

Zhiyi Building Inspection: A UAV-based Building Facade Damage Detection System

Wenxuan Tang*, Hongqi Zhang, Mengqi Li, Shuyu Yang, Furong Liu
College of Education, Zhejiang Normal University, Jinhua 321000, China

*Corresponding email: 1796794131@qq.com

Abstract

Building facade inspection is essential for maintaining structural safety, yet conventional approaches remain limited by high-altitude risks, labour intensity, and inconsistent evaluation. This study develops Zhiyi Building Inspection, an integrated unmanned aerial vehicle-artificial intelligence-building information modeling (UAV-AI-BIM) framework for automated facade damage detection. The system combines UAV-based image acquisition with real-time kinematic (RTK) positioning, an improved You Only Look Once version 5 Large (YOLOv5l) detection model incorporating Complete Intersection over Union (CIoU) loss and Convolutional Block Attention Module (CBAM), Sobel-based defect quantification, severity assessment, and BIM-assisted visualisation. A dataset of 3,982 high-resolution images covering eight facade-defect categories was constructed for model training and evaluation. Experimental results demonstrate that the proposed framework improves detection accuracy while maintaining real-time processing capability and provides engineering-oriented outputs including defect location, severity, and inspection reports. The study demonstrates the feasibility of integrating UAV sensing, artificial intelligence, and digital building management for safer and more efficient facade maintenance.

Keywords

Unmanned aerial vehicle, Building facade damage, Deep learning, You Only Look Once version 5 Large, Building information modeling visualisation

Introduction

Building facade integrity is closely related to structural health monitoring, urban safety management, and the long-term maintenance of the built environment. With the continued growth of high-rise and ageing building stock, facade defects such as cracks, delamination, corrosion, spalling, exposed reinforcement, and tile detachment can create substantial safety risks. Conventional inspection still relies heavily on manual visual examination, hammer sounding, scaffolds, suspended platforms, and rope-access work. Although these methods offer direct engineering interpretation, they are labour-intensive, experience-dependent, difficult to scale, and expose inspectors to high-altitude hazards.

Sensing-based alternatives such as infrared thermography and three-dimensional scanning can improve defect discovery and geometric measurement, but their performance, cost, and operational complexity remain important constraints [1]. Three-dimensional scanning and Scan-to-BIM workflows strengthen

geometric documentation and the connection between measured conditions and digital building information [2]. These sensing approaches are valuable complements to visual inspection, but they also demonstrate that inspection performance depends on both sensing capability and the usability of the resulting information. Unmanned aerial vehicles (UAVs) provide a practical means of improving facade data acquisition because they can rapidly access elevated, irregular, or otherwise difficult-to-reach surfaces while reducing direct human exposure.

Previous research has demonstrated the value of UAV platforms for infrastructure inspection and construction monitoring, particularly when high-resolution imaging, autonomous flight planning, and positioning technologies are combined [3]. A broader review of unmanned aerial system (UAS) applications in the built environment likewise identifies accessibility, repeatability, and rapid visual documentation as major advantages

of aerial inspection [4]. UAV-based infrastructure studies have further shown that coupling aerial imagery with deep-learning analysis can reduce the burden of manual image review [5]. More recent intelligent-construction research has emphasized the value of combining UAV sensing with artificial intelligence to support automated monitoring workflows [6].

However, UAVs alone do not resolve the analytical bottleneck created by large volumes of inspection imagery. Illumination variation, material diversity, background texture, and the small visual footprint of many facade defects make reliable automated interpretation a central technical challenge. Deep learning has therefore become an important component of intelligent inspection. Convolutional neural networks and one-stage object detectors can automatically learn hierarchical visual features and jointly identify defect categories and locations.

The broader deep-learning framework described by Dong et al. explains how representation learning reduces dependence on manually engineered visual features [7]. Residual learning subsequently demonstrated that deeper networks can be trained effectively for visual recognition, providing an important architectural foundation for modern detection backbones [8]. Convolutional neural networks (CNNs)-based construction research has shown that these learned representations can be transferred to concrete crack identification [9]. The You Only Look Once (YOLO) family is especially relevant to UAV applications because of its balance between recognition performance and real-time computational efficiency [10].

The study therefore proposes Zhiyi Building Inspection, an integrated UAV-AI-BIM facade damage inspection framework combining UAV image acquisition, RTK positioning, improved YOLOv5l detection, defect segmentation, severity evaluation, and BIM visualisation. The framework aims to transform aerial images into engineering information that supports facade maintenance decisions. The following sections review related research, establish the conceptual model, present the methodology, evaluate system performance, discuss engineering implications, and summarise limitations and future research directions.

Literature review

Traditional building facade inspection methods and their limitations

Building facade inspection is an essential component of

structural health monitoring and urban safety management. With the rapid growth of high-rise buildings and the ageing of existing structures, ensuring facade integrity has become increasingly important. Defects such as cracks, delamination, corrosion, and tile detachment can pose significant threats to public safety. Traditional facade inspection methods predominantly rely on manual visual examination, hammer sounding, and close-range investigations conducted by experienced inspectors. Although these approaches remain widely used due to their simplicity and capacity for direct engineering interpretation, they are highly dependent on human expertise. Consequently, they are difficult to apply efficiently to large-scale building inventories.

Previous studies have identified several limitations inherent in manual inspection. First, inspectors are often required to work at considerable heights using scaffolds, suspended platforms, or ropes, which creates significant occupational safety risks. Second, the accuracy of defect identification is strongly influenced by an inspector's experience and subjective judgement. Small cracks, hidden delamination, and early-stage deterioration are frequently difficult to recognize through visual observation alone. Third, traditional inspection typically produces fragmented records, complicating the establishment of continuous monitoring and predictive maintenance systems. To overcome these limitations, researchers have explored various sensor-based inspection technologies. Infrared thermography, for instance, has been applied to detect facade defects by identifying abnormal thermal patterns caused by material discontinuities. H. Kim et al. demonstrated that infrared imaging could effectively identify subsurface defects in building envelopes [11]. However, the effectiveness of infrared thermography is heavily influenced by environmental conditions, including solar radiation, temperature differentials, and weather changes. Moreover, defects lacking pronounced thermal characteristics often cannot be accurately detected. Laser scanning and three-dimensional reconstruction technologies have also been introduced into building inspection. Bosché et al. suggested that three-dimensional sensing technologies could improve geometric measurement accuracy and support automated construction analysis.

Nevertheless, these methods often require expensive

equipment and complex data processing procedures, limiting their practical application in routine facade maintenance. Therefore, although both traditional and sensor-based approaches provide valuable inspection capabilities, existing methods still face challenges related to safety, efficiency, cost, and scalability. These limitations highlight the necessity of developing an intelligent inspection framework that integrates autonomous data acquisition with advanced artificial intelligence algorithms.

UAV-based building inspection: Opportunities and challenges

UAVs have attracted considerable attention in construction monitoring due to their flexibility, mobility, and ability to access dangerous or otherwise inaccessible areas. Compared with traditional manual inspection, UAV-based inspection can significantly reduce human exposure to hazardous environments while improving data collection efficiency. Many researchers have investigated UAV applications in infrastructure inspection. For example, Ellenberg et al. demonstrated that UAV imagery could provide an efficient alternative for bridge inspection by collecting high-resolution visual information from difficult-to-access locations.

Similarly, Rakha and Gorodetsky emphasized that UAV platforms could improve construction monitoring efficiency by integrating autonomous flight planning, digital imaging, and automated analysis. In building facade inspection, UAVs offer several advantages. First, they can rapidly collect high-resolution images of large facade areas without requiring complex scaffolding systems. Second, UAVs can capture data from irregular architectural structures, including curved surfaces, narrow spaces, and elevated components. Third, when combined with positioning technologies such as RTK positioning, UAV systems can establish accurate spatial relationships between detected defects and building locations. However, UAV-based inspection still faces several technical challenges. One major challenge is the massive volume of image data generated during inspection missions. Without effective automated analysis methods, human operators must manually review thousands of images, which reduces practical efficiency.

Another challenge is that facade environments involve

complex visual conditions, including illumination variation, material diversity, texture interference, and background noise. These factors may reduce the reliability of conventional image-processing algorithms. Consequently, UAV platforms alone cannot fully overcome the limitations of traditional inspection. The integration of UAV sensing with artificial intelligence-based image understanding has therefore become an important research direction.

Deep learning-based visual defect detection

The development of deep learning has significantly transformed computer vision applications in engineering inspection. Compared to traditional image-processing methods that rely on manually designed features, deep learning models can automatically extract hierarchical visual representations from large-scale datasets. CNNs have become one of the most widely adopted approaches for image-based defect detection. LeCun et al. demonstrated the powerful feature extraction capability of deep neural networks in visual recognition tasks [12]. In construction engineering, Cha et al. applied CNN models for concrete crack detection and achieved improved recognition performance compared to traditional approaches. Object detection algorithms have further expanded the capability of automated inspection by simultaneously identifying defect categories and locating their positions. The YOLO series proposed by Pham et al. introduced a one-stage detection framework that achieves real-time object recognition [13]. Compared to two-stage detection models such as Faster Region-based convolutional network (R-CNN), YOLO models provide higher computational efficiency, making them suitable for embedded or UAV-based applications. Recent studies have applied YOLO-based algorithms to various construction defect detection scenarios. B. Kim et al. developed a deep learning-based crack detection system using object detection techniques and demonstrated the feasibility of automated construction inspection [14]. Zhang et al. further improved detection performance by incorporating attention mechanisms into deep learning models, allowing the networks to focus more effectively on defect-related features [15]. Nevertheless, facade defect detection presents unique difficulties.

Many facade defects are relatively small compared to the whole image, resulting in insufficient feature representation. Additionally, facade materials such as ceramic tiles, glass curtain walls, and concrete surfaces have complex textures that may cause false detections. Therefore, model optimisation strategies, including attention mechanisms, improved loss functions, and domain-specific datasets, are necessary. In this project, the intelligent facade inspection model integrates an improved YOLOv5l architecture with CIoU Loss and CBAM attention mechanisms.

The system also incorporates engineering knowledge by combining defect detection results with damage classification and quantitative assessment. This approach addresses the limitation that conventional deep learning models often focus only on recognition accuracy while lacking engineering applicability. The project establishes a standardized dataset containing multiple facade materials, environmental conditions, and eight defect categories, providing a foundation for domain-specific model training.

Research gap and theoretical foundation of this study

Existing studies have demonstrated the potential of UAVs and deep learning for automated infrastructure inspection. However, several research gaps remain. First, many previous studies focus primarily on defect recognition accuracy but pay insufficient attention to engineering usability. In practical facade maintenance, merely identifying defect categories is insufficient.

Engineers also require information regarding defect location, severity, size, and maintenance priority. Second, existing AI inspection systems often lack integration between data acquisition, defect recognition, spatial positioning, and maintenance decision-making. A practical intelligent inspection system should form a complete workflow from image collection to engineering evaluation. Third, current research datasets are often limited to specific materials or laboratory conditions.

The diversity of facade materials and environmental conditions requires more comprehensive datasets to improve model robustness. Therefore, this study proposes an integrated UAV-AI facade inspection framework. The framework combines autonomous UAV data collection, deep learning-based defect recognition, RTK-based spatial positioning, image segmentation-

based quantitative measurement, and BIM visualisation. The theoretical foundation is based on three perspectives: computer vision theory, which supports automated feature extraction and defect recognition; structural health monitoring theory, which emphasizes continuous condition assessment; and digital construction theory, which promotes integration between inspection data and building information management.

These gaps suggest that intelligent facade inspection should be understood as a chain of interdependent capabilities rather than as a single recognition task. Image quality determines the visual information available to the detector; detection quality influences subsequent segmentation and measurement; spatial positioning determines whether a detected defect can be associated with a physical building location; and information organisation determines whether the result can support maintenance decisions. A weakness at any stage can therefore reduce the value of the entire workflow. This systems perspective also changes the meaning of model performance.

In a generic computer vision benchmark, precision, recall, and mean average precision may be sufficient to compare algorithms. In engineering inspection, however, the consequence of an error depends on the physical meaning of the target. Missing a severe defect may be more consequential than producing an additional false alarm on a minor surface feature. Similarly, correct recognition without usable localization may still fail to support field maintenance.

The proposed framework therefore treats recognition metrics as necessary but not sufficient indicators of practical inspection quality. Structural health monitoring provides a second theoretical rationale for this integration. Inspection becomes more valuable when observations can be repeated, spatially referenced, and compared over time. Standardized UAV trajectories and georeferenced imagery can create a consistent basis for future reinspection, while BIM-oriented information management can associate defect records with building components.

Although long-term deterioration prediction is outside the scope of the present validation, a structured and spatially referenced inspection record is an important prerequisite for future condition-based maintenance.

Model design and hypotheses

Conceptual model development

Based on the literature review and identified research gaps, this study develops a conceptual model for UAV-based intelligent facade damage detection. The proposed model posits that intelligent inspection performance is determined by three core technological capabilities: sensing capability, algorithmic intelligence, and engineering integration capability. The first capability is UAV-based sensing. UAV platforms improve inspection accessibility by replacing risky manual operations and enabling rapid image acquisition. High-resolution cameras combined with RTK positioning provide both visual information and spatial coordinates. Therefore, UAV sensing capability is expected to positively influence inspection efficiency and data quality.

The second capability is artificial intelligence detection. Deep learning models improve automatic defect recognition by extracting complex visual features. In this study, the improved YOLOv5l model introduces CBAM attention mechanisms to enhance feature extraction and applies CIoU Loss to improve localization accuracy. Such optimisation is expected to enhance defect detection performance.

The third capability is engineering integration. Unlike general computer vision tasks, building inspection requires outputs that can support maintenance decisions. Therefore, this study incorporates Sobel-based defect segmentation, damage grading, and BIM visualisation to transform detection results into engineering information. Based on these theoretical relationships, the proposed model assumes that UAV sensing, AI recognition, and engineering integration jointly determine the effectiveness of intelligent facade inspection.

Research hypotheses

Hypothesis 1: UAV-based data acquisition capability positively influences inspection efficiency.

UAV technology provides significant advantages in facade inspection by reducing dependence on manual high-altitude operations and increasing spatial coverage. Compared with traditional inspection approaches, UAV systems can collect facade images rapidly and repeatedly under different inspection scenarios. The effectiveness of AI-based detection depends heavily on the quality and diversity of collected data. High-resolution images,

accurate positioning information, and standardized acquisition procedures provide reliable inputs for deep learning models.

Hypothesis 2: Deep learning-based detection capability positively influences damage recognition accuracy.

Deep learning models have demonstrated strong capabilities in extracting visual patterns from complex images. However, facade defect detection involves small targets, irregular shapes, and complex backgrounds. Model improvements such as attention mechanisms and optimized loss functions can enhance feature discrimination and localization accuracy. The improved YOLOv5l model used in this study introduces CBAM attention mechanisms to strengthen important defect features and adopts CIoU Loss to improve bounding box regression.

Hypothesis 3: Engineering-oriented quantification capability positively influences maintenance decision support.

Previous studies mainly evaluate AI inspection systems based on recognition accuracy, while practical engineering applications require quantitative defect information. Damage area, spatial location, and severity classification are essential for maintenance planning. By combining object detection with Sobel edge segmentation and RTK-based spatial conversion, the proposed system can estimate defect dimensions and classify damage levels.

Hypothesis 4: Integrated UAV-AI-BIM framework improves overall inspection performance.

A complete intelligent inspection system requires not only accurate detection but also effective information management. Integrating UAV acquisition, AI recognition, quantitative analysis, and BIM visualisation enables a closed-loop process from detection to maintenance. The proposed system aims to overcome the fragmentation of traditional inspection workflows by automatically generating standardized inspection reports and visualizing defect distributions.

Method

Overall framework

This study proposes an intelligent building facade inspection framework that integrates UAV sensing, deep learning-based damage detection, image processing, and engineering evaluation. The framework aims to

overcome the limitations of traditional manual inspection methods, which are often constrained by safety risks, low efficiency, and subjective evaluations. The system adopts a three-layer architecture consisting of a data acquisition layer, an algorithm layer, and an application layer. The data acquisition layer is responsible for collecting high-resolution facade images and spatial positioning information.

The algorithm layer performs automated defect recognition and quantitative analysis using deep learning and computer vision techniques. The application layer integrates detection results with spatial information to generate visual inspection outputs and maintenance reports. The complete workflow includes four major steps: UAV-based image acquisition, dataset construction and preprocessing, deep learning-based damage detection and segmentation, and damage assessment and visualisation. This framework establishes an integrated process from data collection to engineering decision support.

UAV-based image acquisition and dataset construction

To obtain comprehensive facade information while reducing inspection blind areas, a UAV-based image acquisition strategy was developed. A vertical layered flight trajectory was designed to ensure sufficient coverage of building exterior surfaces. During data acquisition, adjacent images were maintained with approximately 10.0% overlap, allowing continuous reconstruction of facade information. The UAV platform was equipped with a high-resolution 4K camera and an RTK positioning module.

The camera was used to collect detailed Red Green Blue (RGB) images of facade surfaces, while the RTK module provided accurate spatial coordinates for detected defects. The combination of visual information and positioning data enables the system to locate damage areas more precisely. The dataset was constructed by integrating three types of image sources: UAV-collected facade images, publicly available building damage datasets, and historical inspection images. All images were manually annotated using Labellmg, and eight typical facade damage categories were included in the dataset.

Considering the influence of illumination changes, surface materials, and environmental conditions on visual recognition, several preprocessing methods were

applied. These methods included image cropping, difficult-sample enhancement, and Mosaic data augmentation. Data augmentation is widely used in deep learning-based detection tasks to improve model generalisation ability under complex conditions [16]. After preprocessing, a standardized dataset containing 3,982 high-resolution images was established for model training and evaluation.

The acquisition strategy was designed to improve not only coverage but also the consistency of the visual information supplied to the detector. Facade imagery can vary substantially with camera orientation, viewing distance, illumination, facade geometry, and UAV motion. A layered trajectory provides a more systematic sampling pattern than arbitrary manual flight and reduces the probability that elevated or geometrically irregular regions will be omitted. Image overlap supplies redundant observations near frame boundaries and helps preserve continuity between adjacent facade regions.

RTK positioning is equally important because image recognition by itself does not establish a direct relationship between a detected object and its physical location. Recording spatial information during acquisition creates the basis for later localization and BIM-assisted visualisation. Dataset construction was intended to reflect practical visual diversity rather than a single building environment. Combining UAV-collected images, public damage data, and historical inspection records broadens the range of materials, backgrounds, and acquisition conditions represented during training.

Manual annotation provides a consistent definition of the eight target categories, while augmentation exposes the model to additional combinations of scale and context. Nevertheless, augmentation cannot fully replace genuine diversity in the source data. Rare classes and visually subtle defects remain more difficult to learn, which is reflected in the class-wise results reported later in the paper.

Improved YOLOv5-based damage detection model

Deep learning-based object detection methods have demonstrated strong potential for infrastructure damage identification. Among existing detection algorithms, YOLO series models offer advantages in real-time detection speed and accuracy, making them suitable for UAV-based inspection scenarios [17]. In this study, YOLOv5 was selected as the basic detection framework.

However, facade damage detection presents several challenges, including small defect sizes, complex backgrounds, and large variations in appearance.

Therefore, improvements were introduced to enhance feature extraction and localization performance. First, the CIoU loss function was incorporated into the bounding box regression process. Compared with traditional Intersection over Union (IoU)-based loss functions, CIoU simultaneously considers the overlap ratio, center distance, and aspect ratio between predicted and ground-truth bounding boxes, leading to improved localization accuracy [18]. Second, the CBAM was introduced into the network structure. CBAM enhances feature representation by adaptively assigning weights to channel and spatial information.

Previous studies have demonstrated that attention mechanisms can improve the ability of YOLO-based models to identify small and complex damage features in challenging environments [19]. The improved YOLOv5 model outputs the category, location, and confidence score of detected facade defects. Through these improvements, the model achieves more accurate identification of irregular and small-scale building damage. The selection of YOLOv5 as the baseline framework reflects the deployment requirements of UAV inspection.

The detector must process images quickly enough for an operational workflow while retaining sufficient feature capacity for small defects. An excessively large network may improve benchmark performance at the cost of latency and memory consumption, whereas an overly lightweight model may discard weak crack and texture information. Model selection therefore requires a balance among recognition quality, inference speed, and deployment complexity. The two principal model modifications address complementary error sources.

CIoU-based regression targets geometric localization by considering overlap, center distance, and aspect-ratio consistency during optimisation. This is relevant to cracks and elongated damage regions because a modest displacement can exclude a meaningful portion of a thin target from the predicted box. CBAM addresses feature discrimination by recalibrating channel and spatial responses. Facade imagery often contains joints, windows, reflections, decorative patterns, and material boundaries that can compete with genuine damage

features.

Attention weighting helps the network emphasize informative regions without replacing the underlying convolutional feature extractor. Together, the two mechanisms improve both where the model looks and how precisely it represents the detected region.

Damage segmentation and quantitative evaluation

Although object detection can identify the approximate location of facade defects, further quantitative analysis is required for engineering applications. Therefore, an image segmentation procedure was introduced after the detection stage. The Sobel edge detection algorithm was applied to extract defect boundaries from detected regions. By calculating gradient changes in horizontal and vertical directions, the Sobel operator can effectively identify edge information and separate damaged areas from surrounding facade surfaces.

Based on the extracted boundaries, the system calculates the actual damage area and performs severity classification. According to engineering requirements, detected defects are divided into three levels: Minor damage (damage area $<30 \text{ cm}^2$), Moderate damage ($30 \text{ cm}^2 \leq \text{damage area} \leq 100 \text{ cm}^2$), Severe damage (damage area $>100 \text{ cm}^2$). This quantitative evaluation process converts image-based detection results into measurable engineering indicators, improving the objectivity and usability of inspection results. A separate segmentation stage is necessary because a bounding box is only an approximate spatial envelope and normally contains undamaged surrounding pixels.

Treating the complete rectangular box as the damaged area would systematically overestimate the defect extent. Edge-based processing therefore provides a more detailed representation of the visible contour inside the detected region. The Sobel operator is computationally lightweight and is compatible with the practical emphasis of the system, although its output remains sensitive to texture and image quality. Quantification also performs an important translation function between computer vision and engineering interpretation.

Pixel-level information is not in itself a maintenance decision. Converting a visible contour into a measurable area and then into a severity category provides an output that can be incorporated into inspection reports and prioritization procedures. However, physical-area estimates should only be reported when adequate

calibration and distance information are available. This distinction is maintained in the results section, where a representative segmented area is retained in pixels because the supplied record does not contain complete calibration parameters.

Visualisation and inspection report generation

To enhance practical applicability, the proposed system integrates defect categories, spatial coordinates, damage severity levels, and visualisation data. The application layer automatically generates inspection reports containing defect locations, types, severity classifications, and corresponding images. Furthermore, the system supports BIM-based visualisation, which allows detected defects to be associated with specific building components. This functionality enables engineers to analyse damage distribution more efficiently and provides support for future maintenance planning.

By integrating UAV sensing technology, deep learning algorithms, and engineering evaluation methods, the proposed framework establishes a comprehensive intelligent inspection workflow. The system offers a practical approach for transforming traditional experience-based facade inspection into data-driven building maintenance management.

Results

The evaluation was conducted in two stages. The baseline detector was assessed on the 8.10 Train Dataset, which contained 3,700 training images, 177 validation images, and 100 test images across eight facade-defect classes. The optimized detector and the integrated unmanned aerial vehicle workflow were then evaluated using 1,000 labeled images in laboratory tests and at two pilot buildings: a 100 m residential building and a 180 m commercial complex. The system-level campaign lasted 120 hours and produced 320 valid test records.

Baseline test-set measurements and final system-validation measurements are reported separately because they used different sample sizes and evaluation settings.

Model selection and training performance

The model-selection experiment identified YOLOv5l as the most suitable accuracy-speed compromise for the facade inspection task. The baseline YOLOv5l model achieved an mAP@0.5 of 65.9% at 36.765 frames/s. The larger YOLOv5x variant operated at 19.531 frames/s, whereas YOLOv5m reached 58.140 frames/s but produced lower overall detection quality. YOLOv5l therefore provided sufficient throughput for field deployment while retaining the feature capacity required for small and visually similar defects. Following input augmentation, adaptive anchor calculation, CIoU-based box regression, attention enhancement, and deployment-oriented optimisation, the mAP@0.5 increased from 65.9% to 76.3%, a gain of 10.4 percentage points. The final integrated evaluation recorded an overall recognition accuracy of 95.8%, a recall of 92.5%, and an inference rate of 32 frames/s. The lower deployed throughput relative to the isolated baseline measurement reflects the additional preprocessing and system integration included in the final workflow.

Detection performance across defect classes

On the 100-image baseline test set, which contained 238 labeled defect instances, the detector achieved a precision of 69.4%, a recall of 65.5%, an mAP@0.5 of 65.9%, and an mAP@0.5:0.95 of 39.2%. Performance varied substantially by defect appearance. Exposed reinforcement and tile delamination achieved the highest mAP@0.5 values, whereas tile cracks and rust stains were more difficult to separate from background texture and visually similar facade features. The baseline detection results for each defect class are summarised in Table 1.

Table 1. Baseline detection results on the 100-image test set.

Class	Instances	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
All classes	238	69.4	65.5	65.9	39.2
Crack	52	60.4	73.1	57.5	22.5
Delamination	31	76.5	61.3	70.5	37.7
Exposed reinforcement	47	88.4	95.7	95.7	70.1
Rust stain	23	53.7	52.2	47.0	20.2
Spalling	49	84.0	59.2	77.3	41.8

Class	Instances	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Tile crack	15	57.9	45.9	30.7	12.8
Tile delamination	5	89.9	80.0	93.8	82.5
Tile loss	16	44.7	56.2	54.8	26.3

Note: P and R denote precision and recall. The row totals refer to labeled instances, not image counts.

Figure 1 presents the precision-recall curves for the eight facade defect classes. As shown, these curves exhibit a consistent ordering at the class level. Tile delamination yields the most robust performance, with its curve maintaining the highest precision across nearly the entire recall range (corresponding to an AP of 0.995). In contrast, tile crack remains the weakest-performing class, with its curve dropping precipitously as recall increases (corresponding to an AP of only 0.446). This pronounced

difference in performance is consistent with two key factors: First, limited instances of tile delamination and tile cracks create a long-tailed distribution in the dataset, impairing model learning. Second, tile cracks exhibit a fine, line-like appearance in aerial imagery, which significantly increases the difficulty of feature extraction and localization. Overall, the model achieves an mAP@0.5 of 0.678 across all classes, as indicated by the bold blue curve.

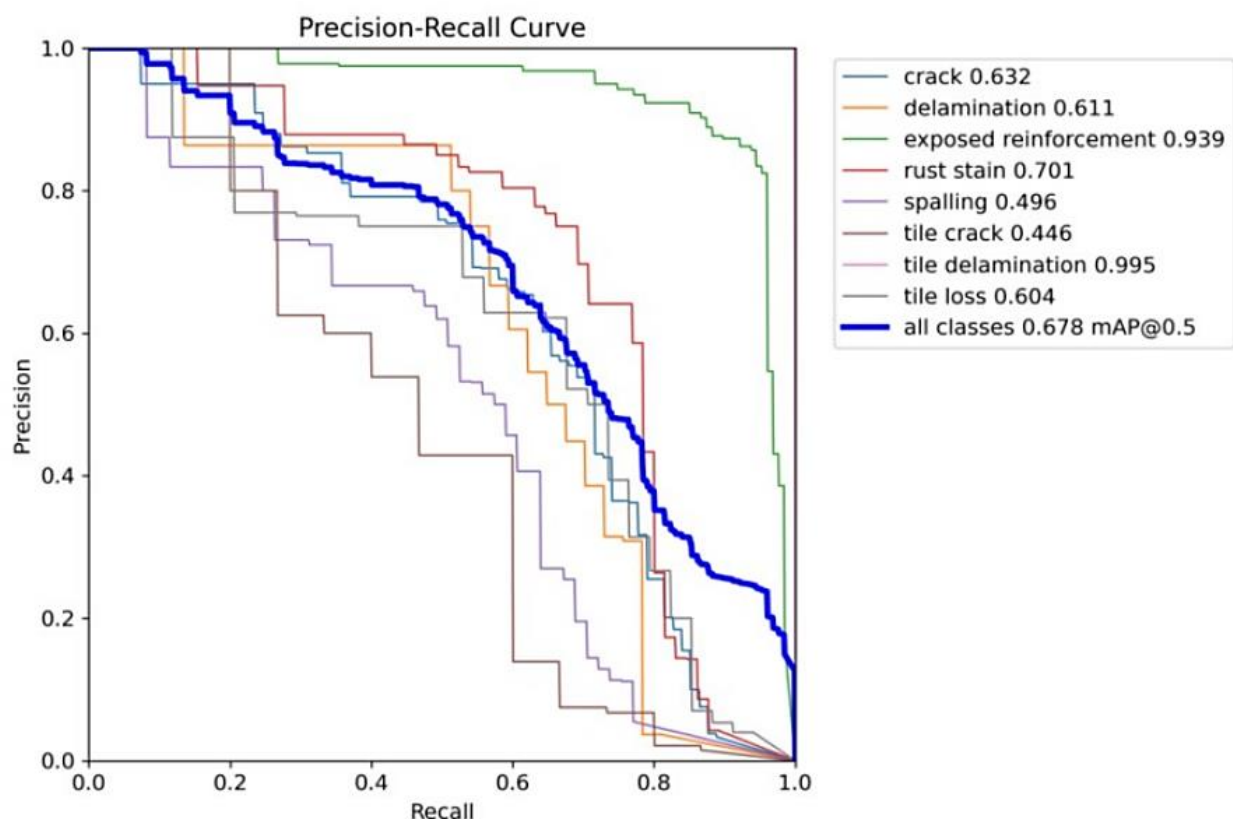


Figure 1. Precision-recall curves for the eight facade-defect classes.

To further analyse model misclassification, the normalized confusion matrix of the baseline detector is provided in Figure 2. This matrix localizes the remaining errors at the class level and reveals several distinct confusion patterns. Most notably, tile cracks were frequently assigned to the background, indicating that the model struggles to distinguish this fine, line-like defect from the surrounding intact surface. Errors also occurred

among spalling, delamination, and exposed reinforcement, as well as among tile cracks, rust stains, and tile delamination. These misclassifications correspond to classes that share similar textures or exhibit small, ambiguous boundaries at the given imaging distance. In contrast, tile delamination achieved a perfect recall of 1.00, which is consistent with its robust performance observed in Figure 1.

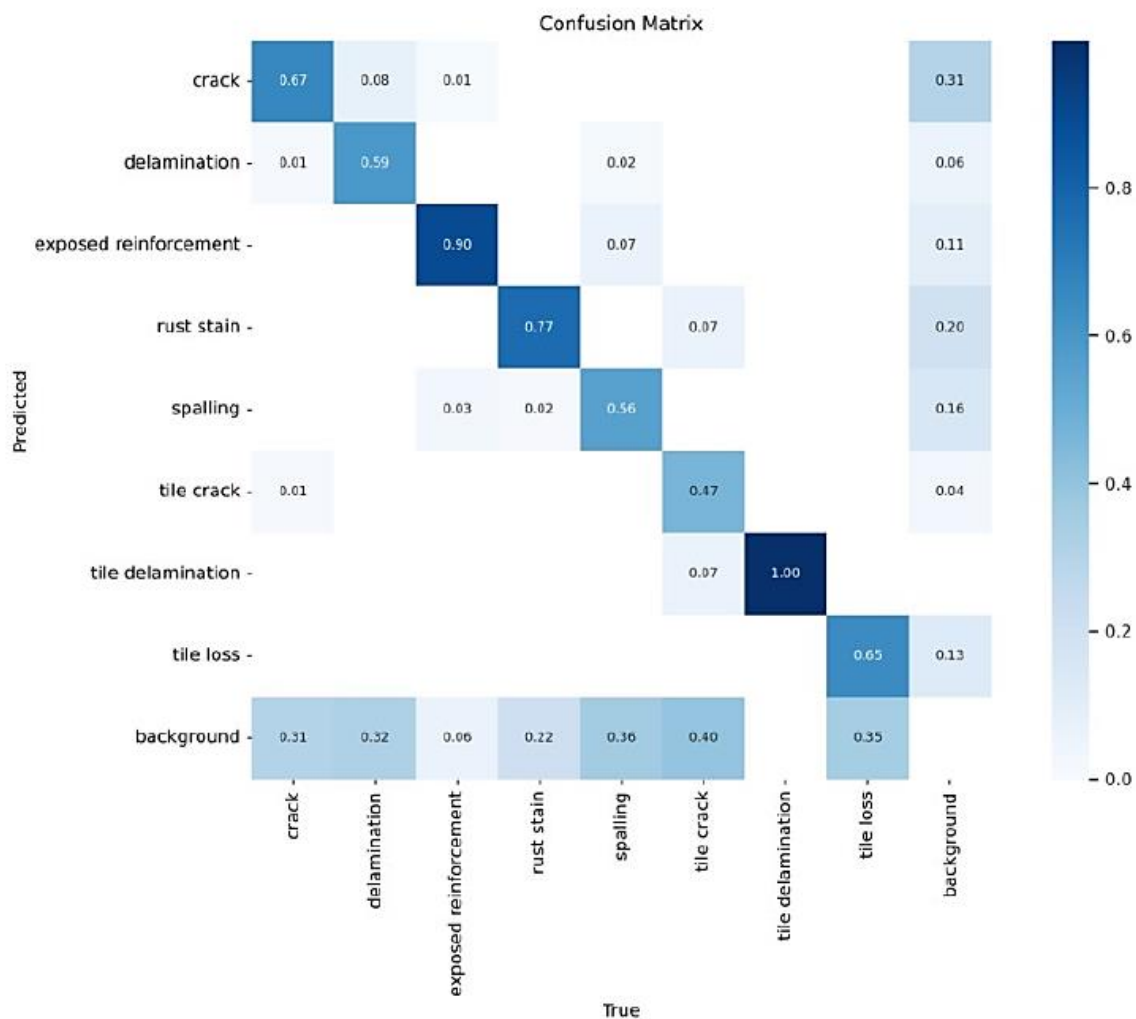


Figure 2. Normalized confusion matrix for the baseline detector.

Integrated system validation

The integrated system passed 96.4% of the 28 functional test cases, exceeding the 95.0% acceptance threshold defined in our project specification. Path coverage, abnormal-return success, multi-material acquisition, small-crack detection, and engineering grading

agreement all met their predefined criteria. The final detector also exceeded the accuracy, recall, and mAP@0.5 thresholds, while the mean three-dimensional localization error remained below 0.5 m. Detailed accuracy and functional validation results are listed in Table 2.

Table 2. Accuracy and functional validation results.

Metric	Measured result	Acceptance criterion
Functional test pass rate	96.4%	$\geq 95.0\%$
Facade path coverage	99.2%	$\geq 99.0\%$
Abnormal-return success	99.6%	$\geq 99.5\%$
Multi-material acquisition success	100.0%	100.0%
Eight-class mean recognition accuracy	95.3%	$\geq 95.0\%$
Fine crack detection below 5 cm	91.2%	$\geq 90.0\%$
Damage-grade agreement with manual review	93.0%	$\geq 92.0\%$
Final overall recognition accuracy	95.8%	$\geq 95.0\%$
Final recall	92.5%	$\geq 92.0\%$
Final mAP@0.5	76.3%	$\geq 75.0\%$
Mean three-dimensional localization error	0.42 m	≤ 0.50 m

Note: Final accuracy metrics were obtained from the 1,000-image integrated-system evaluation.

Efficiency and reliability measurements also met the project criteria. The complete inspection of a 150 m building, including image acquisition, analysis, and report generation, required 2 h 42 min. Continuous operation across three buildings lasted 8.5 h without a program crash. During that run, peak CPU and memory

utilization were 82.0% and 80.0%, and the data-loss rate was 0.3%.

Storage checks covering 10,000 images and 100 reports found no damaged files during the 30-day observation period. The full efficiency and reliability validation metrics are presented in Table 3.

Table 3. Efficiency and reliability validation results.

Metric	Measured result	Acceptance criterion
End-to-end inspection of a 150 m building	2 h 42 min	≤ 3 h
Image inference rate	32 frames/s	≥ 30 frames/s
Report generation time	45 s	≤ 60 s
Peak CPU and memory utilization	82.0% and 80.0%	$\leq 85.0\%$ each
Data-loss rate	0.3%	$\leq 0.5\%$
Single-image read time	0.15 s	≤ 0.20 s
Report export time	50 s	≤ 60 s
Mean fault-recovery time	45 s	≤ 60 s
Unsaved-data recovery rate	92.0%	$\geq 90.0\%$

The segmentation stage converted the detector output into a defect contour and pixel-area measurement. In the representative Matrix Laboratory (MATLAB) record, the segmented region contained 15,536 pixels. The supplied record did not include the complete camera calibration and distance parameters for that image; consequently, the result is retained as a pixel-area measurement rather than being reported as a physical area.

Error analysis and corrective validation

The only functional metric that initially failed its acceptance criterion was the false-positive rate for hollow areas on reflective glass curtain walls. The initial rate was 4.2%, above the required maximum of 3.0%. Inspection of the error cases showed that local overexposure from reflected light produced features similar to the target defects. A polarizing filter was added during acquisition, and Gamma correction was introduced in preprocessing.

In the subsequent test, the false-positive rate decreased to 2.1%, satisfying the criterion. After corrective validation, the system received a composite score of 91, corresponding to grade A under the project's scoring rule. In the two pilot-building comparisons, the integrated workflow increased end-to-end inspection efficiency by 82.0% relative to the recorded manual process, maintained a missed-detection rate below 2.0%, and produced result consistency above 95.0%.

Discussion and implications

Interpreting the detection performance

The proposed UAV-based inspection system achieved an overall recognition accuracy of 95.8% on an independent test set of 1,000 annotated images, with a recall of 92.5% and a mean average precision at IoU=0.5 (mAP@0.5) of 76.3%, while sustaining a real-time inference speed of 32 frames per second on an on-board edge-computing terminal. These figures should not be interpreted in isolation. The deliberate selection of the YOLOv5l backbone, rather than the larger YOLOv5x variant, reflects a conscious engineering trade-off between accuracy and deployability. Although YOLOv5x marginally improved precision and mAP, its throughput of only 19.5 fps is incompatible with near-miss inspection flights, where a blurred or dropped frame could result in a defect being missed entirely.

Consequently, the intermediate configuration offered by YOLOv5l represents the practical optimum for airborne deployment, where network depth and width must be balanced against the latency constraints of a mobile vision pipeline. Three design choices underpin the robustness reflected in these metrics. First, the Complete-IoU (CIoU) loss function explicitly accounts for overlap, center distance, and aspect ratio, thereby improving the localization of thin and elongated targets such as cracks, which standard IoU regression tends to

underfit. Second, the embedded Convolutional Block Attention Module (CBAM) re-weights feature channels and spatial maps, directing the network's focus toward regions of texture discontinuity rather than uniform background. This is particularly valuable for detecting delamination defects, which exhibit subtle visual signatures.

Third, adaptive anchor clustering and letterbox-style minimum-black-border padding eliminated a significant portion of redundant padding computation, contributing approximately a 37.0% increase in speed without altering the receptive field. The sub-meter three-dimensional localization error of 0.42 m, achieved by fusing Real-Time Kinematic (RTK) telemetry with a pixel-to-physical-scale model, is sufficient to direct maintenance crews to a specific facade panel rather than a vaguely described section of the building elevation.

The residual error budget is dominated by long-tail classes

A per-class analysis is arguably the most informative aspect of the results. The precision-recall curves reveal a pronounced performance disparity: Tile delamination was detected with near-perfect accuracy (AP=0.995), whereas tile cracks constituted the weakest performing class (AP=0.446). The confusion matrix clarifies the nature of the remaining errors. A significant proportion of missed tile cracks were simply classified as background because, when viewed from a drone at a distance of several tens of meters, a fine crack resolves to only a handful of pixels and becomes visually indistinguishable from the regular grout joints.

Therefore, accuracy is limited primarily by the available data, not by network capacity. Simply scaling the model would yield diminishing returns. Conversely, the system implements a safety-priority resource allocation strategy, directing the majority of computational resources toward high-risk defect classes. For example, serious hollow or delaminated areas achieved a 98.0% recognition rate, while minor defects are screened more leniently. This asymmetric weighting aligns the algorithm's behavior with the core engineering objective: Failing to detect a severe defect is far more costly than occasionally flagging a minor one. The residual 91.2% recognition rate for sub-5 cm cracks should be regarded as a known, bounded risk that can be mitigated with additional data collection, rather than as an unresolved failure of the

system.

From detection to a closed-loop inspection workflow

A central contribution of this work is not the detector alone, but its connection to civil-engineering decision-making. The pipeline transforms a bounding box into an engineering assessment. Detected regions are segmented using a Sobel edge operator. Their areas are converted to real-world units via RTK geometry and are then mapped onto a three-level scale - minor (below 30 cm²), moderate (30-100 cm²), and severe (above 100 cm²) - which mirrors conventional defect-grading practice. The grading agreed with manual assessment in 93.0% of cases, and the generated reports already comply with relevant Chinese facade standards, so they can be uploaded directly for construction-quality filing. Overlaying the graded defects on a BIM-based thermal map, together with automatic report generation in roughly 45 seconds per building, completes the "acquisition - recognition - localization - assessment - maintenance" loop that manual and infrared inspections have historically failed to close.

The field results quantify the operational value. For a 200-m super-high-rise building, the full inspection time fell from approximately 72 person-hours of rope-access work to 2.5 hours, a 96.5% reduction, and the per-building cost dropped from about RMB 68,000 to RMB 12,000 - an 82.0% saving that removes the economic barrier to routine inspection of small residential compounds. Across the deployed sites, 326 hazards were identified, 17 of them high-risk, and the rectification rate in the supervised district rose from 78.0% to 99.0%. Because no personnel are suspended outside the facade, the system simultaneously eliminates the roughly 1.2% accident rate associated with conventional high-altitude inspection and brings the missed-detection rate below 2.0%.

The implication is that the hardest problem in facade safety is no longer technical feasibility but the institutional adoption of a data-driven, auditable workflow. The closed-loop structure is important because conventional inspection often loses information between observation and maintenance. Photographs and free-text descriptions may document that a defect exists, but they do not always provide a machine-readable relationship among defect type, severity, image evidence, spatial position, and building component. A structured

record can preserve these relationships and can be compared with later inspections.

Repeated observations could eventually reveal whether a crack is stable or expanding, whether delamination is spreading, or whether a repaired area is deteriorating again. The present study does not model longitudinal deterioration, but its georeferenced data structure provides a foundation for that transition. BIM integration also changes the role of the inspection report. Rather than functioning only as a static document produced after a field visit, inspection information can become part of the building's digital information environment.

Associating defects with components can support traceability across inspection, maintenance, and reinspection cycles. This does not remove the need for professional engineering judgement. Ambiguous defects, unusual materials, and high-risk cases still require human review. The more appropriate role of automation is to reduce repetitive acquisition and screening work, standardize the initial representation of defects, and direct expert attention toward cases where professional interpretation has the greatest value.

Positioning against conventional and competing methods

Relative to the status quo, the advantages are threefold. Manual visual inspection and hammer tapping are strongly experience-dependent: The inter-rater disagreement on hollow-sound interpretation can reach 30.0%, and a single 150 m building requires three inspectors and three days. Infrared thermography is capable of revealing subsurface defects, yet it is constrained by solar loading and ambient conditions [20]. Its detectability reportedly drops by more than 40.0% on overcast days, and the technique also incurs high hardware and mobilisation costs. The proposed visible-light UAV platform trades the ability to see deeply subsurface voids for speed, coverage, and all-weather stability, and the vertical-layered flight with 10.0% forward overlap avoids blind zones on setback corners and curved curtain walls that scaffolds cannot reach.

Compared with the two dominant categories of competing UAV-plus-AI products - purpose-built, domain-locked inspection systems and generic off-the-shelf vision detectors - the proposed platform sits between them. Domain-specific systems are highly optimised for a single asset class but cannot transfer

across materials or building types; generic vision tools, while flexible, treat every facade as a natural image and rarely emit engineering-grade severity grading. By contrast, the vertical-layered flight protocol, the material-specific anchor priors, and the standards-linked grading make the system both transferable across tile, concrete, stone, and glass surfaces and interpretable for inspection engineers. Whether this middle position can be defended as detectors commoditise is an open strategic question; the answer probably lies in accumulated, georeferenced defect data and regulatory acceptance rather than in the detector itself.

Interdisciplinary implications

Finally, the project offers a methodological lesson for "artificial intelligence plus civil engineering". Purely generic detectors, trained on internet-scale but engineering-agnostic data, frequently miss material-specific defects because they have no notion of why a defect matters. Here, civil-engineering knowledge was injected at several points: Anchor priors were adapted to the characteristic shape of tile hollowing and concrete cracks, the attention mechanism was tuned to strengthen hollow-spot features, and the output thresholds were anchored to existing grading standards. The resulting system does not merely produce boxes; it produces a language that inspectors and regulators already understand.

The openly released 3,982-image, eight-class dataset - downloaded more than 500 times on Roboflow - is intended to lower this barrier for other teams and to support the broader shift from experience-driven to data-driven facade management.

Conclusion

This study presents an integrated, UAV-mounted, deep-learning system for the automated inspection of building facade damage. By combining a Matrice-class industrial drone with RTK positioning, a Clou - and CBAM-enhanced YOLOv5l detector, Sobel-based defect segmentation, and a BIM-assisted reporting module, the system recognizes eight categories of facade defects with 95.8% accuracy and a below-2.0% missed-detection rate. It completes the inspection of a 150 m building in under three hours and reduces inspection costs by approximately 82.0% while eliminating the need for workers to be exposed to high altitudes. Deployed across

super-high-rise residential compounds, metro entrances, and heritage buildings, it shifts facade management from reactive, subjective, and dangerously slow manual work toward a closed-loop, data-driven, auditable workflow. This work demonstrates that the value of computer vision in construction lies not in standalone benchmark scores, but in the tight coupling of model design, engineering standards, and field operations.

Limitations

Several limitations qualify these conclusions. First, class-wise performance is uneven: The weakest category, tile crack (AP = 0.446), remains constrained by long-tail sampling and the small pixel footprint of hairline cracks at flight altitude. Therefore, the headline accuracy should not be generalized to rare, fine defects without additional training data. Second, the visible-light sensor has an intrinsic physical limitation: It observes surface signatures and cannot reliably detect fully internal delamination or voids behind interfaces, which are the targets of infrared or ultrasonic methods. Third, although the glare-induced false-alarm rate on glass curtain walls was reduced from 4.2% to 2.1% by adding a polarizing filter and Gamma correction, specular overexposure is not yet fully eliminated.

Fourth, Sobel edge segmentation is a classical, noise-sensitive operator; measured defect areas depend on accurate RTK ranging and remain approximate on highly textured facades. Fifth, the evidence base is geographically and climatically narrow. The curated corpus of 3,982 images covers mainly tile, concrete, and glass facades in two Chinese provinces and has not been validated across different seismic zones, climates, or architectural traditions. Thus, generalizability is assumed rather than proven. Sixth, extreme operating conditions - strong winds, rain, fog, night flights, and electromagnetic-interference environments - were only partially tested. The reported comparisons rely largely on vendor benchmarks and pilot deployments rather than on a prospectively blinded trial against certified thermographers.

Finally, routine adoption still depends on airspace approval, wind-field stability near tall structures, and evolving liability and regulatory frameworks. These caveats should temper the optimism suggested by the headline metrics. An additional methodological limitation concerns the interpretation of the four

hypotheses (H_1 - H_4). They are evaluated primarily through engineering performance indicators and integrated system tests rather than through a statistical causal model.

The results demonstrate that the corresponding sensing, recognition, quantification, and integration functions operate effectively within the proposed framework. However, they do not independently estimate the causal effect size of each capability on overall inspection performance. Future controlled experiments could isolate these contributions more rigorously. Furthermore, several system-level outcomes are derived from pilot deployments rather than long-term operational observation. Short-term gains in inspection efficiency and defect identification do not, by themselves, establish long-term reductions in lifecycle maintenance costs or facade-related incidents.

Such claims would require repeated inspection cycles, maintenance records, and comparable control cases over a longer period. Therefore, the present findings should be interpreted primarily as evidence of technical and operational feasibility within the tested scenarios.

Future work

Three specific directions are proposed. First is multi-modal sensing: Fusing RGB imagery with Light Detection and Ranging (LiDAR) point clouds and long-wave infrared thermography could bridge the subsurface-defect gap that visible light alone cannot address. This would enable the direct detection of hollow delamination and moisture ingress, rather than relying on inference. Second, model evolution involves advancing beyond YOLOv5 towards transformer-based or multi-task detectors capable of joint detection and segmentation. When combined with class-balanced sampling and hard-example mining for tail categories such as tile cracks, this approach is expected to improve the lowest average precision (AP) scores without compromising edge throughput. Third is a shift toward predictive rather than merely descriptive analytics.

By accumulating time-stamped, georeferenced defect records and linking them to BIM geometry, a digital-twin platform could model damage progression, forecast maintenance intervals, and transition owners from scheduled to condition-based maintenance. Meanwhile, on-board edge inference would allow results to be generated during the flight itself. Future development

should also incorporate uncertainty-aware inspection. Detector confidence alone does not equate to engineering risk. A decision-support layer could integrate recognition confidence, defect severity, image quality, and the consequence of missed detection to determine whether a result can be accepted automatically or should be referred for manual review.

This human-in-the-loop strategy would be especially valuable for rare classes and safety-critical facade components. Longitudinal inspection is another important direction. Repeated UAV surveys registered to a common BIM or digital-twin coordinate system could distinguish newly emerging defects from existing ones and measure changes in crack length, damaged area, or severity over time. Dataset development should proceed in parallel by increasing the representation of rare categories and difficult negative samples such as joints, shadows, stains, reflections, and decorative textures.

Finally, interoperability deserves greater attention. Consistent coordinate conventions, defect definitions, severity criteria, and reporting schemas would allow inspection records generated by different UAV platforms and models to be integrated into building-management systems rather than remaining isolated project files. Beyond facades, the same acquisition-assessment loop transfers naturally to bridges, tunnels, hydraulic structures, and rapid post-disaster damage assessment, where speed and safe access matter most. The ambition is therefore not a single product but a standards-based ecosystem: an open, annotated, multi-modal defect corpus; lightweight detectors that run on commodity inspection drones; and reporting schemas that plug directly into existing building-safety regulations.

Only when algorithm, sensor, workflow, and regulation mature together can UAV-based facade inspection become a routine public-safety service rather than a specialized demonstration.

Funding

This work was supported by the National Undergraduate Training Program on Innovation and Entrepreneurship (Grant No: 202610345040).

Author Contributions

Conceptualization, Wenxuan Tang; methodology, Wenxuan Tang and Hongqi Zhang; software, Wenxuan Tang; formal analysis, Hongqi Zhang; investigation,

Mengqi Li and Shuyu Yang; data curation, Mengqi Li; validation, Shuyu Yang; visualisation, Hongqi Zhang; writing - original draft preparation, Wenxuan Tang; writing - review and editing, Hongqi Zhang and Furong Liu; project administration, Wenxuan Tang; funding acquisition, Furong Liu. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

The authors would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Puliti, M., Montaggioli, G., Sabato, A. (2021) Automated subsurface defects' detection using point cloud reconstruction from infrared images. *Automation in Construction*, 129, 103829.
- [2] Bosché, F., Ahmed, M., Turkan, Y., Haas, C. T., Haas, R. (2015) The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: the case of cylindrical MEP components. *Automation in Construction*, 49, 201-213.
- [3] Ellenberg, A., Kotsos, A., Moon, F., Bartoli, I. (2016) Bridge related damage quantification using unmanned aerial vehicle imagery. *Structural Control and Health Monitoring*, 23(9), 1168-1179.
- [4] 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.
- [5] Lyu, C., Lin, S., Lynch, A., Zou, Y., Liarokapis, M. (2025) UAV-based deep learning applications for automated inspection of civil infrastructure. *Automation in Construction*, 177, 106285.
- [6] Liang, H., Lee, S. C., Bae, W., Kim, J., Seo, S. (2023) Towards UAVs in construction: advancements, challenges, and future directions for monitoring and inspection. *Drones*, 7(3), 202.
- [7] Dong, X., Wang, Q., Deng, H., Yang, Z., Ruan, W., Liu, W. (2025) From global to hybrid: a review of supervised deep learning for 2-D image feature representation. *IEEE Transactions on Artificial Intelligence*, 6(6), 1540-1560.

- [8] Shafiq, M., Gu, Z. (2022) Deep residual learning for image recognition: a survey. *Applied Sciences*, 12(18), 8972.
- [9] Cha, Y. J., Choi, W., Büyüköztürk, O. (2017) Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5), 361-378.
- [10] Chen, C., Zheng, Z., Xu, T., Guo, S., Feng, S., Yao, W., Lan, Y. (2023) Yolo-based uav technology: a review of the research and its applications. *Drones*, 7(3), 190.
- [11] Kim, H., Lamichhane, N., Kim, C., Shrestha, R. (2023) Innovations in building diagnostics and condition monitoring: a comprehensive review of infrared thermography applications. *Buildings*, 13(11), 2829.
- [12] LeCun, Y., Bengio, Y., Hinton, G. (2015) Deep learning. *Nature*, 521(7553), 436-444.
- [13] Pham, M. T., Courtrai, L., Friguet, C., Lefèvre, S., Baussard, A. (2020) YOLO-Fine: One-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sensing*, 12(15), 2501.
- [14] Kim, B., Yuvaraj, N., Park, H. W., Preethaa, K. S., Pandian, R. A., Lee, D. E. (2021) Investigation of steel frame damage based on computer vision and deep learning. *Automation in Construction*, 132, 103941.
- [15] Zhang, J., Qian, S., Tan, C. (2022) Automated bridge surface crack detection and segmentation using computer vision-based deep learning model. *Engineering Applications of Artificial Intelligence*, 115, 105225.
- [16] Kumar, T., Brennan, R., Mileo, A., Bendeche, M. (2024) Image data augmentation approaches: a comprehensive survey and future directions. *IEEE Access*, 12, 187536-187571.
- [17] Vijayakumar, A., Vairavasundaram, S. (2024) Yolo-based object detection models: a review and its applications. *Multimedia Tools and Applications*, 83(35), 83535-83574.
- [18] Zhang, Y., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T. (2022) Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing*, 506, 146-157.
- [19] Ren, W., Zhong, Z. (2025) Building construction crack detection with BCCD YOLO enhanced feature fusion and attention mechanisms. *Scientific Reports*, 15(1), 23167.
- [20] Ibrahim, A., Faris, N., Zayed, T., Qureshi, A. H., Abdelkhalek, S., Abdelkader, E. M. (2026) Application of infrared thermography in concrete bridge deck inspection: current practices, challenges and future needs. *Nondestructive Testing and Evaluation*, 41(1), 1-44.