ceramic tiling for building façades. Infrared Phys. Technol., 121:104012, 2022.

[13] X. Yang, R. Guo, and H. Li. Comparison of multimodal RGB-thermal fusion techniques for exterior wall multi-defect detection. J. Infrastruct. Intell. Resil., 2(2):100029, 2023.