UAV-Based Structural Health Monitoring Using a Two-Stage CNN Model with Lighthouse Localization in GNSS-Denied Environments

Timothy Scott C. Chu [1*] , Joses Sorilla [1], Alvin Y. Chua [1]

[1] Department of Mechanical Engineering, De La Salle University, 2401 Taft Ave. Malate, Manila, Philippines.

Abstract

This study presents a UAV-based Structural Health Monitoring (SHM) system that combines Lighthouse localization with a two-stage CNN architecture—AlexNet for crack classification and YOLOv4 for segmentation—to enable reliable crack detection and spatial mapping in GNSS-denied environments. This study explores the effectiveness of this combination as a practical and computationally efficient solution for indoor SHM tasks. The UAV was deployed within a 1.5 m × 1.2 m × 1.2 m test volume to inspect synthetic cracks derived from Özgenel's dataset, as well as a real-world wall crack. Two experiments were conducted: evaluating UAV localization accuracy and assessing the system's ability to detect cracks and provide corresponding pose data. The system achieved a 1–2 cm margin of error in pose estimation, alongside 100% precision, 83.33% recall, and 91.89% accuracy in crack detection. This level of localization accuracy supports stable autonomous UAV flight and ensures that cracks are detected and spatially localized with minimal deviation. Beyond classification and segmentation, the system returns pose data tied to each detected crack, allowing users to identify defect locations precisely and use this information to guide inspection or maintenance tasks. Future work includes expanding the dataset, generalization, and evaluating scalability via multi-base station setups.

Keywords: Crack Detection; Crack Segmentation; GNSS-Denied Environments; Lighthouse Localization; Structural Health Monitoring; Two-Stage CNN Model; Unmanned Aerial Vehicles.

1. Introduction

Structural Health Monitoring (SHM) is a systematic approach to assessing the integrity of infrastructures, ensuring public safety, and extending their service life. Modern SHM relies on *Non-Destructive Testing* (NDT) techniques, which evaluate structural conditions without causing damage [1]. Among these, *Unmanned Aerial Vehicles* (UAVs) have emerged as a viable inspection tool, offering a safer, more efficient alternative to manual inspections that traditionally require personnel to ascend to hazardous heights [2, 3]. UAVs enable rapid data acquisition by maneuvering complex structures [4], facilitating real-time or near-real-time defect identification through computer vision-based techniques [5]. A critical aspect of structural assessment is the detection of hairline cracks, which often serve as early indicators of fatigue, stress accumulation, and potential failure [6]. If left undetected, these fractures can propagate over time, compromising the stability of load-bearing structures. Traditional inspection methods struggle with consistently and reliably identifying hairline cracks, particularly in large-scale or hard-to-access structures [7]. By leveraging advanced image processing and deep learning models, UAV-based SHM systems significantly improve early-stage damage detection, enabling timely intervention before structural failures occur.

Effective UAV-based SHM systems rely on two fundamental components: localization and computer vision models. Localization ensures precise UAV positioning, critical for accurate flight operation, crack mapping, and comprehensive inspection coverage, especially in *Global Navigation Satellite System* (GNSS)-denied environments such as tunnels, indoor facilities, and under-bridge structures [8]. UAVs deviate from planned flight paths without a reliable localization system, reducing inspection efficiency and compromising crack tracking accuracy. Researchers have explored various localization techniques to address this, balancing cost, accuracy, and complexity. These techniques include *Radio-Frequency Identification* (RFID), *Ultra-Wideband* (UWB), vision-based tracking, and the emerging Lighthouse localization approach. With the availability of hybrid-localization techniques that integrate multiple positioning systems for improved accuracy and robustness, they introduce greater system complexity, computational overhead, and hardware dependencies—factors that may be unnecessary for this application. Lighthouse localization, however, satisfies the operational requirements for UAV control in GNSS-denied environments while achieving sub-centimeter accuracy. Complementing localization, computer vision models—particularly deep learning-based techniques— automate crack detection with high precision. *Convolutional Neural Networks* (CNNs), including AlexNet, VGG-19, ResNet, and YOLO, have demonstrated effectiveness in crack identification [9], with two-stage CNN transfer learning approaches further enhancing classification and segmentation accuracy [10]. Our previous work [11] has demonstrated the efficacy of a two-stage CNN model that leverages transfer learning in UAV-based crack detection with strong performance in classification and segmentation tasks.

In light of the critical importance of a localization method's ability to generate accurate, precise, and application-appropriate data, evaluating existing techniques for their suitability in UAV-based crack detection within GNSS-denied environments is essential. Improper or low-quality localization can lead to operational drift, where the UAV deviates significantly from its intended path or exhibits unstable movements in response to perceived position errors. This erratic behavior compromises the spatial reference needed for accurate crack mapping. It also introduces safety risks, as abrupt course corrections may cause the UAV to collide with structural elements, damage itself, or endanger nearby personnel. Moreover, logistical constraints such as weight, cost, and setup feasibility often render conventional high-precision systems impractical in real-world deployment. The challenge this study aims to address is the localization of hairline cracks using a system that not only achieves sufficient positional accuracy but also remains lightweight, scalable, and computationally efficient.

RFID localization can be categorized into Antenna Array-Based and *Synthetic Aperture Radar* (SAR)-Based approaches. While the former relies on bulky hardware composed of multiple RFID reader antennas surrounding passive tags, rendering it impractical for UAV deployment, the latter enables real-time tracking using a mobile RFID reader traversing an array of passive tags, albeit with significant infrastructure requirements [12]. Recent developments in SAR-based RFID localization have demonstrated steady improvements in accuracy and efficiency. In a foundational study [13], the proponents flew a UAV equipped with a UHF-RFID reader over a football field embedded with fixed RFID tags, and the resulting RFID-based position estimates exhibited a *Mean Absolute Error* (MAE) of 27.1 cm when compared to GNSS ground truth. To address computational inefficiencies in this setup, subsequent work [14] developed a hybrid method that utilized *Particle Swarm Optimization* (PSO) to significantly reduce processing time while improving accuracy to 15–20 cm. Further advancements have explored more hybrid methods; for instance, Böller et al. [15] incorporated *Angle of Arrival* (AoA) and *Time of Flight* (ToF) measurements, reducing localization error to 12.8 cm. While promising in structured environments such as warehouses, where fixed tag arrays are acceptable, RFID localization remains unsuitable for UAV-based SHM. As noted in Buffi et al. [16] study, its dependence on fixed infrastructure and moderate spatial resolution limits its use in dynamic UAV-based SHM contexts.

UWB localization estimates position using short-duration radio pulses exchanged between fixed anchors and mobile tags, relying on ToF and *Time-of-Arrival* (ToA) measurements to achieve high spatial accuracy [17]. In structured environments such as warehouses, UWB has enabled coordinated multi-UAV operations, with one study [18] demonstrating its application in a warehouse management system using eight fixed anchors to define the flight volume. The system achieved a mean localization error of 18.2 cm, preventing UAV collisions during path planning. In industrial scenarios, UWB has demonstrated robustness under asymmetrical anchor configurations; even under these conditions, the system achieved a localization error below 20 ± 7 cm, performing comparably to vision- and LiDAR-based systems in low-visibility environments [19]. Mobile UWB anchors mounted on *Unmanned Ground Vehicles* (UGV) have been explored, enabling UAVs to localize relative to a moving reference and supporting multi-agent system configurations. These setups have reported sub-centimeter accuracy, with more than 50% of position estimates falling below 10 cm of error [20]. However, observations suggest that UWB performance remains sensitive to anchor geometry and environmental conditions. Sorescu et al. [21] evaluated four configurations of anchor placements, showing that localization error increased from 6.95 cm in a compact $3 \times 3 \times 1$ m setup to 9.34 cm in a more elongated $3 \times 12 \times 1$ m space. This sensitivity underscores the need for careful calibration to mitigate multipath interference and signal degradation, particularly in sparse or cluttered deployments.

Vision-based tracking offers high-precision localization but requires substantial infrastructure and controlled environmental conditions. In Preiss et al. [22] study, a vision system utilizing 24 Vicon Vantage cameras achieved 1.5-2 cm accuracy while tracking 49 UAVs within a confined 6 m × 6 m × 3 m space. Marker-based tracking has also shown

promise; Mu et al. [23] localized a DJI Tello Ryze drone for precision docking using a modified YOLO algorithm, attaining a localization error of 1.03 cm, including successful trials on a moving UGV platform. Despite these results, such systems depend heavily on the system setup, structure settings, and stable lighting. To overcome this, onboard camera-based localization methods, such as *Visual Simultaneous Localization and Mapping* (vSLAM), have gained traction. However, as noted in Murhij et al. [24] study, integrating deep learning into UAV visual localization remains a challenge due to concerns on real-time processing and memory constraints. To enhance robustness under weak lighting, Wu et al. [25] proposed a hybrid Visual SLAM system incorporating ORB-SLAM3, GAN networks, and YOLOv5, achieving an average error of 1.0077 cm, comparable to baseline and hybrid SLAM models. This performance came at the cost of significant computational resources. These constraints highlight a key limitation of vision-based localization: despite their accuracy, the requirements for processing power, lighting stability, and structured environments reduce their practicality for lightweight, real-time UAV-based structural health monitoring. In contrast, Lighthouse localization presents a promising alternative that balances precision, responsiveness, and deployment simplicity, particularly in GNSS-denied environments.

Lighthouse localization offers a cost-effective, high-precision alternative for UAV tracking, particularly in indoor navigation and autonomous landing. Unlike vision-based systems, lighthouse systems provide low-latency, high-accuracy tracking with minimal computational demands. They are independent of ambient lighting conditions, performing comparably to or surpassing UWB-based localization in precision and responsiveness. Studies have demonstrated its feasibility for UGVs, offering low-cost, high-precision tracking in structured environments [26]. In industrial automation, lighthouse localization has achieved sub-millimeter accuracy for product tracking, though its full 3D tracking capabilities remain unexplored [27]. For UAV positioning, Greiff et al. [28] validated its sub-centimeter accuracy through simulations, presenting it as an affordable alternative to motion capture and UWB tracking. Furthermore, Martin et al. [29] applied lighthouse localization for UAV docking, achieving an average landing error of 3.5 cm, outperforming ultrasonic and optical flow sensors in indoor precision landing applications. Table 1 presents a comparative analysis of each localization technique, highlighting the advantages and limitations of various localization methods for UAV-based applications in GNSS-denied environments.

Table 1. Localization Methods for UAV-Based Applications in GNSS-Denied Environments

Method	Accuracy	Key Benefits	Limitations
RFID [10-12]	~12–27 cm	Low-cost, suitable for warehouses	Limited range, infrastructure-dependent, low accuracy for SHM
UWB [15-18]	7–20 cm ± 7 cm	High accuracy, robust in GNSS-denied environments	Requires optimal anchor placement, susceptible to interference
Vision-Based [19, 20, 22]	1–2 cm	High precision, effective for swarm UAVs	Requires multiple cameras, controlled lighting, expensive setup
Lighthouse [26]	~ 3.5 cm	Cost-effective, low-latency, computationally efficient	Requires line-of-sight, limited range (< 7 m)

Each localization technique presents trade-offs in cost, accuracy, and complexity. RFID is suited for structured environments but lacks precision, UWB offers sub-centimeter accuracy but requires careful anchor placement, and vision-based tracking provides high precision but is infrastructure-dependent. Lighthouse localization emerges as a cost-effective alternative, delivering sub-centimeter accuracy with minimal computational demands. While it requires line-of-sight, this limitation is shared by most localization techniques, making it a viable option for autonomous UAV operations in controlled indoor environments. Despite advances in deep learning-based crack detection, precise and reliable localization remains a limiting factor in achieving fully autonomous UAV-based SHM, particularly in GNSS-denied environments. This gap is especially critical when detecting fine, hairline cracks, where even minor localization errors can undermine mapping accuracy and inspection coverage.

Thus, this study presents a novel UAV-based SHM by integrating lighthouse localization as a high-precision alternative for UAV-based inspections in GNSS-denied environments. By addressing this localization challenge, we aim to develop a fully autonomous UAV inspection system that combines state-of-the-art crack detection with precise, real-time positioning, enabling more reliable and scalable SHM solutions. The remainder of this paper is structured as follows: Section 2 outlines the theoretical framework, detailing the two-stage CNN model and the Lighthouse positioning system. Section 3 presents the methodology, including the system architecture, dataset preparation, and physical experimentation. Section 4 discusses the results, highlighting localization performance and crack detection accuracy. Finally, Section 5 concludes the study and outlines directions for future research.

2. Theoretical Considerations

This section presents the theoretical foundations of this study, focusing on the two-stage CNN model with transfer learning for crack detection and segmentation, followed by the Lighthouse positioning system for precise localization in GNSS-denied environments. These components form the framework for integrating vision-based defect identification with high-accuracy UAV positioning, ensuring reliable and scalable SHM applications.

2.1. Two-Stage CNN Model with Transfer Learning

Our previous work [11] presented a two-stage CNN model optimized through transfer learning for UAV-based SHM in GNSS-denied environments. As illustrated in Figure 1, the first stage uses AlexNet for binary crack classification, crack-positive or crack-negative, reducing computational load for the second stage, which employs YOLOv4 for precise crack boundary detection, leveraging its speed and real-time processing capabilities. Transfer learning with ImageNet weights enabled faster convergence and robust generalization. Evaluation results showed 99.42% classification accuracy and 85.71% segmentation accuracy, outperforming standalone YOLOv4 by reducing false positives while maintaining high detection rates. This approach optimized computational efficiency, allowing real-time deployment on UAVs with limited processing power.

This pairing was selected for its balance of speed and accuracy—ideal for UAVs with limited onboard resources. While deeper models like ResNet or SENet may yield higher accuracy, their computational cost makes them less suitable for real-time deployment. A comparative evaluation of traditional and two-stage CNN architectures, supporting this rationale, is summarized in our prior study.
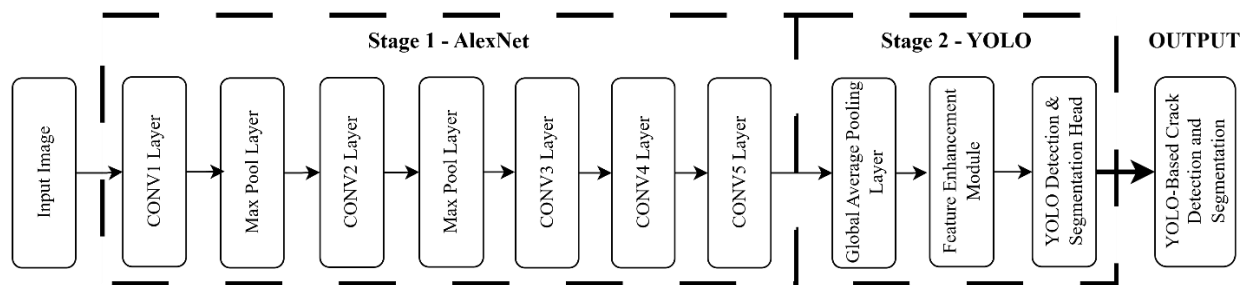

Figure 1. Architecture of the two-stage CNN model for UAV-based crack detection

2.2. Lighthouse Positioning System

Lighthouse localization is a laser-based positioning system that determines an object's position through laser sweeps from static base stations, like the SteamVR Base Station. This method offers high accuracy, low latency, and minimal computational overhead, making it suitable for UAV navigation in GNSS-denied environments. The base station emits a structured infrared laser sweep at a known angular velocity, detected by photodiode sensors mounted on the UAV. By analyzing the time difference between when the sensor receives signals from the base station and between sweeps, the system can determine sensor's position and track its movement relative to the base stations. With at least two base stations, the UAV's 3D position and orientation are estimated using triangulation and sensor fusion algorithms [30].

The objective of the Lighthouse localization system is to determine the rotation angle (α), defined as when the structured infrared light from the base station strikes the sensor. The system then employs ToF and Euclidean Distance calculations to estimate the sensor's x, y, and z coordinates in a 3D space. The predicted rotation angle (α_p) is derived from the sensor rotation angle (α_s) relative to the base station and the rotation angle from the intersection line to the sensor (α_t) that accounts for the inclination of the light plane. This relationship is defined in Equation 1, while Equation 2 provides the expression for calculating the sensor rotation angle. The variables x and y represent the measured distances along their respective axes between the UAV-mounted sensor and the base station. Figure 2a visually represents this localization process, illustrating the interaction between the sensor, the structured light plane, and the base station.

$$\alpha_p = \alpha_s - \alpha_t \tag{1}$$

$$\alpha_s = \tan^{-1}\left(\frac{y}{x}\right) \tag{2}$$

where: α – rotation angle; α_p – predicted/calculated rotation angle; α_s – measured rotation angle from the sensor; α_t – rotation angle from the intersection line to the sensor; t – tilt angle; x, y, and z – Euclidean distance between the sensor and the tag.

The tilt correction term is expressed in Equation 3. The variable r represents the Euclidean distance along the XY-plane between the base station and the sensor, while the variable d is determined using the relationship illustrated in Figure 2b. The variable t represents the tilt angle of the laser ray, depicted by the red line in Figure 2b, as it intersects with the UAV-mounted sensor.

$$\alpha_t = \sin^{-1}\left(\frac{d}{r}\right) = \sin^{-1}\left(\frac{-z \tan t}{\sqrt{x^2 + y^2}}\right) \tag{3}$$

where: t – tilt.

Substituting this into the original equation provides the final expression for the predicted rotation angle (α_p).

$$\alpha_p = \alpha_s - \alpha_t = \tan^{-1}\left(\frac{y}{x}\right) + \sin^{-1}\left(\frac{z \tan t}{\sqrt{x^2+y^2}}\right) \tag{4}$$

Thus, the lighthouse localization system can accurately determine the UAV's position relative to the base station by leveraging Euclidean distance calculations and the predicted rotation angle. The sensor's spatial orientation in 3D space is established through reference frame transformations, while ToF measurements provide precise distance estimates. These combined methodologies ensure robust and reliable real-time tracking, making the system well-suited for GNSS-denied environments.

(a)   (b)
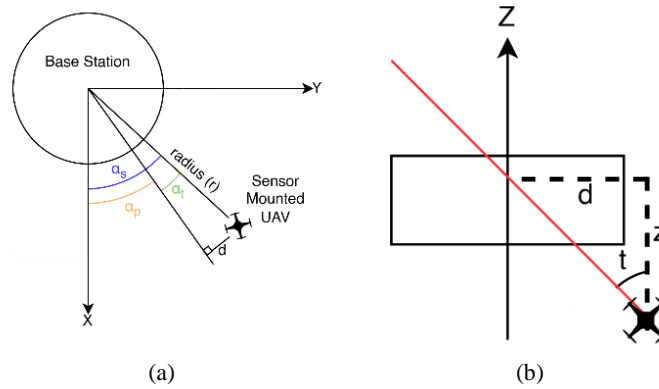
Figure 2. Lighthouse Localization for UAV Tracking, (a) Top view illustrating the sensor's predicted rotation angle (α_p), measured rotation (α_s), and tilt correction (α_t) relative to the base station. (b) Side view depicting the perpendicular distance (d) from t.

3. Methods

This section presents the methodology in three parts: System Overview, detailing the localization framework; Neural Network and Dataset, covering data acquisition and model architecture; and Physical Experimentations, describing the experimental setup and performance evaluation.

3.1. System Overview

Figure 3 illustrates the overall workflow of the crack detection and localization system. A Crazyflie micro-UAV, equipped with a Lighthouse Positioning Deck and a HIMAX HM01B0 low-power monochrome camera, was utilized for experimentation. The Lighthouse system, developed by Bitcraze, consists of base stations that define the experimental space and provide real-time positional tracking of the UAV. Within this controlled environment, printed cracked images are placed against a white background to simulate structural defects.
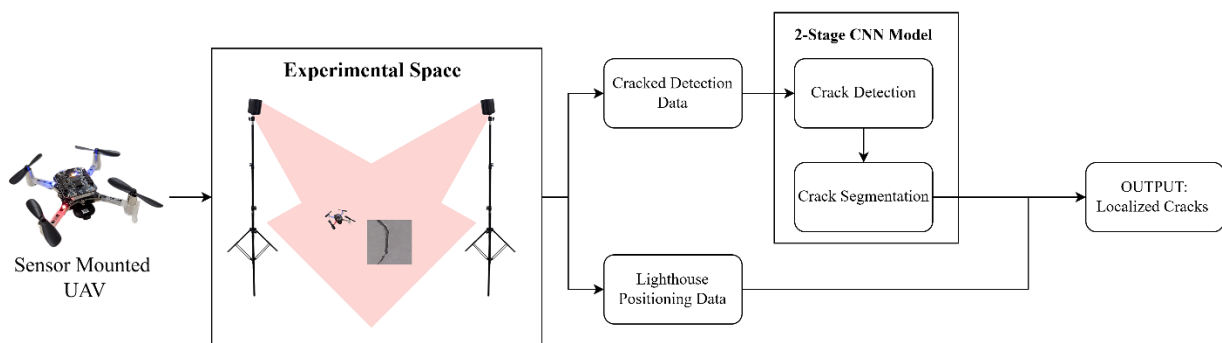
Figure 3. System Overview of UAV-Based Crack Detection and Localization

During operation, the UAV navigates the space while capturing images of the inspection area. These images are then processed through a 2-Stage CNN model to detect the presence of cracks. The Lighthouse Positioning System provides precise X, Y, and Z coordinates in meters to ensure accurate localization of the UAV and the detected cracks. The final output consists of a localized image of the detected crack, providing visual and positional information, which is crucial for UAV-based SHM applications in GNSS-denied environments.

3.2. Neural Network and Dataset

This study utilizes the same model as in our previous work [11]. The classification dataset used for model training was sourced from Özgenel et al. [31], which includes 20,000 and 20,000 non-crack images extracted from 458 high-resolution structural photos. For segmentation, 704 samples were extracted and augmented by varying orientation, position, and brightness, then manually annotated with bounding boxes. Both datasets followed a 60:20:20 train-validation-test split.. The laptop used for this study possessed a Ryzen 5 5500 processor and an RTX 3060 12GB GPU for efficient computation. To evaluate the model's ability to detect previously unseen cracks, the researchers developed synthetic defects by merging multiple cracked images, representing data beyond the training dataset. These synthetic cracks are then strategically placed, with different orientations, in the inspection area for detection. This experimentation was conducted together with the crack localization experiment. By combining these controlled tests, the study evaluated the detection performance of the system and the model's ability to map crack positions using the Lighthouse positioning system precisely. The system was further validated on actual wall cracks following controlled environment experiments to assess its viability for real-world inspections.

3.3. Physical Experimentation

The physical experimental phase consists of two primary components: the localization experiment and the crack detection experiment. The localization experiment aims to evaluate the performance of the Lighthouse positioning system after calibration. On the other hand, the crack detection experiment assesses the system's capability to detect and localize cracks within the inspection area. The experiment space, bounded by the Lighthouse positioning system, measures $1.5 \times 1.2 \times 1.2$ meters, with the base stations positioned 30 cm away from the experimental area and mounted at 1.65 meters above the ground to ensure optimal coverage. The inspection area consists of an 81.28 cm $\times$ 101.6 cm white foam board, placed at the edge of the test space, simulating a structural surface for crack detection. A visualization of the experiment setup is presented in Figure 4.
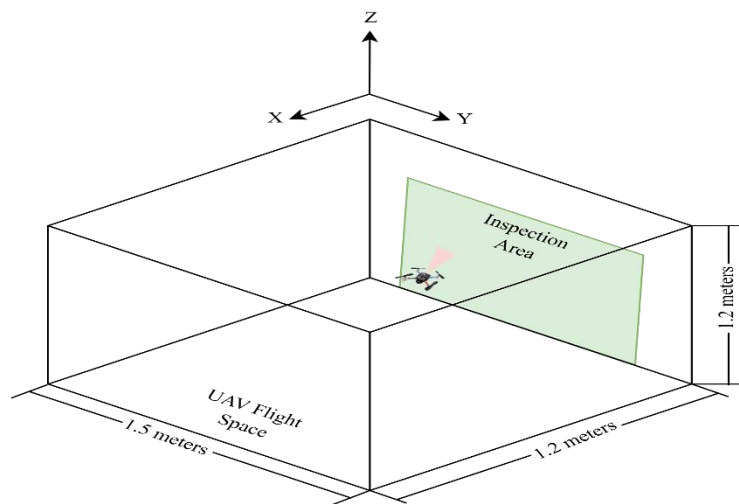

Figure 4. Experimental Volume for UAV Operation and Inspection

3.3.1. Localization Experiment

Before conducting localization tests, a calibration phase was performed to validate the accuracy of the Lighthouse positioning system. This process involved positioning the UAV at known reference points within the experimental area and comparing the system's measured coordinates with the actual positions. Following calibration, a test trajectory was executed, during which the UAV autonomously followed a predefined flight path within the experimental space. Error readings were collected throughout the trajectory to assess the system's ability to maintain the expected path. These results provided critical insights into the positioning system's accuracy, informing the subsequent crack detection experiments. As depicted in Figure 5, the predefined flight path was structured around a targeted inspection trajectory, covering an area of 55 cm $\times$ 80 cm, with 16 designated waypoints where the UAV hovered. This setup allowed for controlled movement along the Y and Z axes while assessing the UAV's ability to maintain its position along the X axis. This evaluation offers a comprehensive understanding of the Lighthouse positioning system's performance.
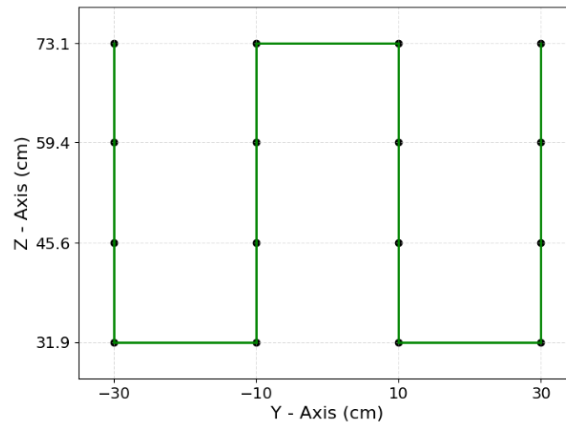
Figure 5. Predefined UAV Trajectory (in cm) for Lighthouse Localization Assessment

3.3.2. Crack Detection Experiment

The crack detection experiment began with a calibration process to determine the optimal UAV-to-target distance for image capture. Based on prior findings, this X-distance was set to 13 cm to maximize detection accuracy, a 15.5 cm by 9 cm field-of-view. The researchers experimented with three key scenarios: (1) crack localization at different positions within a 40×45 cm inspection area to assess detection performance across various regions; (2) synthetic crack scenarios, one long vertical crack (14×80 cm), one long horizontal crack (80x14 cm) and another with a long diagonal crack (60×60 cm) to evaluate the model's ability to detect larger, more complex formations; and (3) real-world crack detection, where the system was tested on actual structural. Performance was measured based on successful crack identification and localization accuracy. Three of the five total case scenarios involved synthetic cracks, one with printed crack images positioned in several areas, and one used real cracks. Figure 6 visualizes the case scenarios, where white areas indicate crack presence, and black areas represent non-crack regions.
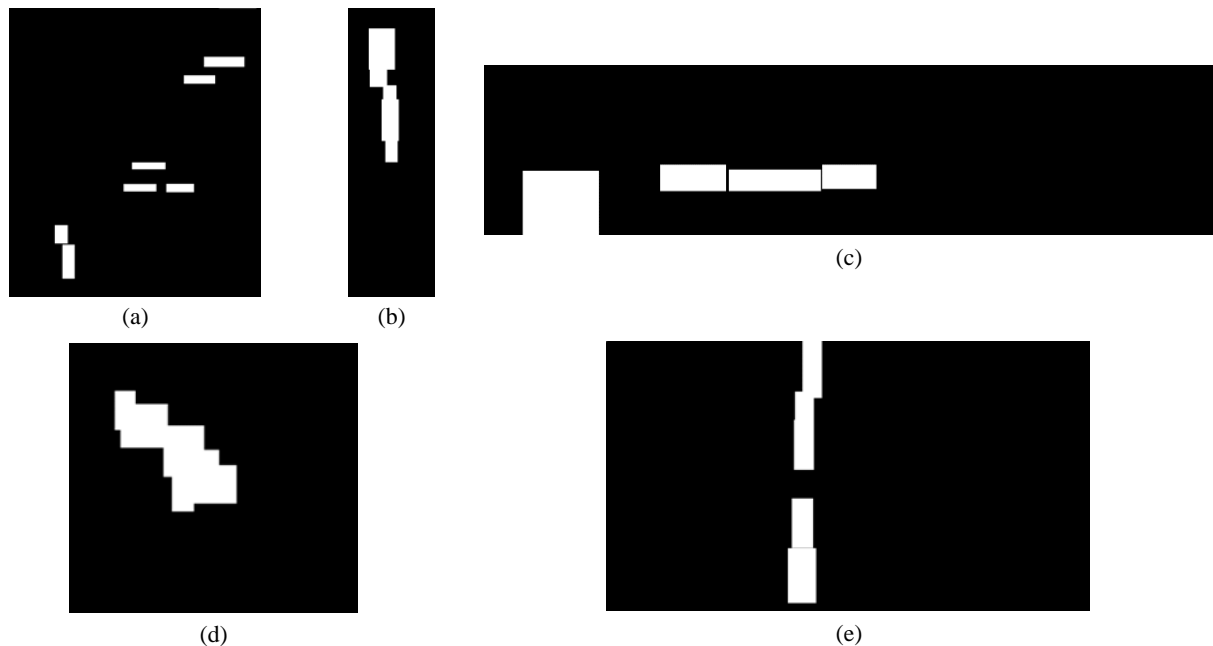
Figure 6. Crack Case Scenarios in the Inspection Area, (a) Distributed cracks across the inspection area, (b) combined vertical crack, (c) combined horizontal crack, (d) real diagonal crack, and (e) actual crack on a structural wall

4. Results and Discussion

This section presents an evaluation of the system's performance across three key aspects: Localization Accuracy, Crack Localization, and Analysis of Findings. The results highlight the system's precision in positioning, its effectiveness in detecting and mapping cracks, and key insights derived from the experiments.

### 4.1. Localization Results Evaluation

Figure 7 illustrates the actual flight path of the UAV, revealing a distinct hovering pattern at each waypoint. The actual line represents the system's measured values, while the command line represents the drone's theoretical path. The observed deviations indicate that the drone experienced slight oscillations around the waypoints during the three-second

hover period. This wobbling effect, inherent to the UAV's state estimation and correction mechanisms, contributed to variations in image capture locations, potentially shifting the recorded crack images away from the intended waypoint centers.
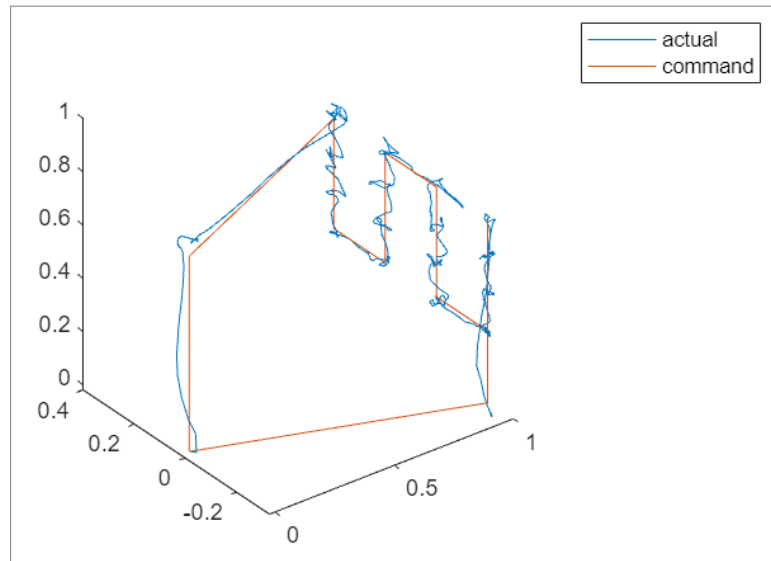
Figure 7. Actual vs. Command UAV Trajectory Using the Lighthouse Localization System (units in cm)

Figure 8 presents the offset measurements of the UAV along the X, Y, and Z axes, providing a detailed evaluation of localization performance over time, represented in frames. The recorded maximum offsets reached approximately 4 cm along the X-axis, 3.5 cm along the Y-axis, and 6 cm along the Z-axis. The average offsets across all test flights were measured at 1.08 cm, 1.73 cm, and 1.48 cm, respectively. These values indicate that, on average, the positioning system maintained the UAV's position within an accuracy of 1−2 cm, demonstrating a high level of precision in localization.
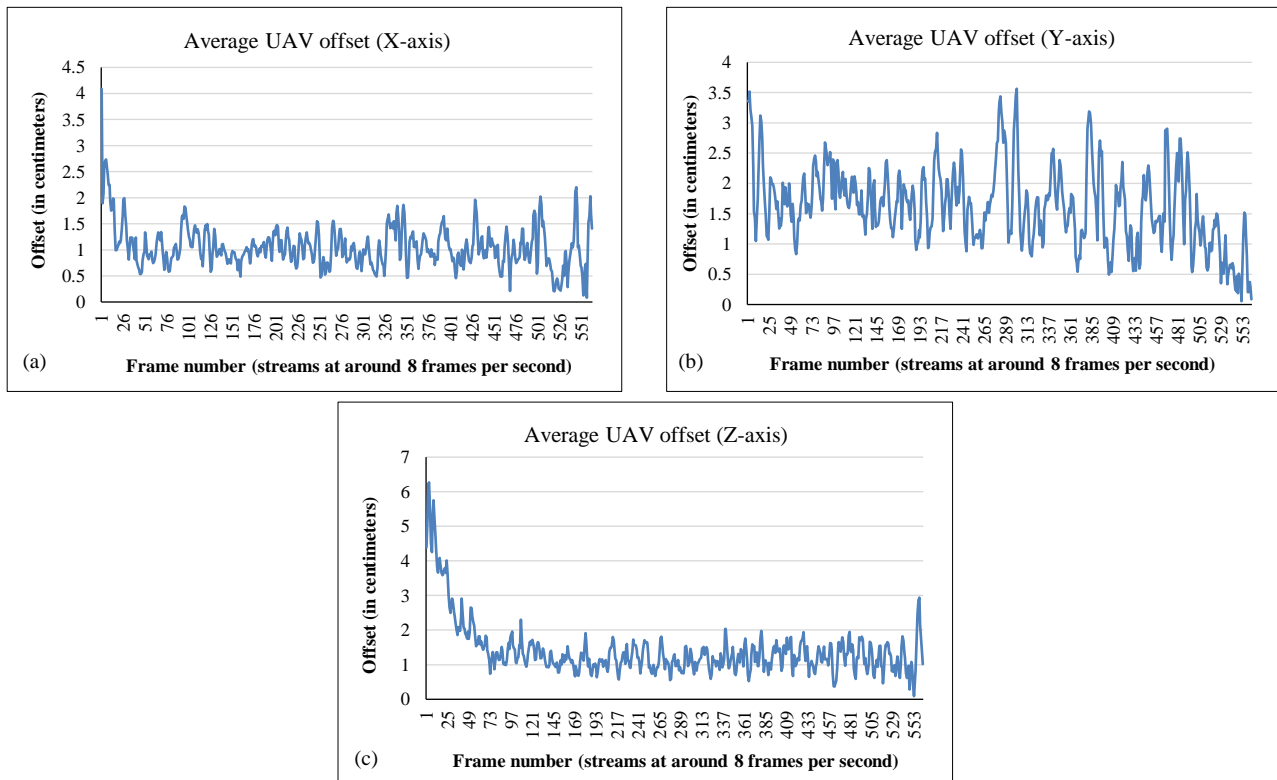
Figure 8. Average UAV Offset in cm Across (a) the X-axis, (b) the Y-axis, and (c) the Z-axis Over Time

Notably, the initial two waypoints exhibited larger deviations than the rest of the trajectory, particularly along the X and Z axes. This behavior likely resulted from the UAV's transition from takeoff to its first waypoint, introducing minor inaccuracies during early image capture. The observed oscillations measured around 3.5 to 6 cm but stabilized within approximately three seconds. Once stabilized, the system maintained an average error of approximately 1−2 cm, within acceptable limits for UAV-based crack localization. Given that the target cracks span several centimeters, this level of

accuracy ensures the UAV's field of view remains sufficient to detect and segment cracks, even with minor deviations. Beyond precision, this stability supports consistent UAV behavior, allowing rapid waypoint settling and minimizing motion blur during image capture. Compared to alternative localization methods (Table 1), the Lighthouse system offers a compelling balance of accuracy (< 2 cm), low system complexity, and ease of deployment. It requires minimal setup, does not rely on GPS or external cameras, and operates without intensive onboard computation. These findings confirm that the system enables reliable centimeter-level crack localization while remaining practical and scalable for structural inspections in GNSS-denied environments.

4.2. Crack Localization

The captured images, classification results, segmentation outputs, and estimated crack locations are presented in the figure. Each image includes position data in meters and a classification score ranging from 0 to 1, where values closer to 1 indicate a detected crack. This score represents the average classification prediction during the UAV's hover. Positive classifications undergo segmentation, with detected cracks highlighted using yellow bounding boxes. The background color of each image represents its classification outcome: red indicates no detected cracks, green signifies correctly classified and segmented cracks, and yellow highlights instances where classification and segmentation results do not align. Beneath each image, the ground truth is provided for direct comparison. Crack locations are estimated by masking bounding boxes in white against a black background, visually representing detected cracks. These masks are then stitched together using absolute coordinates to generate a composite map, where black represents the surveyed area and white denotes identified cracks.

Figure 9 illustrates the system's performance in detecting a long diagonal crack that is not part of the training dataset. The UAV correctly localized the crack spanning from the upper left region towards the center, with position data provided in meters. Among the 12 samples, only one false negative was observed, indicating strong generalization despite unseen orientations. This spatial localization was made possible through coordinate-based stitching of segmented outputs, providing detection and positional insight into defect extent and orientation. A key challenge in this experiment was the camera's limited field of view, which resulted in overlapping detections. While the system performed well overall, the same crack appeared across multiple viewpoints, complicating precise localization. To address this, a 30% overlap was applied when stitching segmented masks, allowing bounding boxes to converge into a more accurate crack representation. This approach introduces redundancy for robustness but may require tuning to optimize and avoid over-segmentation.
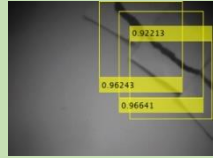
Figure 9. Crack Detection Performance Across Y and Z Coordinates for a Long Diagonal Crack

Figure 10 presents a real-world scenario where the UAV successfully identified a crack near the inspection region's center (approximately Y = -0.10 meters). The vertical extent of the crack, estimated to be around 28 cm, was effectively segmented through multiple inspection windows. Though false negatives were present in both figures, analysis suggests that misclassifications occurred when prediction scores fell below the set threshold (0.8). Notably, true positives were consistently identified around the misclassified areas, indicating that the crack remained within the UAV's field of view, which is acceptable for this application. However, future work may explore dynamic thresholding mechanisms.
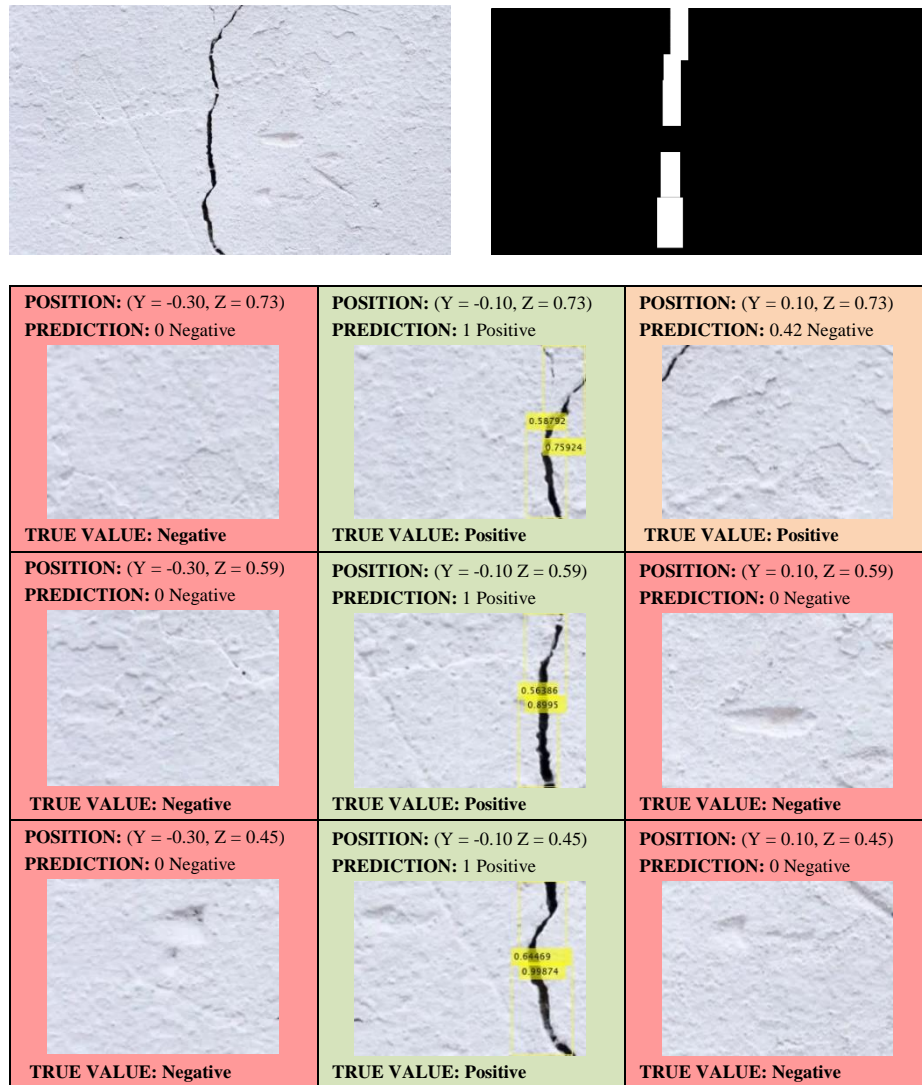

Figure 10. Crack Detection Performance on an Actual Wall Surface Across Different Y and Z Coordinates

Table 2 summarizes the model performance across all scenarios, reporting 100% precision and 83.33% recall, with an overall accuracy of 91.89%. These metrics highlight the system's robustness in classifying and segmenting cracks. When paired with coordinate mapping, this enables the generation of reliable spatial crack representations, supporting targeted maintenance and data-driven repair prioritization. These results are consistent with our previous study [11], where the proposed two-stage model achieved an F1-score of 88.7% for classification and 95.2% for segmentation. Compared to other approaches, such as ResNet50+SENet (mPA: 92.7%, IoU: 88.3%) and CNN-based classification with CTV2 (F1: 92.23%), the proposed method remains competitive while offering advantages in real-time deployment and computational efficiency.

Table 2. Confusion Matrix for the Model's Performance Across All Case Scenarios

	Prediction Negative	**Prediction Positive**	
Actual Positive	False Negative 4 (8.16%) 4	True Positive 20 (40.81%)	*__Recall__* 83.33%
Actual Negative	True Negative 25 (51.02%)	False Positive 0	
		__Precision__ 100%	*__Accuracy__* 91.89%

Unlike most crack detection studies focusing on image-level classification or segmentation [32, 33], this study emphasizes spatial crack localization using a Lighthouse-based UAV framework. The segmentation masks are tied to Y and Z position data, enabling physical mapping of defect locations in real-world coordinates. While a few studies attempt image-based feedback [34], they often lack detailed spatial context. The present approach addresses this gap, reinforcing the value of pose-informed crack detection in UAV-based SHM workflows. Moreover, the pairing of AlexNet and YOLOv4 was chosen to balance detection speed and precision. AlexNet is a lightweight classifier that reduces processing load, while YOLOv4 performs accurate segmentation, making this two-stage architecture particularly suitable for real-time UAV deployment under resource constraints.

Although the system demonstrated reliable crack detection in indoor environments, its performance under more challenging surface or lighting conditions, such as wetness, shadows, or textured concrete, was not evaluated. Nonetheless, the grayscale conversion step in the preprocessing pipeline may help mitigate minor lighting variations, such as shadows, provided sufficient illumination is maintained. Evaluating the system's robustness under real-world conditions remains valuable for future work. While the model effectively classifies and localizes cracks, it is currently limited to binary classification (crack vs. no crack) and does not distinguish between different crack types or severities. However, the segmentation masks produced by the model provide sufficient spatial information to enable post-processing for crack length estimation. Further crack width or depth analysis remains outside the current scope but represents a valuable direction for future development.

5. Conclusion

This study demonstrates a novel Lighthouse localization for UAV positioning, integrated with a two-stage CNN crack detection model. The system achieved effective and precise crack detection by leveraging AlexNet for lightweight binary classification and YOLOv4 for real-time segmentation. The localization component provided sub-centimeter accuracy, stabilizing within a 1-2 cm margin, ensuring reliable UAV tracking in GNSS-denied environments. Spatial coordinate assignment (Y and Z) further enhanced the system's utility for mapping defects, supporting targeted maintenance, and reducing human intervention.

The study confirms the viability of the Lighthouse system as a cost-effective and precise solution for indoor infrastructure inspection. Future work may improve segmentation overlap, expand the dataset for greater generalization across structural conditions, and investigate scalability through multi-base station setups. Since the Lighthouse system calculates UAV position based on proximity to visible base stations, increasing the number of base stations could extend coverage, provided consistent line-of-sight is maintained [35]. These directions open new possibilities for deploying UAV-based SHM systems in larger or multi-room environments. This research reinforces the potential of combining efficient localization with deep learning for scalable, precise defect detection in GNSS-denied settings.

6. Declarations

6.1. Author Contributions

Conceptualization, J.S., A.C., and T.C.; methodology, J.S., A.C., and T.C.; software, J.S.; validation, J.S. and T.C.; formal analysis, J.S.; investigation, J.S.; resources, A.C. and T.C.; data curation, J.S. and T.C.; writing—original draft preparation, J.S. and T.C.; writing—review and editing, A.C. and T.C.; visualization, T.C.; supervision, A.C. and T.C.; project administration, A.C.; funding acquisition, A.C. and T.C. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. References

[1] Kot, P., Muradov, M., Gkantou, M., Kamaris, G. S., Hashim, K., & Yeboah, D. (2021). Recent advancements in non-destructive testing techniques for structural health monitoring. Applied Sciences (Switzerland), 11(6), 2750. doi:10.3390/app11062750.

[2] Say, M. F. Q., Sybingco, E., Bandala, A. A., Vicerra, R. R. P., & Chua, A. Y. (2021). 2D Position Control of a UAV Using Fuzzy Logic Control. 2021 IEEE/SICE International Symposium on System Integration, SII 2021, 679–683. doi:10.1109/IEEECONF49454.2021.9382784.

[3] Piquero, J. (2019). A new sliding mode controller implementation on an autonomous quadcopter system. International Journal of Automation and Smart Technology, 9(2), 53-63. doi:10.5875/ausmt.v9i2.1876.

[4] Aniceto1, S. B. P., McGrah, R. V. S., Ochengco, C. J. I., Regalado, M. G., & Chua, A. Y. (2020). A novel low-cost obstacle avoidance system for a quadcopter UAV using fuzzy logic. International Journal of Mechanical Engineering and Robotics Research, 9(5), 733–738. doi:10.18178/ijmerr.9.5.733-738.

[5] Rathod, V. V., Rana, D. P., & Mehta, R. G. (2024). Deep learning-driven UAV vision for automated road crack detection and classification. Nondestructive Testing and Evaluation, 1–30. doi:10.1080/10589759.2024.2446650.

[6] Jin, T., Zhang, W., Chen, C., Chen, B., Zhuang, Y., & Zhang, H. (2023). Deep-Learning- and Unmanned Aerial Vehicle-Based Structural Crack Detection in Concrete. Buildings, 13(12), 3114. doi:10.3390/buildings13123114.

[7] Duan, Z., Liu, J., Ling, X., Zhang, J., & Liu, Z. (2024). ERNet: A Rapid Road Crack Detection Method Using Low-Altitude UAV Remote Sensing Images. Remote Sensing, 16(10), 1741. doi:10.3390/rs16101741.

[8] Beguni, C., Done, A., Căilean, A. M., Avătămăniței, S. A., & Zadobrischi, E. (2023). Experimental Demonstration of a Visible Light Communications System Based on Binary Frequency-Shift Keying Modulation: A New Step toward Improved Noise Resilience. Sensors, 23(11), 5001. doi:10.3390/s23115001.

[9] Krishnan, S. S. R., Karuppan, M. K. N., Khadidos, A. O., Khadidos, A. O., Selvarajan, S., Tandon, S., & Balusamy, B. (2025). Comparative analysis of deep learning models for crack detection in buildings. Scientific Reports, 15(1), 2125. doi:10.1038/s41598-025-85983-3.

[10] Qayyum, W., Ehtisham, R., Bahrami, A., Camp, C., Mir, J., & Ahmad, A. (2023). Assessment of Convolutional Neural Network Pre-Trained Models for Detection and Orientation of Cracks. Materials, 16(2), 826. doi:10.3390/ma16020826.

[11] Sorilla, J., Chu, T. S. C., & Chua, A. Y. (2024). A UAV Based Concrete Crack Detection and Segmentation Using 2-Stage Convolutional Network with Transfer Learning. HighTech and Innovation Journal, 5(3), 690–702. doi:10.28991/HIJ-2024-05-03-010.

[12] Xie, D., Wang, X., Tang, A., & Zhu, H. (2022). A Portable RFID Localization Approach for Mobile Robots. IEEE Internet of Things Journal, 9(23), 23332–23347. doi:10.1109/JIOT.2022.3202136.

[13] Buffi, A., Motroni, A., Nepa, P., Tellini, B., & Cioni, R. (2019). A SAR-Based Measurement Method for Passive-Tag Positioning with a Flying UHF-RFID Reader. IEEE Transactions on Instrumentation and Measurement, 68(3), 845–853. doi:10.1109/TIM.2018.2857045.

[14] Motroni, A., & Nepa, P. (2023). UAV-based 3D localization of passive UHF-RFID tags empowering outdoor stock management. 2023 35th General Assembly and Scientific Symposium of the International Union of Radio Science, URSI GASS 2023, 1–4. doi:10.23919/URSIGASS57860.2023.10265434.

[15] Böller, S., Bödder, A., Riegler, J., & Grenter, T. (2024). SHF RFID based Localization: Combining Time-of-Flight and Angle-of-Arrival. 2024 IEEE International Conference on RFID Technology and Applications, RFID-TA 2024, 70–73. doi:10.1109/RFID-TA64374.2024.10965186.

[16] Buffi, A., Nepa, P., & Cioni, R. (2017). SARFID on drone: Drone-based UHF-RFID tag localization. 2017 IEEE International Conference on RFID Technology and Application, RFID-TA 2017, 40–44. doi:10.1109/RFID-TA.2017.8098872.

[17] Kabiri, M., Cimarelli, C., Bavle, H., Sanchez-Lopez, J. L., & Voos, H. (2023). A Review of Radio Frequency Based Localisation for Aerial and Ground Robots with 5G Future Perspectives. Sensors, 23(1), 188. doi:10.3390/s23010188.

[18] De Guzman, C. J. P., Chua, A. Y., & Chu, T. S. C. (2024). Path Planning of Multiple Quadrotors with Ultrawideband Localization for Warehouse Inventory Management. Journal of Robotics, 2566619. doi:10.1155/2024/2566619.

[19] Yang, B., Yang, E., Yu, L., & Loeliger, A. (2022). High-Precision UWB-Based Localisation for UAV in Extremely Confined Environments. IEEE Sensors Journal, 22(1), 1020–1029. doi:10.1109/JSEN.2021.3130724.

[20] Queralta, J. P., Martinez Almansa, C., Schiano, F., Floreano, D., & Westerlund, T. (2020). UWB-based system for UAV localization in GNSS-Denied environments: Characterization and dataset. IEEE International Conference on Intelligent Robots and Systems, 4521–4528. doi:10.1109/IROS45743.2020.9341042.

[21] Sorescu, T. G., Militaru, A. V., Chiriac, V. M., Comsa, C. R., & Alecsandrescu, I. E. (2024). UWB Indoor Localization: Accuracy Evaluation in a Controlled Environment. Proceedings of the 16th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2024, 1–6. doi:10.1109/ECAI61503.2024.10607552.

[22] Preiss, J. A., Honig, W., Sukhatme, G. S., & Ayanian, N. (2017). Crazyswarm: A large nano-quadcopter swarm. Proceedings - IEEE International Conference on Robotics and Automation, 3299–3304. doi:10.1109/ICRA.2017.7989376.

[23] Mu, L., Cao, S., Zhang, Y., Zhang, X., Feng, N., & Zhang, Y. (2025). Autonomous Landing Guidance for Quad-UAVs Based on Visual Image and Altitude Estimation. Drones, 9(1), 57. doi:10.3390/drones9010057.

[24] Murhij, Y., Gavrilov, D., Buzdin, V., Kafa, W., Tatarinova, E., Fateev, A., Shkatula, S., & Krichevets, D. (2024). Visual Localization system for GPS-Blind Environments in Unmanned Aerial Vehicles. Proceedings of the 2024 8th International Conference on Information, Control, and Communication Technologies, ICCT 2024, 1–5. doi:10.1109/ICCT62929.2024.10875019.

[25] Wu, P., Tong, P., Zhou, X., & Yang, X. (2024). Dyn-DarkSLAM: YOLO-based Visual SLAM in Low-light Conditions. Proceedings of 2024 IEEE 25th China Conference on System Simulation Technology and Its Application, CCSSTA 2024, 346–351. doi:10.1109/CCSSTA62096.2024.10691775.

[26] Gibson, J., Haseler, C., Lassiter, H., Liu, R., Morrow, G., Oslund, B., Zrimm, M., & Lewin, G. (2016). Lighthouse localization for unmanned applications. 2016 IEEE Systems and Information Engineering Design Symposium, SIEDS 2016, 187–192. doi:10.1109/SIEDS.2016.7489296.

[27] Campos, F. M. R., Schindler, C. B., Kilberg, B. G., & Pister, K. S. J. (2020). Lighthouse Localization of Wireless Sensor Networks for Latency-Bounded, High-Reliability Industrial Automation Tasks. IEEE International Workshop on Factory Communication Systems - Proceedings, WFCS, 2020-April, 1–8. doi:10.1109/WFCS47810.2020.9114443.

[28] Greiff, M., Robertsson, A., & Berntorp, K. (2019). Performance Bounds in Positioning with the VIVE Lighthouse System. FUSION 2019 - 22nd International Conference on Information Fusion, 1–8. doi:10.23919/fusion43075.2019.9011242.

[29] Martin, T., Blanco, J. R., Mouret, J. B., & Raharijaona, T. (2024). Compact Docking Station for Sub-150g UAV Indoor Precise Landing. International Conference on Unmanned Aircraft Systems, ICUAS 2024, 1267–1274. doi:10.1109/ICUAS60882.2024.10557124.

[30] Bitcraze (2024). Lighthouse Kalman Measurment Model. Bitcraze, Malmö, Sweden. Available online: https://www.bitcraze.io/documentation/repository/crazyflie-firmware/master/functional-areas/lighthouse/kalman_measurement_model/ (accessed on May 2025).

[31] Özgenel, Ç. F. (2019), Concrete Crack Images for Classification. Mendeley Data, V2. doi:10.17632/5y9wdsg2zt.2.

[32] Yang, Z. (2024). Consider Computer Vision for Building Component Recognition and Defect Detection. 2024 5th International Conference on Machine Learning and Computer Application, 457–460. doi:10.1109/ICMLCA63499.2024.10754392.

[33] Zaman, F., Xu, Z., Hussain, A., Tariq, Aslam, A., & Abdullahi, M. R. (2024). Concrete and Pavement Cracks Detection Leveraging Vision Transformer (ViT) Model. Proceedings - IEIT 2024, International Conference on Electrical and Information Technology, 90–95. doi:10.1109/IEIT64341.2024.10762971.

[34] Bakirci, M., & Avran, M. N. (2025). Automated Detection and Localization of Wind Turbine Deformations Through YOLO11 and UAV Systems. Proceedings - 2025 International Russian Smart Industry Conference, SmartIndustryCon 2025, 109–114. doi:10.1109/SmartIndustryCon65166.2025.10986124.

[35] Bitcraze (2024). Lighthouse more than 4 base stations. Bitcraze, Malmö, Sweden. Available online: https://www.bitcraze.io/documentation/repository/crazyflie-firmware/master/functional-areas/lighthouse/multi_base_stations/ (accessed on May 2024).