sustainability

MDPI

Article

Real-Time Multi-Damage Detection and Risk Prioritisation for Aging Buildings Using YOLOv11 and a Damage Criticality Index

Jongnam Ho [1], Yonghan Ahn [2] and Hyunkyu Shin [3,*]

[1] Department of Smart City Engineering, Hanyang University, Ansan 15588, Republic of Korea; ghwhdska@hanyang.ac.kr
[2] Department of Architectural Engineering, Hanyang University ERICA, Ansan 15588, Republic of Korea; yhahn@hanyang.ac.kr
[3] Division of Architecture, Mokwon University, Daejeon 35349, Republic of Korea
* Correspondence: shinhk@mokwon.ac.kr

Abstract

Ageing building stock, shrinking budgets, and inspector shortages hinder timely façade safety inspections. This research develops an automated damage detection and risk prioritization system for aging concrete structures. Five YOLOv11 variants were trained on 130,838 high-resolution images from 25 Seoul districts to detect three critical damage types: cracks, exposed rebar, and spalling. The proposed framework integrates YOLOv11 detection with a novel Damage Criticality Index (DCI) that transforms five visual-spatial cues—area, multiplicity, confidence, density, and spread—into continuous severity scores, subsequently categorized into low, medium, and high risk via K-means clustering. YOLOv11x achieved 0.78 mAP@0.5 at 101 FPS, enabling real-time processing suitable for field deployment. Field trials confirmed robust detection and consistent risk ranking in both uncluttered and cluttered urban environments, substantially reducing inspection time and minimizing missed defects compared to conventional manual methods. The framework provides scalable, data-driven support for city-wide monitoring and transparent, risk-prioritized maintenance of aging infrastructure.

Keywords: YOLOv11; damage detection; structural inspection; safety assessment; DCI

check for updates

Academic Editor: Antonio Caggiano

Received: 24 September 2025
Revised: 20 October 2025
Accepted: 21 October 2025
Published: 22 October 2025

Citation: Ho, J.; Ahn, Y.; Shin, H. Real-Time Multi-Damage Detection and Risk Prioritisation for Aging Buildings Using YOLOv11 and a Damage Criticality Index. *Sustainability* **2025**, *17*, 9390. https://doi.org/10.3390/su17219390

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As of 2022, buildings over 30 years old constitute more than 41% of the total building stock in South Korea. Notably, buildings older than 35 years make up 32.22% of this figure [1]. The increasing number of aging structures has led to a growing demand for visual safety inspections. However, according to a survey conducted by Associated General Contractors of America (AGC), 80% of contractors reported difficulties in hiring skilled technicians [2]. The shortage of skilled technicians has become an ongoing issue affecting the entire construction industry [3].

Amid this shortage of skilled technicians, the growing number of aging buildings has led to an increase in safety inspection costs. If accurate safety inspections and maintenance are not carried out, it can result in serious safety issues and accidents [4]. To ensure efficient safety inspections and maintenance of aging buildings, there is a need for a system that can proactively inspect and manage aging structures [5]. Studies have shown that both routine and in-depth visual inspections consistently provide inaccurate results [6], necessitating automated detection systems.

Visual inspection has been the most widely used method for monitoring and investigating structural damage, but it suffers from fundamental limitations. This approach heavily relies on the expertise of the inspector and their empirical knowledge of concrete structures [7,8], typically performed without specialized equipment and directly by the human eye, visual inspections require a high degree of attention and scrutiny [7]. However, the subjective judgment inherent in this process makes it inefficient in terms of both cost and accuracy [9]. A comprehensive study conducted by the NDEVC (Nondestructive valuation Validation Center) of the U.S. Federal Highway Administration in 2001 empirically demonstrated that both routine and in-depth visual inspections consistently provided inaccurate results [6]. Furthermore, this method is time-consuming and heavily reliant on the competence of individual inspectors [10,11], requiring significant human resource allocation and specialized training to conduct safety assessments of increasingly aging buildings. These cumulative drawbacks—high labor costs, slow inspection speeds, inspector-dependent variability, and limited accuracy—underscore the urgent need for automated detection systems [12,13].

Building damage can lead to numerous problems. Cracks should be evaluated during concrete structure safety inspections for both durability and aesthetic reasons. In Massachusetts's Metropolitan Boston area, more than 70% of 135 surveyed homes showed heat loss through cracks when investigated using infrared technology. Approximately 40% of energy loss in buildings is attributed to heat transfer through cracks and ducts, as well as air leakage [14]. Early crack identification therefore offers immediate energy-efficiency benefits alongside durability gains. In the case of exposed rebar, it corrodes more quickly than cement-protected rebar. This corrosion significantly reduces the cross-sectional area of the rebar and impairs its mechanical properties, thereby reducing its load-bearing capacity [9,15]. Consequently, timely detection and remediation of rebar exposure are critical to prevent progressive structural failure.

The continued occurrence of such cracks and rebar exposure ultimately leads to spalling and delamination on the concrete surface. Furthermore, there have been cases in the past where concrete fragments fell from the underside of overpasses, posing a danger to drivers and even resulting in fatalities. Therefore, early detection and remedial measures for concrete spalling are necessary to prevent uncontrolled deterioration and damage in buildings [16]. Despite the significance of these critical defects, current safety assessments heavily rely on error-prone and subjective visual inspections.

Recent advances in deep learning have shown promising results. However, existing methods face three critical gaps: (1) limited performance in noisy real-world conditions, (2) lack of severity assessment beyond binary detection, and (3) absence of risk prioritization frameworks for maintenance planning. To conduct safety assessments of increasingly aging buildings, it is essential to cultivate a skilled workforce, and this necessitates specialized training. In particular, visual inspections for safety assessments are time-consuming and heavily reliant on the competence of the inspector [10,11]. The drawbacks of such visual inspections include high costs, time requirements, and the need for significant human resource allocation.

Encoder-decoder architectures have demonstrated effectiveness in pavement crack segmentation, with hierarchical feature learning approaches achieving F1-scores above 0.93 [17]. Comprehensive reviews of AI applications in structural health monitoring have identified K-means clustering, support vector machines, and deep learning as primary pattern recognition techniques [13]. However, these methods typically focus on binary detection or pixel-level segmentation without quantifying damage severity or providing risk-based prioritization frameworks for maintenance planning.

To address these limitations, this study offers four key contributions: The primary objective is to achieve real-time detection accuracy exceeding 0.75 mAP while maintaining inference speeds suitable for field deployment (>30 FPS). The results demonstrate that YOLOv11x achieves 0.78 mAP@0.5 at 101 FPS, while lightweight models (n, s) deliver 0.688–0.757 mAP@0.5 at >350 FPS. (1) a YOLOv11-based real-time detector that recognises three critical damage categories—cracks, rebar exposure, and spalling, (2) a Damage Criticality Index (DCI) that integrates five visual–spatial indicators into a continuous severity scale, (3) a publicly released field dataset of 130,838 images from aging residential buildings, (4) an end-to-end, risk-prioritised maintenance framework that operates without expert intervention. Collectively, these contributions establish a scalable, data-driven pipeline for the safety monitoring of aging infrastructure.

The remainder of this paper is organised as follows: In Section 2, we review previous research related to safety assessments, examining the directions and limitations of existing studies. Section 3 describes the proposed dataset, deep-learning framework, and DCI equation. Section 4 presents the training and testing results, confirms the model's performance based on identified evaluation metrics, and verifies the actual detection outcomes. Section 5 concludes this paper and outlines avenues for future research.

2. Literature Review

In recent years, there have been significant advancements in computer vision technology in the field of surface damage detection [18–20]. This literature review aims to explore a wide range of studies, including research on the identification and assessment of structural damage in buildings and infrastructure. This review encompasses studies that employ both traditional computer vision methods for safety assessments and those utilizing deep learning technology, with a focus on assessing their applicability.

The endeavor to detect specific objects from images has a history of ongoing research, with various studies harnessing computer vision technology to identify surface damage [21]. O'Byrne, et al. [22] explored methods for detecting damage by analyzing various factors such as brightness, color, and texture. Karimi and Asemani [23] also developed methods for damage detection through shape and repetitive pattern recognition. Nishikawa, et al. [7], Fujita and Hamamoto [24], Sinha and Fieguth [25] investigated filtering images to remove noise or emphasize specific areas to detect damage with images that have noise or share similarities with the surrounding environment. However, traditional methods that binarize images to distinguish cracks are challenging to apply in noisy data with various characteristics [26].

In recent years, with the growing popularity of deep learning technology, there has been a noticeable shift towards using neural networks for surface damage detection. LeCun, et al. [27] and Cha, et al. [28] demonstrated the potential of deep learning approaches to achieve higher accuracy compared to traditional methods. Munawar, et al. [29], Kim and Cho [30], Dorafshan, et al. [31], Li and Zhao [32] applied an algorithm that specifically uses Convolutional Neural Networks (CNN) to effectively identify concrete damage, showing high performance and applicability. Perez, et al. [33], Dung [34] demonstrated that it is possible to identify defects in buildings based on the VGG-16 CNN or FCN (Fully Convolutional Network), allowing for a rapid assessment of building conditions. Ren, et al. [35], Ali, et al. [36], Chen, et al. [37] proved the reliability and efficiency of using deep learning models in the context of maintaining the exterior of buildings. Alipour, et al. [38] successfully detected cracks at the pixel level in images with low noise. Cumbajin, et al. [39] reviewed literature across various fields on surface defect detection and found that the use of CNN models is increasing annually and will continue to be considered an excellent research topic.

Various deep learning models have emerged recently, including models with high accuracy, fast speed, and those considering both accuracy and speed. Among them, YOLO has been widely used for detection and real-time detection due to its fast speed and decent detection capability. Sun, et al. [40] proposed the detection of defects in boiler inner walls based on YOLOv5 and data augmentation techniques and demonstrated that the YOLOv5 model achieved better results compared to Faster R-CNN, SSD, and RetinaNet. Ping, et al. [41] compared the performance of SSD, HOG with SVM, and Faster R-CNN for the detection of potholes, showing that the YOLO model performed best in delivering faster and more reliable detection results. Ahmed [42] in his study, comparing the performance of Faster R-CNN with a modified VGG16 as the backbone, YOLOv5 (Large, Medium, Small) models with ResNet101 as the backbone, and Faster R-CNN using ResNet50, VGG16, MobileNetV2, InceptionV3, revealed that the YOLOv5 Small model is more suitable for detecting potholes.

The YOLO architecture has continuously evolved since 2016, with each iteration improving real-time detection capabilities for structural health monitoring applications. Recent comparative benchmarking studies have demonstrated YOLO11's state-of-the-art performance relative to both previous YOLO versions and alternative architectures Sharma, et al. [43]. conducted a comprehensive comparison of YOLOv8, YOLOv9, YOLOv10, YOLOv11, and Faster R-CNN, revealing that YOLO11 variants achieved superior speed-accuracy trade-offs, with the YOLO11l model demonstrating particularly strong performance in balancing detection precision with inference speed (>200 FPS on GPU). A systematic benchmark across diverse datasets by Jegham, et al. [44] confirmed that YOLO11m, YOLO11n, and YOLO11s emerged as the most consistent performers when considering accuracy, computational efficiency (GFLOPs), and model size—outperforming both YOLOv9 and YOLOv10 across multiple metrics.

While transformer-based detectors such as RT-DETR Zhao, et al. [45] temporarily challenged YOLO's dominance in real-time detection, subsequent YOLO iterations including v11 and v12 have reclaimed performance leadership through architectural innovations such as the C3k2 block, C2PSA attention mechanism, and optimized detection heads Khanam and Hussain [46]. The availability of five model variants (n, s, m, l, x) in YOLO11 enables flexible deployment ranging from embedded devices (YOLO11n at ~170 FPS) to server-based systems (YOLO11x at ~240 FPS with highest accuracy) Kishor [47], addressing the practical requirement for scalable infrastructure inspection solutions. In structural damage detection, YOLO11-based approaches have demonstrated robust performance on concrete crack datasets, achieving >93% mAP@0.5 while maintaining real-time inference speeds Huang, et al. [48].

These findings collectively establish YOLO11 as a validated architecture for field deployment in aging infrastructure monitoring, where the balance between detection accuracy, inference speed, and deployment flexibility is critical. Although direct comparison is difficult due to the varying characteristics of the datasets, it is evident that datasets with complex and noisy data tend to show relatively lower accuracy. However, this study recorded improved MAP results compared to previous studies using YOLO models, despite utilizing relatively complex and noisy data.

Senthilnathan [49] has indicated that with these advancements, deep learning algorithms have the potential to replace the human visual inspection process. However, despite these developments, applying this technology in real-world environments remains a challenging task, as many studies have primarily relied on data collected in controlled conditions, which may not fully encompass the complexity and variability of real-life environments. These limitations could pose challenges when applying such methods to safety assessments in actual buildings. Moreover, another notable limitation in prior studies lies

in their binary approach to damage detection—primarily focusing on whether damage exists, rather than evaluating its severity. Most existing research has emphasized improving detection accuracy and robustness under varied conditions but has seldom addressed how detected damages should be prioritized for maintenance planning or risk mitigation.

Luo, et al. [50] conducted a systematic review of CV-based bridge inspection methods and identified three critical limitations: (1) susceptibility of detection systems to hardware and environmental factors, (2) computational inefficiency hindering real-time detection despite improvements through lightweight models, and (3) insufficient focus on severity assessment beyond binary detection. These findings align with our observations that existing methods primarily address whether damage exists rather than quantifying its criticality for maintenance prioritization.

To overcome these gaps, we introduce a novel metric, the Damage Criticality Index (DCI), which integrates multiple visual and spatial features—such as damage type, area, confidence, multiplicity, and density—into a unified score that quantifies the severity and urgency of the detected damage. This scoring system enables a more prioritized and risk-informed safety assessment, especially valuable in aging infrastructure management. Inspired by Rashidi, et al. [51], who proposed a multi-criteria Priority Index (PI) for bridge condition ranking integrating structural and functional efficiency, the proposed DCI integrates multiple weighted features of damage for automated prioritization in safety diagnostics. The probabilistic risk ranking model presented by Dong and Frangopol [52] provides a foundation for mapping detection features to actionable risk priorities, a concept integrated in the proposed DCI equation for automated inspection feedback. Following Gattulli and Chiaramonte [53] probabilistic treatment of condition indices under inspection uncertainty, our DCI score quantifies severity through normalized statistical deviations of area, density, and confidence metrics.

In addition, to enhance applicability under real-world complexities, this study uses a dataset consisting of approximately 130,000 images collected from residential buildings in Seoul. Azimi, et al. [19] and Azhari, et al. [54] have highlighted the importance of training with data from noisy and diverse environments. Therefore, this work not only addresses model robustness but also incorporates a scoring mechanism that enhances the interpretability and practical usefulness of automated damage assessments.

In summary, this literature review reveals critical gaps in current damage detection research. First, existing methods primarily focus on binary detection (damage/no damage) without assessing severity levels necessary for maintenance prioritization. Second, most studies validated their models in controlled environments with limited noise, failing to address the complexity of real-world conditions. Third, while detection accuracy has improved, there remains an absence of systematic frameworks for translating detection results into actionable risk categories.

This study addresses these limitations by developing a comprehensive framework that: integrates multi-class damage detection robust to noisy field conditions using YOLOv11, introduces the Damage Criticality Index (DCI) to quantify severity based on multiple visual-spatial features, and employs K-means clustering validated through statistical metrics to automatically categorize damages into risk-based maintenance priorities. This progression from detection to quantification to prioritization represents a complete pipeline for practical infrastructure management.

3. Utilizing Deep Learning for Safety Assessment

This section proposes a damage detection method using the YOLOv11 model to address the limitations of conventional visual inspection. Unlike traditional approaches that rely on inspectors' visual assessment, the proposed automated method offers an efficient al-

ternative or supplementary solution. Furthermore, a quantitative metric DCI is introduced to assess the severity of detected damages, allowing for risk-based decision-making. This enables inspectors and facility managers to perform objective risk management based on quantified indicators.

3.1. YOLOv11

The YOLO object detection model has undergone continuous improvements, culminating in the official release of YOLOv11. This version enhances feature extraction capabilities by incorporating C3k2 blocks, Spatial Pyramid Pooling–Fast (SPPF), and Convolutional Block with Parallel Spatial Attention (C2PSA) [46].

Compared to earlier YOLO versions, YOLOv11 demonstrates superior performance in both mean Average Precision (mAP) and recall metrics [55].

YOLOv11 was selected as the detection backbone for this framework based on validated performance in recent benchmarking studies [43,44]. As demonstrated in Literature Review, YOLO11 achieves optimal balance between detection accuracy and inference speed compared to alternative architectures including Faster R-CNN and transformer-based detectors. The availability of five model variants (n, s, m, l, x) enables flexible deployment across different hardware constraints, from resource-limited field devices to high-performance server systems—a critical requirement for scalable infrastructure inspection programs.

3.2. Proposed Inspection Framework

As illustrated in Figure 1, safety inspections begin with data acquisition via drones or handheld cameras. This data collection is a fundamental aspect of safety assessments, as it is crucial in obtaining accurate information from the field, ensuring safety, and accurately understanding the condition of the building.

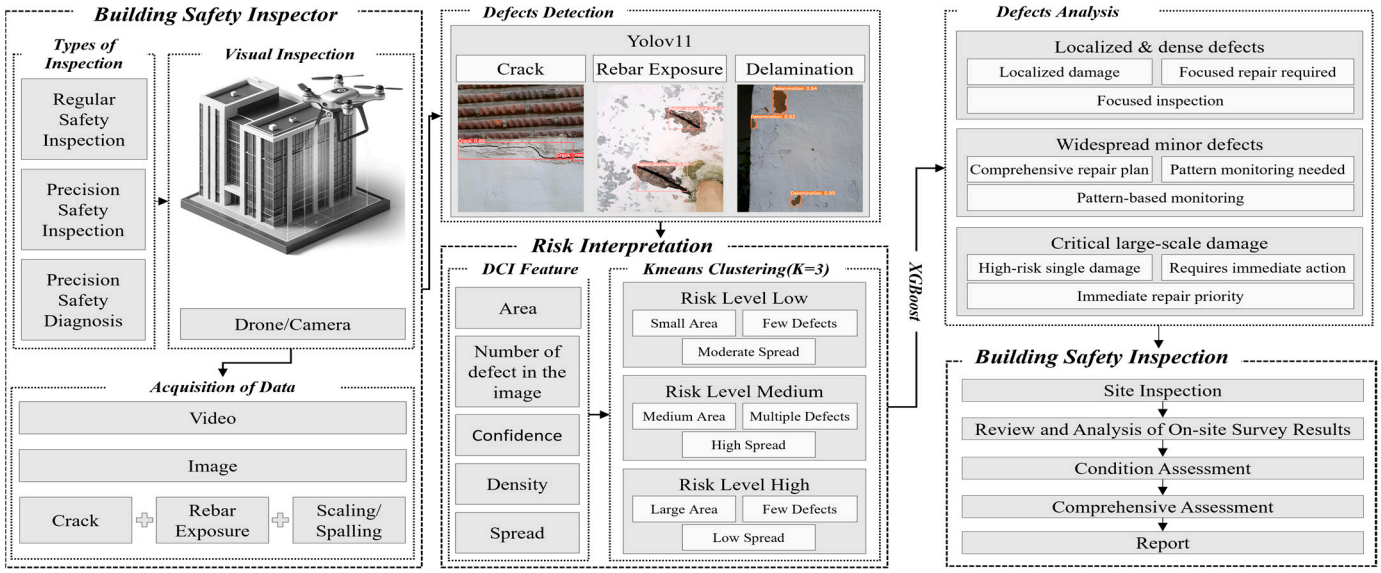


Figure 1. The framework of the proposed model for safety inspection.

The data collected in this manner is then utilized in conjunction with a deep learning engine to identify and detect various types of damage. This process is automated, minimizing human error and providing rapid results. By leveraging deep learning algorithms, even damage that is not easily detectable by the human eye can be detected, enabling a more accurate assessment of the types of damage present in the building.

Following the detection of damage, the subsequent step, "Review and analysis of on-site survey results," involves a thorough examination and analysis of the collected data.

This deepens the understanding of the building's safety status and aids in determining the necessary actions. In this phase, human judgment and expertise are combined to derive the final results.

Next, in the "Condition assessment" and "Comprehensive assessment" phases, the building's condition is evaluated comprehensively. In these phases, the building's structure, materials, facilities, and more are assessed based on the safety inspection results, allowing for the identification of the current state and potential risks. This assessment provides crucial information for the long-term safety and maintenance of the building.

Finally, in the "Report" phase, the safety inspection results are documented, and a report is created. This report is provided to building owners, operators, government agencies, and others, informing them of the safety inspection results and serving as valuable information when planning necessary actions and repair work.

The ultimate goal of this framework is to reduce the time required for visual inspections and minimize human error. This ensures that the safety assessment process operates more efficiently and that building safety and maintenance are carried out with greater precision.

3.3. Dataset Construction

To achieve accurate damage detection for safety assessment, having a sufficient quantity and quality of training data is crucial. Hence, this study, we utilized various data sources from AIhub to obtain image data for safety assessment of apartments, row houses, and non-residential houses that are over 20 years old and located in 25 districts within Seoul, South Korea. The data provided by AIhub is available for use by individuals of South Korean nationality and can be used at any time as long as the source of the data is cited.

The collected data consists of a total of 130,838 high-resolution images (1080 $\times$ 1440), with each image accurately representing damage types such as cracks, exposed rebar, and delamination. The dataset contains 327,372 damage instances across all images: 113,297 instances of cracks, 121,416 instances of delamination, and 92,659 instances of rebar exposure, with an average of 2.50 instances per image. Additionally, for object detection, bounding boxes are used to clearly indicate the locations of the damage. This dataset plays a vital role in enabling the model to identify various damage types and pinpoint their locations accurately.

Out of the entire dataset, 104,670 images (80%) were used as training data for the model, and the remaining 26,168 images (20%) were randomly selected for validation data to evaluate the model's performance. The training set contains 261,809 damage instances (90,688 cracks, 97,219 delaminations, and 73,902 rebar exposures), while the validation set contains 65,563 instances (22,609 cracks, 24,197 delaminations, and 18,757 rebar exposures). Splitting the data into training and validation sets is essential for improving the model's generalization capability and ensuring the accuracy of safety assessments.

This data collection and partitioning are critical steps for model training and evaluation, which are essential for guaranteeing the accuracy and efficiency of safety assessments. Sufficient data and training processes will enhance the quality of safety assessments and lead to more accurate results.

3.4. Risk Interpretation

For each damage instance detected by YOLOv11, both spatial and probabilistic features were quantitatively extracted to interpret severity. Five DCI features were defined and used for risk classification: (1) Area, (2) Number of detections in the image, (3) Confidence, (4) Density, and (5) Spread. These features, aggregated per image, reflect the

number, spatial distribution, and confidence of damage detections. After normalization, the optimal number of clusters was determined using both the Silhouette Score and the Elbow Method analysis. The average feature values of each cluster were analyzed to define Risk Levels 0 to 2. The interpretation and management strategies for each cluster are discussed in the following section. This procedure extends beyond mere damage detection and serves as a foundation for risk-informed structural assessment and future maintenance decision-making.

$$DCI = \omega_1 \times A_{norm} + \omega_2 \times M + \omega_3 \times \mu_{\ conf} + \omega_4 \times \rho + \omega_5 \times \sigma^2 \quad (1)$$

$$\begin{aligned} A_{norm} &= Normalized\ area\ (0-1) \\ M &= Number\ of\ detect\ in\ the\ image \\ \mu_{\ conf} &= Confidence \\ \rho &= Density \\ \sigma^2 &= Spread \\ \omega &= DCI\ feature\ importance \end{aligned}$$

To quantify the relative importance of each DCI feature, a supervised learning approach was employed. K-means cluster assignments (k = 3) served as pseudo-labels for training an XGBoost classifier, https://github.com/dmlc/xgboost (accessed on 27 March 2024), which then provided feature importance scores based on information gain. This approach ensures that importance weights reflect actual contribution to risk level separation.

4. Results

4.1. YOLOv11 Detection

Accuracy is an indicator of how the model predicts across all classes. It is the number of correctly predicted data divided by the total amount of data. Accuracy is defined by Equation (2) as follows:

$$accuracy = \frac{True\ Postive + True\ Negative}{True\ Postive + False\ Negative + True\ Negative + False\ Postive} \quad (2)$$

But in general, accuracy has a characteristic in which the effects of *True Positive* (*TP*) and *True Negative* (*TN*) on accuracy are not the same. Therefore, it is difficult to represent the performance of the model only with accuracy; so, it should be confirmed through additional indicators.

To overcome these limitations, we utilized the *F*1-*score*, which combines *precision* and *recall*, and the *Average Precision* (*AP*), which is the product of *precision* and *recall*. Each equation is as follows:

$$Precision\ = \frac{True\ Postive}{True\ Postive + False\ Postive} \quad (3)$$

$$Recall = \frac{True\ Postive}{True\ Postive + False\ Negative} \quad (4)$$

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$AP = \int_0^1 P(R)dR \quad (6)$$

Table 1 summarizes and compares the performance of five YOLOv11 variants (n, s, m, l, x) in detecting three types of building damage: Crack, Rebar Exposure, and Delamination. Evaluation metrics include class-wise and overall scores for *Precision*, *Recall*, *F*1-*score*,

mAP@0.5, and mAP@0.5:0.95. Although some variation existed across classes, YOLOv11x achieved the highest overall performance (Precision: 0.788, Recall: 0.723, F1-score: 0.753, mAP@0.5: 0.78, mAP@0.5:0.95: 0.592). Performance gains were most pronounced up to YOLOv11m, after which marginal increases in mAP@0.5:0.95 (less than 1.4 percentage points) suggested diminishing returns. In terms of inference speed, YOLOv11x achieved 101 FPS (9.9 ms), enabling real-time processing even for 60 FPS video streams. Lightweight models YOLOv11n and YOLOv11s reached approximately 370 FPS (2.7 ms) and 357 FPS (2.8 ms), respectively, making them suitable for resource-constrained field devices. Inference speeds were measured on NVIDIA RTX 3090Ti GPU (24 GB VRAM) with CUDA 12.6, Intel i9-10900K CPU, and 64 GB RAM, representing a typical high-performance workstation configuration. Therefore, model selection can be guided by application requirements YOLOv11x for high-accuracy server-based inspections and YOLOv11n/s for, e.g., embedded deployments.

Table 1. Training Results of YOLOv11 Models.

Model	Class	Precision	Recall	F1-Score	mAP [0.5]	mAP [0.5:0.95]
YOLOv11n	Crack	0.75	0.62	0.676	0.622	0.422
	Rebar exposure	0.79	0.548	0.646	0.714	0.479
	Delamination	0.759	0.644	0.697	0.728	0.562
	All	0.766	0.604	0.673	0.688	0.488
YOLOv11s	Crack	0.768	0.648	0.703	0.705	0.509
	Rebar exposure	0.824	0.707	0.760	0.78	0.544
	Delamination	0.782	0.724	0.752	0.785	0.628
	All	0.791	0.693	0.738	0.757	0.56
YOLOv11m	Crack	0.775	0.667	0.717	0.723	0.536
	Rebar exposure	0.83	0.73	0.776	0.8	0.567
	Delamination	0.792	0.726	0.757	0.795	0.647
	All	0.799	0.708	0.75	0.773	0.583
YOLOv11l	Crack	0.769	0.671	0.717	0.728	0.545
	Rebar exposure	0.827	0.733	0.777	0.8	0.57
	Delamination	0.791	0.732	0.76	0.799	0.654
	All	0.796	0.712	0.751	0.776	0.59
YOLOv11x	Crack	0.759	0.686	0.72	0.731	0.544
	Rebar exposure	0.829	0.733	0.778	0.802	0.571
	Delamination	0.776	0.749	0.762	0.806	0.66
	All	0.788	0.723	0.753	0.78	0.592

The selection of YOLOv11 for this framework was based on comprehensive benchmarking studies reported in the literature. Sharma, et al. [43] demonstrated YOLOv11's superior speed-accuracy trade-offs compared to YOLOv10, YOLOv9, and Faster R-CNN across multiple datasets, while Jegham, et al. [44] confirmed YOLOv11's consistent performance advantages over its predecessors. Direct comparison on our specific dataset was not conducted as the primary contribution of this study lies in the development of the DCI framework and risk prioritization methodology rather than architectural benchmarking. The proven performance of YOLOv11 in similar structural damage detection tasks [48] provided sufficient justification for its adoption as the detection backbone.

In this study, which aims to replace visual inspections, in diverse environments, damage detection was conducted using the YOLOv11n~x model. As shown in Figure 2, the results demonstrate accurate damage detection in a relatively clean environment with no noise.

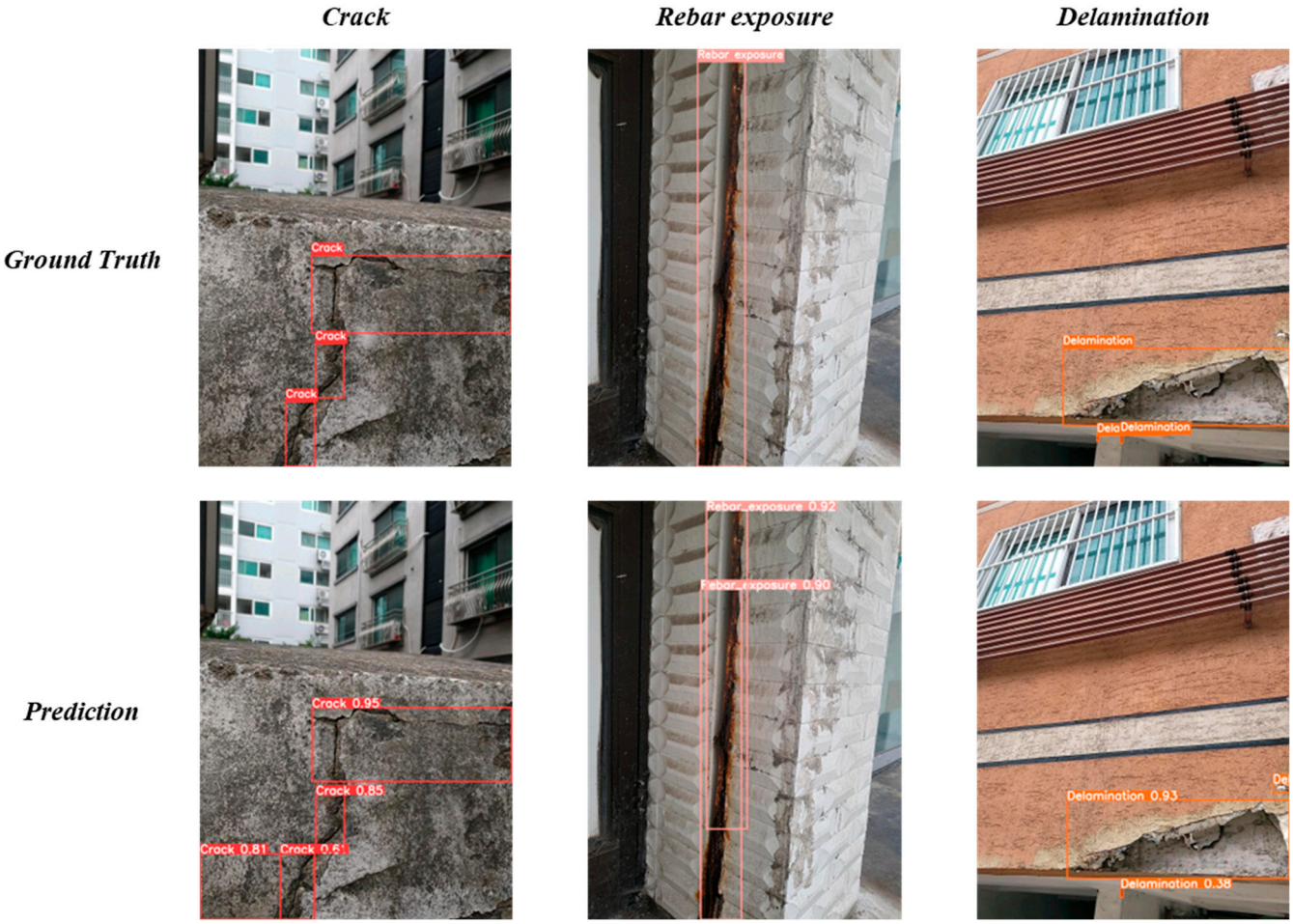


Figure 2. Damage Detection in Low-Noise Images.

Furthermore, Figures 3 and 4 compare the ground truth annotations with the model predictions across various environmental settings. The detection results demonstrate the capability to identify damage relatively accurately in a range of environments, including (a) contaminated surroundings, (b) areas where false positives can easily occur due to proximity to adjacent parts of the building, (c) complex situations involving rebar exposure and delamination, (d) environments with brick patterns that make false positives easy, (e) damage of various sizes and shapes, and (f) environments with surroundings resembling rebar exposure due to wires.



Figure 3. Damage detection in various environments (Ground Truth).



Figure 4. Damage detection in various environments (Prediction).

4.2. DCI Feature Importance

Figure 5 illustrates the relative importance of the five DCI components—Normalized Area, Multiplicity, Average Confidence, Density Score, and Spread Score—across YOLOv11 models (n, s, m, l, x).

- Multiplicity consistently showed the highest importance across all models, with YOLOv11x recording the highest weight (0.543). This underscores the critical role of the number of detections in assessing structural risk. Even smaller models maintained an average weight above 0.40, confirming its dominant contribution.
- Spread Score was the second most influential feature, especially in YOLOv11n (0.361) and YOLOv11x (0.378), suggesting that wider damage dispersion is strongly associated with severity.

- Normalized Area contributed moderately (typically 0.17–0.20), though YOLOv11l showed a notably low value (0.012), indicating limited impact of physical damage size in that model.
- Average Confidence was generally low across models but spiked in YOLOv11l (0.312), implying that medium-sized models rely more on prediction certainty, whereas YOLOv11x showed reduced dependency (0.063).
- Density Score ranked lowest in importance, remaining between 0.01 and 0.05 for most models except YOLOv11l (0.093), suggesting that average inter-damage distance contributes less than damage count or spread in severity estimation.

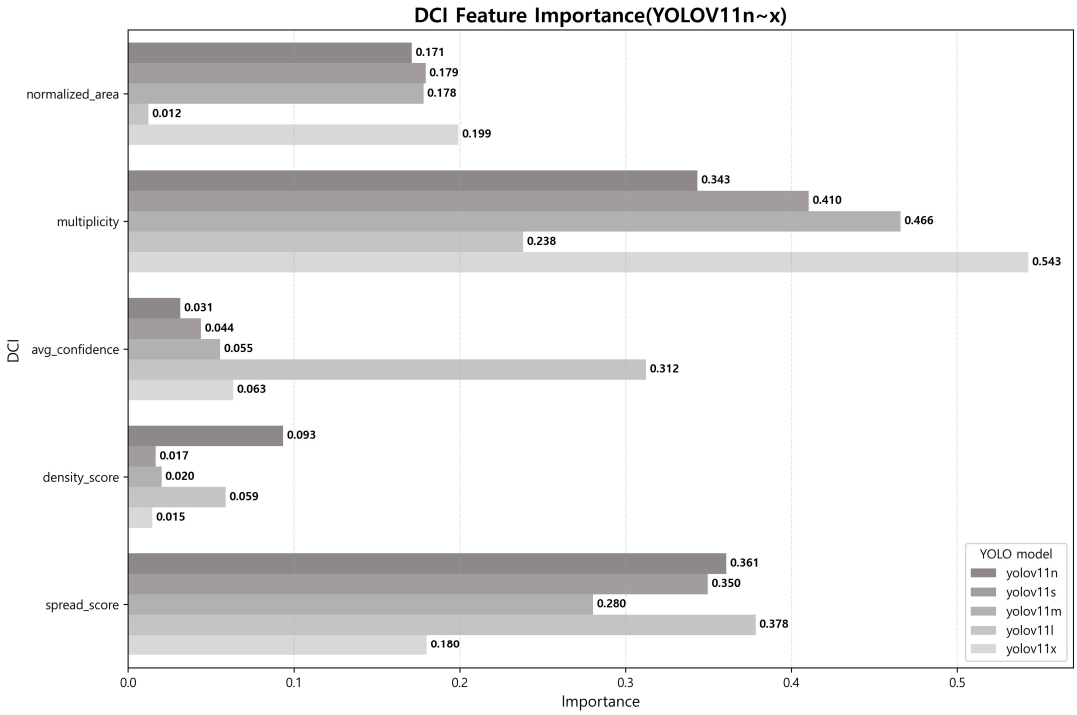


Figure 5. DCI Feature Importance (YOLOv11n~x) calculated using XGBoost with K-means derived risk levels as training targets.

4.3. K-Means Based Risk Level Distribution

Prior to risk level analysis, optimal cluster number was determined through comprehensive evaluation across all YOLOv11 variants. Figure 5 presents the Elbow curves (inertia) and Silhouette scores for k values ranging from 2 to 6.

The Elbow Method consistently showed significant inertia reduction at k = 3 across all models, with an average reduction of 31.2% from k = 2 to k = 3, compared to only 15.8% from k = 3 to k = 4. While Silhouette scores varied by model—YOLOv11n and YOLOv11x favoring k = 4 (0.534, 0.511), YOLOv11s preferring k = 3 (0.488), and YOLOv11m/l showing highest scores at k = 2 (0.495, 0.782)—the decision to adopt k = 3 was based on three considerations:

1. Clear elbow points at k = 3 in all inertia curves
2. Alignment with industry-standard three-tier risk classification (Low, Medium, High) [56]
3. Practical interpretability for maintenance prioritization

Notably, YOLOv11l exhibited anomalously high silhouette scores due to overemphasis on confidence features, creating artificial separation that did not reflect actual damage

severity. Therefore, the consensus k = 3 was applied uniformly across all models to ensure consistent risk assessment.

Figure 6 shows cluster validation metrics for optimal k selection across YOLOv11 variants. The left panels show Elbow curves (within-cluster sum of squares), while the right panels display Silhouette coefficients. The selected k = 3 (highlighted) balances statistical optimization with practical applicability.

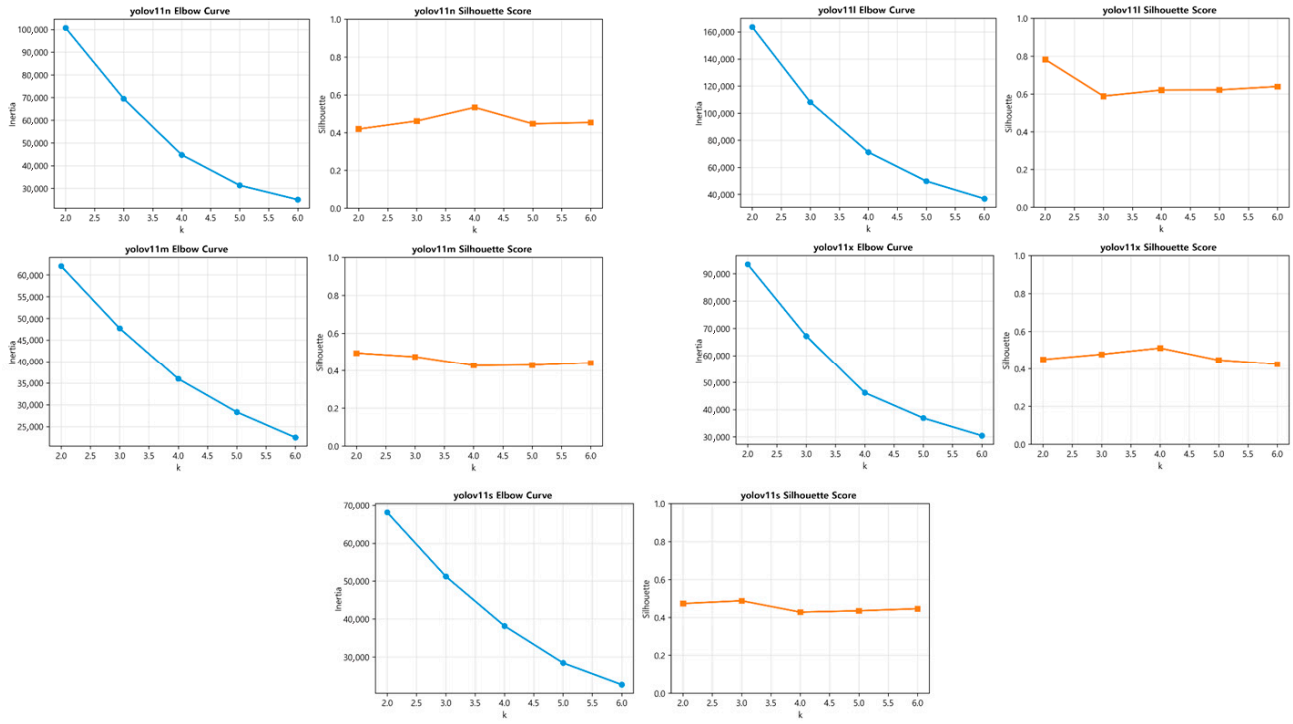


Figure 6. Elbow, Silhouette score.

Figure 7 visualizes the PCA projection (PC1 and PC2) of DCI values and the K-means (k = 3) clustering results. PC1 is largely explained by Multiplicity and Spread Score, while PC2 is influenced by Normalized Area and Average Confidence—thus representing damage count/distribution and area/confidence, respectively.

- Risk Level 0 (Low Risk) is located along the negative PC1 axis, corresponding to small, sparsely distributed damages with limited area.
- Risk Level 1 (Medium Risk) shows compact clusters for YOLOv11n, s, and x, but overlaps with Risk 0 in models m and l due to elevated confidence scores, which blur cluster boundaries.
- Risk Level 2 (High Risk) is concentrated along the positive PC1 axis in all models, with YOLOv11s and x exhibiting the clearest separation.
- YOLOv11l showed degraded clustering quality, forming “tails” along the PC2 axis due to overemphasis on confidence values.

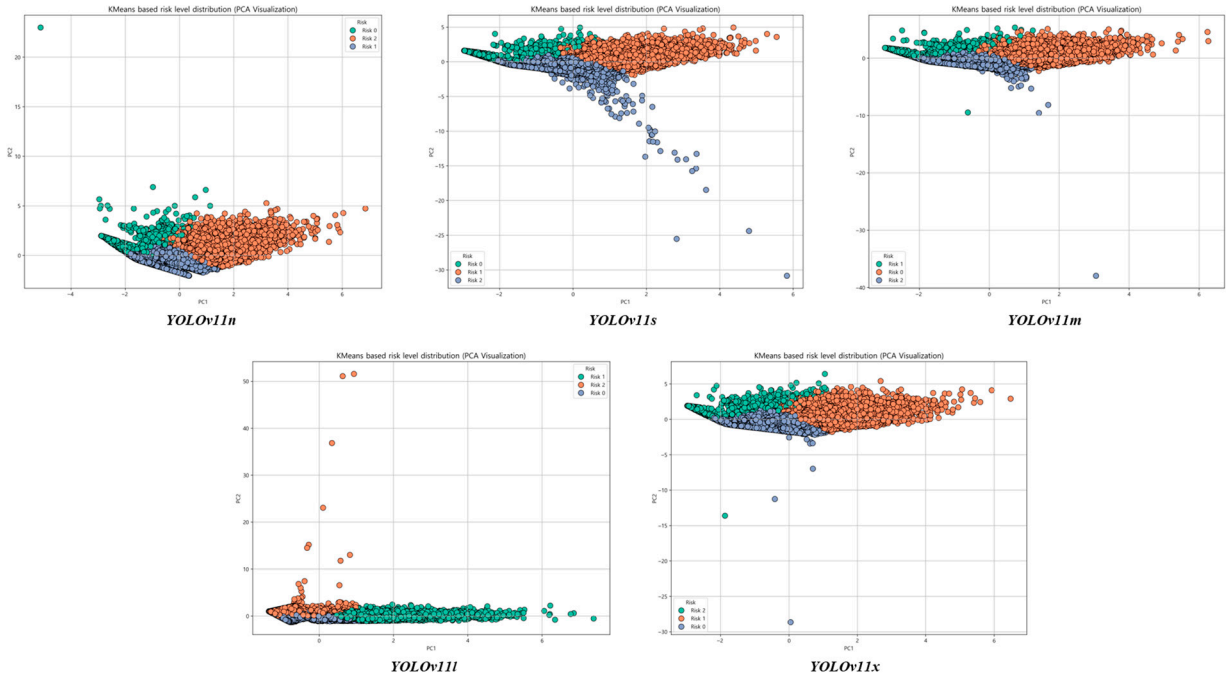


Figure 7. K-Means based risk level distribution.

5. Conclusions

This study presented an automated damage detection and risk prioritization framework for aging concrete structures by integrating YOLOv11 deep learning models with a novel Damage Criticality Index (DCI).

The key achievements include: (1) multi-class damage detection achieving 0.78 mAP@0.5 (YOLOv11x) with real-time processing capability (101 FPS), demonstrating practical feasibility for field deployment; (2) development of the DCI incorporating five visual-spatial features, with multiplicity (weight: 0.543) and spread (0.378) emerging as primary severity indicators; (3) successful risk stratification through K-means clustering (k = 3), validated by silhouette analysis (0.478–0.488) and aligned with industry-standard three-tier classification.

The framework demonstrated robust performance across diverse environmental conditions including contaminated surfaces, complex backgrounds, and varying lighting conditions, addressing the limitations of traditional visual inspection methods that suffer from subjectivity and human error.

Future research will focus on three key areas: (1) enhancing detection precision through segmentation-based approaches, which can provide pixel-level damage boundaries for more accurate area quantification and improved DCI calculation; (2) conducting comprehensive architectural comparisons including YOLOv10 and transformer-based detectors (e.g., RT-DETR) on our specific dataset to validate the optimal detection backbone; and (3) establishing correlations between DCI scores and real-world damage progression through longitudinal field studies. This transition from bounding-box detection to instance segmentation is expected to significantly improve damage assessment accuracy, particularly for irregular crack patterns and overlapping damage regions. Additionally, empirical validation of the DCI against actual structural deterioration patterns will strengthen the framework's practical applicability, ultimately enabling more precise and evidence-based maintenance planning for aging infrastructure.

Author Contributions: Conceptualization, J.H.; Methodology, J.H.; Validation, Y.A. and H.S.; Investigation, J.H.; Writing—original draft, J.H.; Writing—review & editing, J.H.; Supervision, H.S.; Project administration, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2021R1I1A1A01059736).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This paper used datasets from Crack Data of Old Houses in Seoul (AI-Hub, S. Korea). All data information can be accessed through AI-Hub, https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=567 (accessed on 2 August 2022).

Conflicts of 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.

References

1. Statistics on the Status of Buildings as of 22 Years, 17 November 2023. Available online: https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?lcmspage=1&id=95087983 (accessed on 27 March 2024).
2. AGC. Eighty Percent of Contractors Report Difficulty Finding Qualified Craft Workers to Hire as Association Calls for Measures to Rebuild Workforce. 2018. Available online: https://www.agc.org/news/2018/08/29/eighty-percent-contractors-report-difficulty-finding-qualified-craft-workers-hire-0 (accessed on 27 March 2024).
3. Menches, C.L.; Abraham, D.M. Women in Construction—Tapping the Untapped Resource to Meet Future Demands. *J. Constr. Eng. Manag.* **2007**, *133*, 701–707. [CrossRef]
4. Kang, T.W. Study on 3D image scan-based MEP facility management technology. *J. KIBIM* **2016**, *6*, 18–26. [CrossRef]
5. Ham, N.; Lee, S.-H. Empirical study on structural safety diagnosis of large-scale civil infrastructure using laser scanning and BIM. *Sustainability* **2018**, *10*, 4024. [CrossRef]
6. Nishikawa, T.; Yoshida, J.; Sugiyama, T.; Fujino, Y. Concrete crack detection by multiple sequential image filtering. *Comput.-Aided Civ. Infrastruct. Eng.* **2012**, *27*, 29–47. [CrossRef]
7. Flah, M.; Suleiman, A.R.; Nehdi, M.L. Classification and quantification of cracks in concrete structures using deep learning image-based techniques. *Cem. Concr. Compos.* **2020**, *114*, 103781. [CrossRef]
8. Hoang, N.-D. Detection of surface crack in building structures using image processing technique with an improved Otsu method for image thresholding. *Adv. Civ. Eng.* **2018**, *2018*, 3924120. [CrossRef]
9. Moore, M.; Phares, B.M.; Graybeal, B.; Rolander, D.; Washer, G. *Reliability of Visual Inspection for Highway Bridges, Volume I: Final Report*; Report No. FHWA-RD-01-020; Federal Highway Administration: Washington, DC, USA, 2001.
10. Kim, B.; Cho, S. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Struct. Control Health Monit.* **2019**, *26*, e2381. [CrossRef]
11. Zinno, R.; Haghshenas, S.S.; Guido, G.; VItale, A. Artificial intelligence and structural health monitoring of bridges: A review of the state-of-the-art. *IEEE Access* **2022**, *10*, 88058–88078. [CrossRef]
12. Shao, E.C. Detecting Sources of Heat Loss in Residential Buildings from Infrared Imaging. Bachelor's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2011.
13. Papadopoulos, M.; Apostolopoulos, C.A.; Zervaki, A.; Haidemenopoulos, G. Corrosion of exposed rebars, associated mechanical degradation and correlation with accelerated corrosion tests. *Constr. Build. Mater.* **2011**, *25*, 3367–3374. [CrossRef]
14. Raja, B.N.K.; Miramini, S.; Duffield, C.; Sofi, M.; Zhang, L. Infrared thermography detection of delamination in bottom of concrete bridge decks. *Struct. Control Health Monit.* **2022**, *29*, e2886. [CrossRef]
15. Liang, X. Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 415–430. [CrossRef]
16. Sajedi, S.; Liang, X. A convolutional cost-sensitive crack localization algorithm for automated and reliable RC bridge inspection. In *Risk-Based Bridge Engineering: Proceedings of the 10th New York City Bridge Conference*; CRC Press: Boca Raton, FL, USA, 2019; Volume 2019, p. 229.
17. Fan, Z.; Li, C.; Chen, Y.; Wei, J.; Loprencipe, G.; Chen, X.; Di Mascio, P. Automatic crack detection on road pavements using encoder-decoder architecture. *Materials* **2020**, *13*, 2960. [CrossRef] [PubMed]
18. Deng, J.; Singh, A.; Zhou, Y.; Lu, Y.; Lee, V.C.-S. Review on computer vision-based crack detection and quantification methodologies for civil structures. *Constr. Build. Mater.* **2022**, *356*, 129238. [CrossRef]
19. Azimi, M.; Eslamlou, A.D.; Pekcan, G. Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors* **2020**, *20*, 2778. [CrossRef]
20. Feng, D.; Feng, M.Q. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review. *Eng. Struct.* **2018**, *156*, 105–117. [CrossRef]

21. Xie, X. A review of recent advances in surface defect detection using texture analysis techniques. *ELCVIA Electron. Lett. Comput. Vis. Image Anal.* **2008**, *7*, 1–22. [CrossRef]
22. O'Byrne, M.; Schoefs, F.; Ghosh, B.; Pakrashi, V. Texture analysis based damage detection of ageing infrastructural elements. *Comput.-Aided Civ. Infrastruct. Eng.* **2013**, *28*, 162–177. [CrossRef]
23. Karimi, M.H.; Asemani, D. Surface defect detection in tiling Industries using digital image processing methods: Analysis and evaluation. *ISA Trans.* **2014**, *53*, 834–844. [CrossRef] [PubMed]
24. Fujita, Y.; Hamamoto, Y. A robust automatic crack detection method from noisy concrete surfaces. *Mach. Vis. Appl.* **2011**, *22*, 245–254. [CrossRef]
25. Sinha, S.K.; Fieguth, P.W. Automated detection of cracks in buried concrete pipe images. *Autom. Constr.* **2006**, *15*, 58–72. [CrossRef]
26. Kim, H.; Ahn, E.; Cho, S.; Shin, M.; Sim, S.-H. Comparative analysis of image binarization methods for crack identification in concrete structures. *Cem. Concr. Res.* **2017**, *99*, 53–61. [CrossRef]
27. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
28. Cha, Y.J.; Choi, W.; Büyüköztürk, O. Deep learning-based crack damage detection using convolutional neural networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [CrossRef]
29. Munawar, H.S.; Ullah, F.; Heravi, A.; Thaheem, M.J.; Maqsoom, A. Inspecting buildings using drones and computer vision: A machine learning approach to detect cracks and damages. *Drones* **2021**, *6*, 5. [CrossRef]
30. Kim, B.; Cho, S. Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. *Sensors* **2018**, *18*, 3452. [CrossRef] [PubMed]
31. Dorafshan, S.; Thomas, R.J.; Maguire, M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Constr. Build. Mater.* **2018**, *186*, 1031–1045. [CrossRef]
32. Li, S.; Zhao, X. Image-based concrete crack detection using convolutional neural network and exhaustive search technique. *Adv. Civ. Eng.* **2019**, *2019*, 6520620. [CrossRef]
33. Perez, H.; Tah, J.H.; Mosavi, A. Deep learning for detecting building defects using convolutional neural networks. *Sensors* **2019**, *19*, 3556. [CrossRef]
34. Dung, C.V. Autonomous concrete crack detection using deep fully convolutional neural network. *Autom. Constr.* **2019**, *99*, 52–58. [CrossRef]
35. Ren, Y.; Huang, J.; Hong, Z.; Lu, W.; Yin, J.; Zou, L.; Shen, X. Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Constr. Build. Mater.* **2020**, *234*, 117367. [CrossRef]
36. Ali, L.; Alnajjar, F.; Jassmi, H.A.; Gocho, M.; Khan, W.; Serhani, M.A. Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors* **2021**, *21*, 1688. [CrossRef]
37. Chen, K.; Reichard, G.; Xu, X.; Akanmu, A. Automated crack segmentation in close-range building façade inspection images using deep learning techniques. *J. Build. Eng.* **2021**, *43*, 102913. [CrossRef]
38. Alipour, M.; Harris, D.K.; Miller, G.R. Robust pixel-level crack detection using deep fully convolutional neural networks. *J. Comput. Civ. Eng.* **2019**, *33*, 04019040. [CrossRef]
39. Cumbajin, E.; Rodrigues, N.; Costa, P.; Miragaia, R.; Frazão, L.; Costa, N.; Fernández-Caballero, A.; Carneiro, J.; Buruberri, L.H.; Pereira, A. A Systematic Review on Deep Learning with CNNs Applied to Surface Defect Detection. *J. Imaging* **2023**, *9*, 193. [CrossRef] [PubMed]
40. Sun, X.; Jia, X.; Liang, Y.; Wang, M.; Chi, X. A defect detection method for a boiler inner wall based on an improved YOLO-v5 network and data augmentation technologies. *IEEE Access* **2022**, *10*, 93845–93853. [CrossRef]
41. Ping, P.; Yang, X.; Gao, Z. A deep learning approach for street pothole detection. In Proceedings of the 2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, UK, 3–6 August 2020; pp. 198–204.
42. Ahmed, K.R. Smart pothole detection using deep learning based on dilated convolution. *Sensors* **2021**, *21*, 8406. [CrossRef]
43. Sharma, A.; Kumar, V.; Longchamps, L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. *Smart Agric. Technol.* **2024**, *9*, 100648. [CrossRef]
44. Jegham, N.; Koh, C.Y.; Abdelatti, M.; Hendawi, A. Evaluating the evolution of yolo (you only look once) models: A comprehensive benchmark study of yolo11 and its predecessors. *arXiv* **2024**, arXiv:2411.00201. [CrossRef]
45. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 16965–16974.
46. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
47. Kishor, R. Performance Benchmarking of YOLOv11 Variants for Real-Time Delivery Vehicle Detection: A Study on Accuracy, Speed, and Computational Trade-offs. *Asian J. Res. Comput. Sci.* **2024**, *17*, 108–122. [CrossRef]
48. Huang, S.; Liu, Q.; Chen, C.; Chen, Y. A Real-time Concrete Crack Detection and Segmentation Model Based on YOLOv11. *arXiv* **2025**, arXiv:2508.11517. [CrossRef]

49. Senthilnathan, R. Deep Learning in Vision-Based Automated Inspection: Current State and Future Prospects. In *Machine Learning in Industry*; Management and Industrial Engineering; Springer: Berlin/Heidelberg, Germany, 2021; pp. 159–175. [CrossRef]
50. Luo, K.; Kong, X.; Zhang, J.; Hu, J.; Li, J.; Tang, H. Computer vision-based bridge inspection and monitoring: A review. *Sensors* **2023**, *23*, 7863. [CrossRef]
51. Rashidi, M.; Samali, B.; Sharafi, P. A new model for bridge management: Part A: Condition assessment and priority ranking of bridges. *Aust. J. Civ. Eng.* **2016**, *14*, 35–45. [CrossRef]
52. Dong, Y.; Frangopol, D.M. Probabilistic time-dependent multihazard life-cycle assessment and resilience of bridges considering climate change. *J. Perform. Constr. Facil.* **2016**, *30*, 04016034. [CrossRef]
53. Gattulli, V.; Chiaramonte, L. Condition assessment by visual inspection for a bridge management system. *Comput.-Aided Civ. Infrastruct. Eng.* **2005**, *20*, 95–107. [CrossRef]
54. Azhari, F.; Sennersten, C.; Milford, M.; Peynot, T. PointCrack3D: Crack detection in unstructured environments using a 3D-point-cloud-based deep neural network. *arXiv* **2021**, arXiv:2111.11615.
55. He, Z.; Wang, K.; Fang, T.; Su, L.; Chen, R.; Fei, X. Comprehensive Performance Evaluation of YOLOv11, YOLOv10, YOLOv9, YOLOv8 and YOLOv5 on Object Detection of Power Equipment. *arXiv* **2024**, arXiv:2411.18871. [CrossRef]
56. Qazi, A.; Dikmen, I. From risk matrices to risk networks in construction projects. *IEEE Trans. Eng. Manag.* **2019**, *68*, 1449–1460. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.