Journal of Sustainable Civil Engineering and Technology
e-ISSN: 2948-4294 | Volume 4 Issue 21 (September 2025), 39-46
https://doi.org/10.24191/jscet.v4i2.JSCET_ICoSCID2025_006

Evaluation of a Pretrained Deep Learning Model for Indoor Crack Detection Using DSLR and Mobile Phone Cameras

Mohd Ashraf Ahmad Zubir[1], Khairulazhar Zainuddin[1*], Abd Wahid Rasib[2], Zulkepli Majid[2], Norbazlan Mohd Yusof[3], Azizul Faiz Abdul Aziz[3]

[1] Faculty of Built Environment, Universiti Teknologi MARA Perlis Branch, 02600, Arau, Malaysia
[2] Faculty of Built Environment & Surveying, Universiti Teknologi Malaysia, 81310, Malaysia
[3] Innovation & Centre of Excellence, PLUS Berhad, 47301, Petaling Jaya, Malaysia

*Corresponding author: khairul760@uitm.edu.my

Received: 10 July 2025 / Accepted: 08 August 2025 / Published online: 30 September 2025

Abstract

Pretrained deep learning models have shown strong potential in automating crack detection for structural health monitoring. Most of these models are trained using datasets captured in outdoor environments under natural lighting. In addition, many crack detection models operate on two-dimensional images, which lack geometric context and limit the spatial interpretation of defects. The iTwin Capture Modeler by Bentley Systems addresses this limitation by integrating pretrained detection models with photogrammetric processing, enabling cracks to be detected and visualised directly on three-dimensional (3D) models. However, the pretrained model was developed using outdoor environments with image resolution of around 1 cm/pixel. Hence, this study aims to evaluate its performance under indoor conditions, where lighting and surface texture may differ significantly. Images were collected using a Digital Single Lens Reflex (DSLR) camera and a mobile phone. The DSLR produced native high-resolution images, whereas the mobile phone relied on pixel binning to improve image clarity in low-light situations. Both sets of images were used to generate 3D models through photogrammetric techniques, and crack detection was performed inside the iTwin software. The performance of the crack detection model was then evaluated by calculating its precision, recall, and F1-score. The DSLR camera recorded higher scores across all performance measures due to its superior optical quality and greater manual control. The mobile phone also provided satisfactory results despite having hardware limitations. These findings indicate that the pretrained model remains effective for detecting cracks in indoor environments and can be applied using a variety of image capture devices for three-dimensional inspection workflows.

Keywords: 3D Crack Detection, Deep Learning, Photogrammetry, iTwin Capture Modeler

1. Introduction

Structural Health Monitoring (SHM) plays a crucial role in ensuring the long-term integrity and safety of engineering structures, including tunnels and buildings (Abdul Razak et al., 2022). A core element of SHM involves detecting cracks that may appear on concrete surfaces due to factors such as excessive stress, ground movement, or material deterioration. Identifying these cracks at an early stage can prevent serious structural failures and contribute to more effective and timely maintenance strategies (Zhang et al., 2025). To support such early detection efforts, advanced digital tools such as digital twins have been increasingly adopted.

Advances in digital technology have led to the emergence of the digital twin concept in Structural Health Monitoring (SHM). A digital twin refers to a three-dimensional (3D) representation of a physical structure that incorporates sensor data and artificial intelligence to assess its condition (Radek Zhunek, 2025). This technology improves crack detection by integrating real-time structural data with deep learning models, allowing for analysis that is not only automated but also more precise and predictive (Sacks et al., 2020). The integration

approach enables early detection of defects, reduces manual inspection efforts, and supports proactive maintenance strategies using virtual representations of the structure (Boje et al., 2020).

The iTwin Capture Modeler (iTwin) by Bentley System is widely recognised within the engineering field as a prominent digital twin platform. It provides tools for generating reality-based data, including 3D modelling, spatial measurements, and automated classification of objects and regions through machine learning algorithms. Its adoption continues to grow across infrastructure and smart city projects, with both industry practitioners and academic literature acknowledging its role as a leading solution (D. Li et al., 2022).

Pretrained deep learning models used for crack detection are typically designed and validated under controlled outdoor conditions with specific image resolutions. The model integrated into iTwin, for instance, has been optimised for structural assessments in outdoor environments using imagery at approximately 1 cm/pixel resolution. In contrast, indoor structural health monitoring often demands the detection of finer cracks, frequently measuring less than 1 mm in width. This discrepancy prompts a critical evaluation of the model's robustness and adaptability when deployed in settings that differ significantly from its original training context.

This paper aims to evaluate the performance of a pretrained crack detection model provided in the iTwin software when applied to indoor environments, where lighting and surface texture conditions differ significantly from the trained specifications. A key objective of this study was to examine how variations in camera specifications, such as image resolution, sensor quality, and optical characteristics, impact the model's ability to accurately detect fine cracks on indoor surfaces. The evaluation involved comparing detection results obtained from a Digital Single-Lens Reflex (DSLR) camera and a mobile phone camera. Both cameras were utilised to capture high-resolution images for photogrammetric reconstruction and automated crack segmentation.

2. Related Work

Detecting cracks is a fundamental part of structural health monitoring, especially when dealing with critical infrastructure such as buildings and bridges. Cracks may develop due to factors like material wear, environmental conditions, including soil movement, or mechanical stress. If not identified and addressed at an early stage, they can compromise structural integrity and pose a threat to public safety (Tello-Gil et al., 2024). Conventional inspection methods, which rely heavily on manual visual assessments, tend to be time-consuming, labour-intensive, and prone to human error. These limitations may reduce the accuracy and reliability of evaluations, ultimately affecting the effectiveness of maintenance planning (Tello-Gil et al., 2024; D. Li et al., 2022).

Photogrammetry has gained recognition as a reliable method for structural monitoring, particularly in identifying cracks within built infrastructure. The review by H. Li et al. (2022) and Deng et al. (2024) demonstrated how photogrammetric techniques can process images to generate high-quality 3D models when combined with machine learning or deep learning approaches. These models support detailed inspections by capturing the actual condition of structural surfaces. The resulting 3D reconstructions offer a precise and non-destructive means of examining concrete surfaces, which enhances both crack detection and damage assessment (Mett and Eder, 2019). This approach not only improves the accuracy of evaluations but also helps streamline the inspection process with minimal disruption.

Previous studies have highlighted the potential of photogrammetry in crack detection, particularly in structures where traditional inspection methods fall short (Attard et al., 2018). In many of these studies, deep learning models were applied to interpret the captured images for identifying cracks. However, most of the research focused on two-dimensional images without incorporating camera calibration, orientation correction, or spatial scaling. As a result, these systems could not measure the actual size or location of cracks, instead offering outputs such as bounding boxes or segmented areas that lacked geometric context (Jha et al., 2023; Deng et al., 2024).

Journal of Sustainable Civil Engineering and Technology
e-ISSN: 2948-4294 | Volume 4 Issue 21 (September 2025), 39-46
https://doi.org/10.24191/jscet.v4i2.JSCET_ICoSCID2025_006

The use of pretrained models has proven to be beneficial when analysing large annotated datasets, particularly for scalability. While these models are typically designed for outdoor conditions, applying them indoors presents additional challenges. Key issues include the quality and resolution of the images, which can be affected by inconsistent lighting and shadows. Li et al. (2022) pointed out that camera configuration significantly influences detection accuracy, with elements such as resolution and lighting playing critical roles.

3. Methodology

The data collection for this study took place within the Meru Tunnel in Ipoh, Perak (Figure 1). A specific panel, approximately 10 meters long and exhibiting existing surface cracks, was selected as the test subject, as it was already undergoing routine structural monitoring. This panel was chosen due to its accessibility and the presence of varied crack types, which reflect common indoor deterioration patterns. This provided a realistic and controlled environment for evaluating crack detection performance using high-resolution photogrammetric imaging.

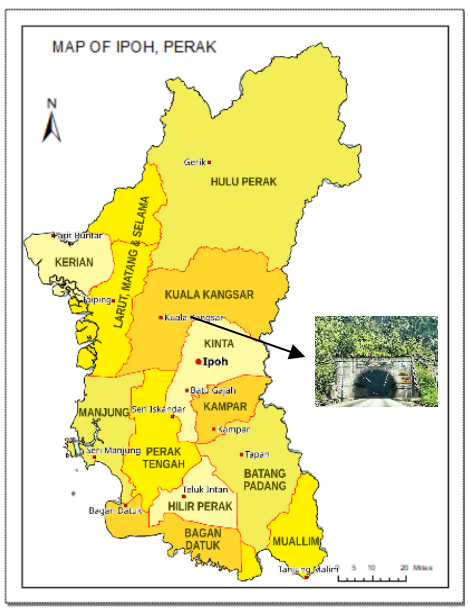


Figure 1. Study area conducted at Meru, Perak.

Two types of cameras were used for image acquisition, as shown in Figure 2. The first camera was a Sony Alpha 7 III equipped with a 29 mm focal length lens and a 24-megapixel full-frame sensor. The second camera was a Xiaomi 11T Pro with a 6 mm focal length lens and a 108-megapixel sensor. Images were captured at an approximate distance of 3 metres from the tunnel walls. This configuration was estimated to produce a ground sampling distance (GSD) ranging between 0.3 mm and 1 mm, which is suitable for detecting cracks within that resolution range.

A total of 246 images were acquired using the DSLR camera, achieving an average ground sampling distance (GSD) of approximately 0.7 mm/pixel. Meanwhile, the mobile phone camera captured 125 images with a finer GSD of approximately 0.4 mm/pixel. This variation in GSD reflects the differences in focal length and image processing pipeline, particularly pixel binning applied in the mobile phone.

The tunnel had an existing lighting system that provided sufficient illumination for image capture. Additional lighting was not required, as the Sony camera operated effectively in low-light conditions due to its larger sensor and high sensitivity. Meanwhile, the Xiaomi 11T Pro employed pixel binning technology to enhance image

brightness in low-light environments. Although this allowed the mobile camera to function without external lighting, it remained susceptible to image noise, which could affect crack detection accuracy.

(a) (b)

Figure 2. Cameras used for image acquisition, (a) Sony Alpha 7 III, and (b) Xiaomi 11T Pro mobile phone.

The acquired images were then processed using iTwin software, where all the image datasets underwent photogrammetric alignment. During this stage, iTwin identified common feature points between overlapping images to calculate the position and orientation of each image based on the Structure-from-Motion (SfM). The estimated camera orientation parameters were then used to generate a dense point cloud based on the Multiview Stereo (MVS) method. This was followed by meshing and texturing the dense point cloud to produce a photorealistic 3D model of the tunnel wall.

The next stage was automated object segmentation using deep learning to detect cracks on the wall surface. The pretrained model provided in iTwin was applied directly without modification. Although the model had originally been trained on drone and handheld image datasets under outdoor conditions, with a resolution of approximately 1 cm/pixel, it was used in this study to assess its applicability in an indoor condition.

The crack detection model works by analysing each photograph to identify visible surface cracks and generating vector representations in the form of polylines. These vectors were then projected onto the corresponding locations on the 3D model using orientation information derived from the previously performed image alignment process. This enabled spatial visualisation of cracks over a 3D model, and allowed dimensional analysis, including the crack length, width, and spatial position on the tunnel wall. Figure 3 illustrates the segmented crack lines overlaid in cyan on both the reconstructed 3D model and the 2D oriented image.

The performance of the pretrained crack detection model was evaluated using both qualitative and quantitative measures. Quantitative evaluation involved calculating standard classification metrics, comprising precision, recall, and F1-score. These metrics were computed based on the number of correctly identified crack pixels (true positives), incorrectly identified non-crack pixels (false positives), and missed crack pixels (false negatives). The evaluation considered all detected cracks regardless of size, and performance was assessed based on the accuracy of segmentation across the entire visible crack set. The analysis was carried out separately for images acquired using the DSLR and mobile phone cameras to examine the influence of image quality and sensor differences on detection outcomes.

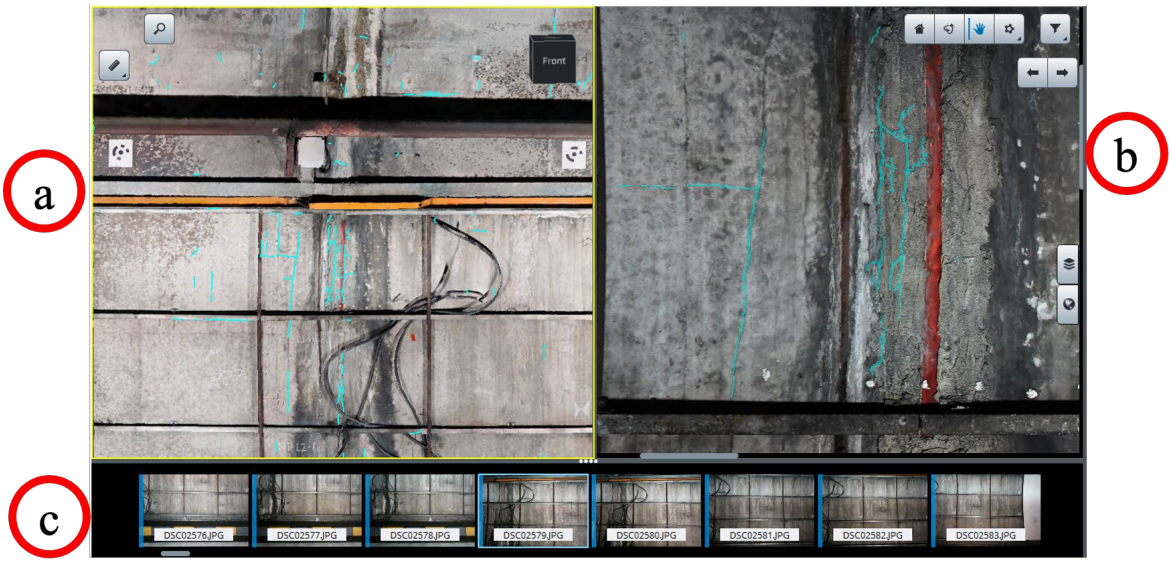


Figure 3. Crack detection visualisation in the iTwin interface. (a) Reconstructed 3D model of the tunnel wall showing segmented cracks overlaid in cyan, (b) Corresponding 2D oriented photograph with segmentation overlay, and (c) Image thumbnails used for model reconstruction.

Precision in crack detection refers to the proportion of detected crack pixels that truly represent actual cracks. High precision helps reduce false positives, which is particularly important in safety-critical infrastructure inspections where unnecessary interventions can be costly or disruptive. Recall indicates the model's ability to detect all existing cracks within an image. A high recall reduces the likelihood of missing critical defects, which is essential for structural safety and maintenance planning. The F1-score provides a single, balanced metric by combining both precision and recall. This is especially useful when dealing with imbalanced datasets, where cracks occupy only a small portion of the image. While the F1-score assumes equal importance of precision and recall, in practical scenarios, one may be prioritised over the other depending on operational needs.

The formulas used to compute the three metrics are presented in Equations (1) – (3):

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{1}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2}$$

$$F1\text{-}score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \tag{3}$$

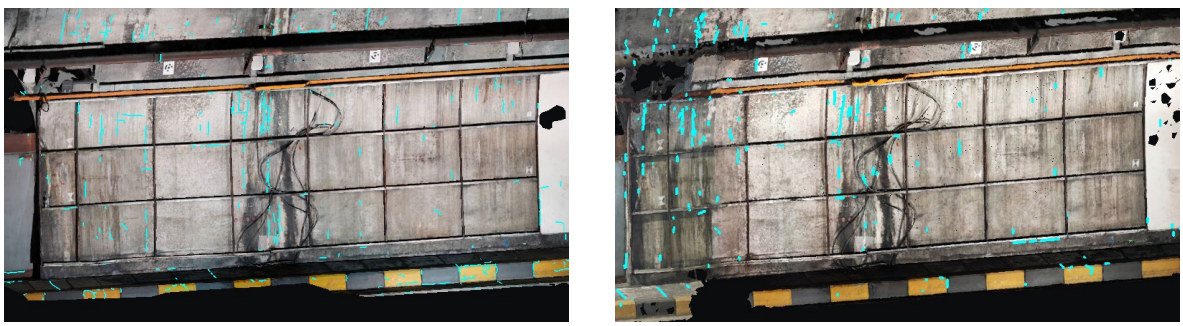
3. Results and Discussion

Figure 4 presents a comparison of 3D models reconstructed from DSLR and mobile phone images. Detected cracks are highlighted in cyan, clearly indicating areas where the model identified potential structural damage. In both cases, the most severe damage appears near the centre of the tunnel wall. While the overall detection pattern is consistent between the two models, differences in reconstruction quality are evident due to the capabilities of each camera.

The DSLR camera produced images with higher visual fidelity, capturing fine surface details with better

dynamic range and minimal noise. These characteristics contributed to sharper texture mapping and more accurate 3D geometry, especially under challenging indoor lighting conditions. Furthermore, the DSLR allows manual control over focus and exposure, which enhances photogrammetric accuracy and improves the input quality for deep learning algorithms.

In contrast, the mobile phone camera offered a lightweight and accessible solution, producing high-resolution images despite its smaller sensor. However, the resulting model exhibited slight distortions and increased image noise, particularly in lower-light areas. Although the phone-based model retained the main structural features and detected key cracks, its texture quality and geometric consistency were comparatively lower than the DSLR output. The model detected a total of 150 cracks from the Sony camera's images and 138 cracks from the Xiaomi mobile phone images, indicating slightly higher sensitivity with higher-quality image input.



(a) (b)

Figure 4. Reconstructed 3D models with detected cracks from (a) DSLR and (b) mobile phone images.

Detection performance was assessed using standard segmentation metrics, which consist of precision, recall, and F1-score, which reflect the accuracy and completeness of the model in identifying crack regions.

Table 1. Performance metrics of crack detection using DSLR and mobile phone images.

Camera Types	Precision %	Recall %	F1-score %
DSLR	91.43	94.12	92.75
Mobile phone	83.78	81.31	82.01

The detection performance, as summarised in Table 1 above, shows that the pretrained crack detection model achieved strong results with both camera types. The DSLR camera produced higher values across all three metrics, with a precision of 91.4%, a recall of 94.12%, and an F1-score of 92.75%. The mobile phone camera also recorded relatively high scores, with a precision of 83.78%, a recall of 81.31%, and an F1-score of 82.01%.

The results indicate that the pretrained crack detection model provided in iTwin software performed reliably on high-resolution images captured in an indoor tunnel environment, even though it was originally trained on outdoor datasets. This highlights the robustness of the pretrained model, showing that it can generalise to different lighting and structural conditions with minimal degradation in performance.

While both cameras demonstrated good performance, several technical differences explain the score variation. The DSLR camera produced natively high-resolution images with greater sharpness, lower noise, and better dynamic range, making it more suited for image capture under low-light conditions. This improved clarity enhanced crack segmentation, leading to more accurate detection.

In contrast, the mobile phone camera employed pixel binning, where images though nominally 108 megapixels were effectively down sampled to 12 megapixels. This reduced the level of detail and introduced minor

distortions, especially under limited lighting, which likely contributed to the lower scores. Nevertheless, the mobile phone still enabled successful crack detection in the indoor setting, further reinforcing the pretrained model's versatility and practicality for field applications using accessible devices.

5. Conclusion

This study evaluated the application of a pretrained deep learning model for crack detection in an indoor tunnel setting using images captured by both DSLR and mobile phone cameras. The investigation focused on assessing the model's ability to generalise beyond its original outdoor training context by testing its performance on high-resolution images acquired from different camera sources.

The results demonstrated that the model successfully detected cracks in both image sets, confirming its potential for use in indoor environments. The DSLR camera achieved higher precision, recall, and F1-score, primarily due to its native high-resolution output, better low-light sensitivity, and manual control capabilities. The mobile phone, despite relying on pixel binning and automated processing, also delivered strong detection performance.

While the dataset used in this study was limited to one section of the tunnel panel and 371 images, but reflective of typical indoor inspection conditions and was sufficient to demonstrate the model's capability. These findings highlight the robustness of pretrained crack detection models and support their practical deployment in constrained environments. Future research with larger and more diverse datasets is recommended to further validate these results across different indoor structural contexts.

These findings support the feasibility of deploying pretrained crack detection models in new application domains without additional retraining. The ability to use both professional and consumer-grade imaging devices broadens the practical scope of automated crack inspection workflows, particularly in resource-constrained or complex infrastructure environments.

Acknowledgments

The authors would like to acknowledge Universiti Teknologi MARA (UiTM), Perlis Branch, for providing research facilities and support under research grant 100-TNCPI/GOV 16/6/2 (018/2024). Appreciation is also extended to the Faculty of Built Environment and Surveying (FABU), Universiti Teknologi Malaysia (UTM), for their technical contributions and financial support through grants VOT R.J130000.7652.4C773 and Q.J130000.3052.04M83. Special thanks are due to PLUS Malaysia Berhad (PMB) for their financial assistance and technical support throughout the data collection phase.

Declaration of Conflicting Interests

All authors declare that they have no conflicts of interest.

Author Contributions

Conceptualisation, Khairulazhar Zainuddin. Methodology, Mohd Ashraf Ahmad Zubir. Validation, Norbazlan Mohd Yusof and Azizul Faiz Abdul Aziz. Analysis, Mohd Ashraf Ahmad Zubir and Khairulazhar Zainuddin Investigation, Khairulazhar Zainuddin and Zulkepli Majid. Resources, Abd Wahid Rasib, Norbazlan Mohd Yusof and Azizul Faiz Abdul Aziz. Data Curation, Mohd Ashraf Ahmad Zubir, Khairulazhar Zainuddin, Abd Wahid Rasib and Zulkepli Majid. Writing-Draft Preparation, Mohd Ashraf Ahmad Zubir. Writing-Review & Editing, Khairulazhar Zainuddin. Visualisation, Mohd Ashraf Ahmad Zubir. Supervision, Khairulazhar Zainuddin and Abd Wahid Rasib. Project Administration, Norbazlan Mohd Yusof and Azizul Faiz Abdul Aziz Funding Acquisition, Abd Wahid Rasib. All authors have reviewed and approved the final version of the manuscript for publication.

References

Razak, A., Hadi, A., Abdullah, N. S., Al Junid, S. A., Halim, A. K., Idros, M. F., ... & Nazamuddin, F. (2022). Structural Crack Detection System Using Internet of Things (IoT) for Structural Health Monitoring (SHM): A Review. Jurnal Kejuruteraan34, 6, 983-98.

Attard, L., Debono, C. J., Valentino, G., & Di Castro, M. (2018). Tunnel inspection using photogrammetric techniques and image processing: A review. ISPRS journal of photogrammetry and remote sensing, 144, 180-188.

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. Automation in construction, 114, 103179.

Deng, L., Yuan, H., Long, L., Chun, P. J., Chen, W., & Chu, H. (2024). Cascade refinement extraction network with active boundary loss for segmentation of concrete cracks from high-resolution images. Automation in Construction, 162, 105410.

Jha, H., Mukherjee, A., Banerjee, S., Roy, T., Das, R., Middya, A. I., & Roy, S. Automated Concrete Structure Defect Detection Using Vision Transformer. Available at SSRN 5096179.

Li, D., Liu, J., Hu, S., Cheng, G., Li, Y., Cao, Y., ... & Chen, Y. F. (2022). A deep learning-based indoor acceptance system for assessment on flatness and verticality quality of concrete surfaces. Journal of Building Engineering, 51, 104284.

Li, H., Wang, W., Wang, M., Li, L., & Vimlund, V. (2022). A review of deep learning methods for pixel-level crack detection. Journal of Traffic and Transportation Engineering (English Edition), 9(6), 945-968.

Mett, M., Kontrus, H., & Eder, S. (2019). 3D tunnel inspection with photogrammetric and hybrid systems. Proceedings of the Shotcrete for Underground Support XIV: Pattaya, Thailand, 17-20.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., & Girolami, M. (2020). Construction with digital twin information systems. Data-centric engineering, 1, e14.

Tello-Gil, C., Jabari, S., Waugh, L., Masry, M., & McGinn, J. (2024). Crack detection and dimensional assessment using smartphone sensors and deep learning. Canadian Journal of Civil Engineering, 51(11), 1197-1211.

Zhang, X., Yu, Y., Yu, Z., Qiao, F., Du, J., & Yao, H. (2025). A Scoping Review: Applications of Deep Learning in Non-Destructive Building Tests. Electronics (2079-9292), 14(6).