
International Journal of
Applied Data Science in Engineering and Health
https://ijadseh.com

# Data-driven Detection and Evaluation of Damages in Concrete Structures: Using Deep Learning and Computer Vision

Saeid Ataei[1*], Saeed Adibnazari[2], Seyyed Taghi Ataei[3]

[1] *Stevens Institute of Technology, Hoboken, NJ, USA*

[2] *Sharif University of Technology, Tehran, Iran*

[3] *University of Tehran, Tehran, Iran*

Received date: October 23, 2025; Accepted date: January 2, 2026

**Abstract**

Maintaining the structural integrity of concrete infrastructure—such as buildings, bridges, and tunnels—is critical to ensuring public safety and uninterrupted operations. However, traditional inspection methods remain labor-intensive, inconsistent, and impractical for large-scale deployment. This research investigates the application of cutting-edge, vision-based deep learning models to automate damage detection in concrete structures, with a focus on cracks and spalling. Two state-of-the-art architectures—YOLO-v7 instance segmentation and Mask R-CNN—are evaluated using an augmented dataset of 10,995 annotated images derived from an initial set of 400 real-world samples. Both models were trained via transfer learning on the COCO dataset and assessed using precision, recall, mean average Precision at IoU 0.5 (mAP50), and inference speed (FPS). YOLO-v7 achieved a superior mAP50 of 96.1% and a real-time processing rate of 40 FPS, making it ideal for rapid field deployment. In contrast, Mask R-CNN delivered a strong mAP50 of 92.1% at 18 FPS, favoring high segmentation fidelity for offline analysis. YOLO-v7's Efficient Layer Aggregation Networks (E-ELAN) enable efficient real-time inference, while Mask R-CNN's Region Proposal Network (RPN) enhances detailed damage localization. These findings suggest a dual-use framework: YOLO-v7 for proactive, on-site monitoring, and Mask R-CNN for post-event forensic evaluation. This study advances the integration of AI in structural health monitoring and paves the way for future research on hybrid architectures, broader damage typologies, and extension to other infrastructure domains, such as steel bridges and composite structures.

**Keywords*:*** Concrete Structures, Structural Health Monitoring, Damage Detection, Deep Learning, YOLO-v7, Mask R-CNN, Data-Driven

## Introduction

The integrity and reliability of infrastructure systems such as buildings, bridges, and tunnels are fundamental to societal functionality, supporting commerce, transportation, and daily life. Concrete structures, despite their durability, are susceptible to damage caused by environmental exposure, aging, and heavy usage. Damage in these structures, particularly cracks and spalls, can compromise safety and lead to costly repairs or catastrophic failures if not addressed promptly. Consequently, regular inspections are critical to identifying early signs of deterioration and preventing significant

* Corresponding author. e-mail: sataei@stevens.edu.

structural issues. However, traditional methods of inspection rely heavily on manual evaluations, which are labor-intensive, costly, time-consuming, and prone to human error [1, 2]. The evaluation of infrastructures subjected to over-height impacts is typically informed by visual inspection; however, some damage mechanisms having significant effects on the strength and durability of prestressed concrete structures may not be visually apparent and therefore can be difficult to assess [3]. This highlights the need for more effective assessment methods. Efficient data management [4] and advanced vision-based methodologies [5, 6] —such as automated image analysis, sensor integration, and data-driven assessment— can be used to extend automated detection and evaluation frameworks for structures. Recent advancements in machine learning (ML) present transformative potential for infrastructure management, offering automated solutions that leverage large datasets to uncover complex patterns among explanatory variables [7]. These innovations underscore the necessity for automated solutions to overcome the limitations of traditional inspection approaches.

ML, a subset of Artificial Intelligence (AI), is revolutionizing structural health monitoring (SHM), which focuses on ensuring the safety and longevity of infrastructure. SHM refers to the continuous or periodic assessment of structural integrity using sensor networks, computer vision, and data-driven analytics. In bridges and concrete infrastructure, SHM involves collecting data from vibration sensors, strain gauges, accelerometers, and high-resolution imagery to monitor structural behavior under operational and environmental loads. The primary objectives are to detect, localize, and evaluate damage, predict potential failures, and optimize maintenance strategies [8]. SHM methods are primarily divided into vibration-based and vision-based approaches. Vibration-based methods involve sensors that measure dynamic properties like natural frequencies and strain to detect structural anomalies. While effective, these methods can be complex and less scalable due to the need for sophisticated sensor networks and extensive data processing. In contrast, vision-based SHM uses imaging technology to identify visible damage, such as cracks, in a non-invasive and scalable manner, allowing for extensive data collection over large areas. Recent advancements in computational power and imaging technology, along with the integration of machine learning and deep learning, have enhanced the effectiveness of vision-based SHM, making it a key area of research.

The development of vision-based methods has transitioned through several stages. Early approaches relied on heuristic techniques such as edge detection using Sobel, Prewitt, and Canny filters. While these methods were computationally efficient, they struggled with complex and nuanced damage patterns. The emergence of machine learning marked a paradigm shift, enabling models to learn features directly from data, thereby improving the accuracy and robustness of damage detection. Deep learning models, particularly CNNs, have revolutionized image analysis. CNNs are adept at learning hierarchical features from raw images, making them particularly effective for tasks such as crack detection, spall identification, and damage classification. By training on large datasets, CNN-based methods achieve significant improvements over traditional approaches.

Recent advances in instance segmentation and object detection have introduced state-of-the-art architectures such as You Only Look Once (YOLO) [9] and Mask R-CNN [10], which have been successfully applied to structural damage detection. YOLO models, particularly YOLOv7, are renowned for their real-time performance, achieving both high detection accuracy and fast inference speeds, making them suitable for field-deployable SHM systems. Mask R-CNN, on the other hand, provides pixel-level segmentation, enabling precise delineation of cracks and spalls, which is critical

for evaluating damage severity. Recent research has demonstrated the effectiveness of data-driven and deep learning methods for detecting and evaluating damage in concrete and wind turbine structures, highlighting YOLOv7's superior performance and real-time capability compared to other models such as Mask R-CNN [47, 48]. The application of AI in damage detection facilitates a shift from manual inspections to automated systems that are faster, more accurate, and scalable. By utilizing data-driven methods, AI analyzes large datasets to identify patterns and predict potential failures in real time. This study evaluates the effectiveness of advanced deep learning models, specifically YOLO-v7 instance segmentation and Mask R-CNN, for detecting and analyzing damage in concrete structures. Manual inspection methods, though common, have significant drawbacks, including subjectivity and high costs, especially for large-scale projects. They also struggle with hard-to-reach areas and real- time monitoring. AI-powered systems address these challenges through advanced image recognition, allowing for rapid damage detection and reduced maintenance costs.

This research emphasizes vision-based deep learning methods, which have shown exceptional performance in automating damage detection using images and videos. YOLO-v7 focuses on real-time object detection, while Mask R-CNN excels in precise damage localization. By comparing these models, the study aims to highlight their strengths and limitations for application in SHM. The implications of these models extend to real-time monitoring and proactive maintenance, enhancing infrastructure management. The broader goal is to develop robust, scalable, and cost-effective solutions for SHM, bridging traditional methods with AI-driven approaches. This research underscores the transformative potential of AI and deep learning in SHM, improving accuracy, reducing costs, and enhancing monitoring capabilities for safer infrastructure systems.
Key contributions of this research include:

- By combining and augmenting datasets from various sources and leveraging transfer learning, the study optimizes the performance of both models, setting a benchmark for future automated SHM studies.

- Successfully optimized both models using transfer learning and advanced training techniques and fine tuned on our dataset.

- The research evaluates YOLO-v7 against Mask R-CNN, offering a comprehensive analysis of their strengths and limitations in detecting and segmenting structural damage.

The literature highlights several advancements in the application of machine learning and deep learning to structural damage detection.

ML and deep learning (DL) have rapidly reshaped how civil-infrastructure damage is detected, classified, and quantified. Vision-based approaches now outperform traditional manual inspection in both accuracy and speed, while sequence models extend predictive capabilities to dynamic hazards such as earthquakes. This literature review synthesizes the main advances from 2018 to mid-2025, highlighting methodological progress, performance trends, and persistent research gaps. Figure 1 shows different types of existing detection methods.

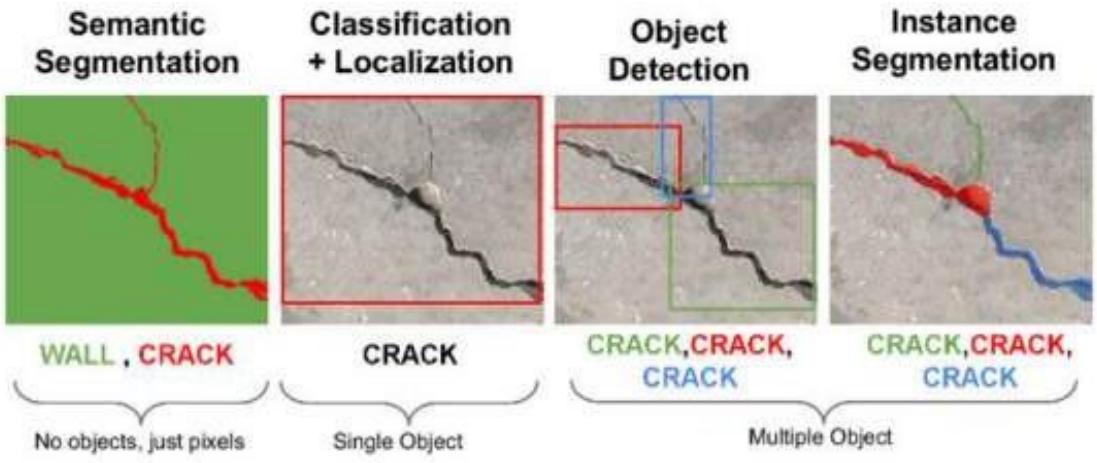


*Figure 1 Different Types of Existing Detection Methods [44]*

Silva et al. first demonstrated the efficacy of transfer learning by training VGG16 on a two-class crack/no-crack dataset, achieving 92.27% accuracy [11]. Later hybrid preprocessing—combining Otsu thresholding, Difference-of-Gaussian filtering, and homomorphic enhancement—boosted accuracy while cutting false positives [12, 13]. Although deeper backbones like DenseNet121 and ResNet50V2 offer improved robustness, lightweight networks such as a shallow 2D-CNN reach 99.53% accuracy with one-third the inference time [14]. Current limitations arise more from data heterogeneity than model depth, making noise-robust preprocessing and augmentation critical.

Edge-oriented studies now prioritize parameter reduction. Integrating Ghost Convolution and GSConv into YOLOv8 cut weights by 17.3% and raised mAP50 by 4.4 points [15]. Mobile-friendly networks like EfficientNetB0 and MobileNet achieve sub-80 ms/frame inference with ≤2 million parameters [16]. Fusion strategies also improve interpretability: a ResNet-50 + Curvelet model enhanced F1 to 96.1% and produced saliency maps for regulatory transparency [17].

For seismic damage, stacked LSTMs predict post-earthquake states with 80–95% accuracy [18], while a Tsinghua pipeline delivers real-time district-level damage maps within 10 s (F1 = 0.88) [19]. Hybrid controllers embedding LSTM reduce interstory drift by 46% [20]. Vision models also advance: Faster R-CNN detects column spalling at 80% accuracy [21], R-FCN improves precision at ∼3× computation [21, 22], and cascading detectors (C-Mask R-CNN) push pavement mAP to 95.4% [23]. UAV-mounted R-CNNs shorten inspection time by 40% [21], while AlexNet → YOLO pipelines segment cracks in 85.71% of aerial frames [24].

The YOLO family evolved through major upgrades (v3→v12), with v7's E-ELAN improving feature aggregation [25]. Although YOLOv8 suits lightweight tasks, v7
excels on resource-rich servers. Enhanced variants like YOLOv8-VOS and YOLOv8-PCD refine small-target detection with custom layers and Biformer attention [26, 27]. Segmentation networks such as Mask R-CNN with ResNeXt achieve 96.5% IoU [21], DeepLabV3+ enhances multi-scale context (94.2% IoU) [28], and APLCNet adds semantic branches for 92.21% precision [29]. Transformer-based models, including CrackFormer [30], CrackFormer-II [31], and hybrid CNN-Transformer designs like PCTC-Net and CCTNet, surpass 96% detection accuracy on large concrete datasets [32, 33].

Pure transformer models are now viable despite small datasets. SwinCrack achieves 0.861 ODS with hierarchical windows [34], while Locally Enhanced Transformers improve fine-scale connectivity [35, 36]. FTN-ResNet50 integrates ViT tokens, raising F1 by 5% on CrackTree200 [37]. Adaptive attention mechanisms (LSKA, Biformer, BiFPN-L) dominate recent work, crucial for detecting both hairline cracks (0.1 mm) and large spalls (10 cm). LSTMs remain vital for structural monitoring, cutting pier displacement errors from 62.88% to 19.44% [38], and regional frameworks achieve ¡3 s latency [19]. Although transformers show promise for long-term earthquake forecasting, adoption remains limited due to sparse seismic datasets [39, 40].

From the first VGG16 transfer-learning experiments in 2018 to transformer-driven dual-encoder architectures in 2025, machine-learning-based structural damage detection has attained near-expert accuracy, real-time inference on embedded devices, and expanding predictive reach to dynamic hazards. While CNNs remain indispensable for local feature extraction, transformers offer unmatched global context, and hybrid models increasingly marry the two. Persistent dataset, explainability, and edge-deployment challenges signal fertile ground for the next wave of research aimed at truly autonomous, trustworthy structural health monitoring.

Comparative studies highlight the speed–accuracy trade-off in vision-based SHM. Mask R-CNN provides highly detailed segmentation masks but suffers from slower inference, making it less suitable for real-time applications. In contrast, YOLO models, particularly YOLO-v7 instance segmentation, achieve a practical balance between fast detection and high accuracy.

Vision-based SHM offers significant benefits, including rapid data acquisition, non-destructive testing, large-scale monitoring, and reduced reliance on subjective human assessment. However, its effectiveness depends on large annotated datasets and is sensitive to lighting, occlusion, and resolution variations. Addressing these challenges requires robust data augmentation and models capable of generalizing across diverse conditions.

Overall, AI-driven SHM demonstrates transformative potential. Advanced deep learning methods provide high accuracy and scalability, but selecting the appropriate model involves balancing precision and computational efficiency for specific applications. This study evaluates YOLO-v7 and Mask R-CNN, emphasizing their applicability to real-world infrastructure monitoring.

## Methodology

The methodology outlines the systematic approach taken to investigate and compare the efficacy of YOLO-v7 instance segmentation and Mask R-CNN for detecting damages in concrete structures. It comprises three key components: dataset preparation, algorithm implementation, and the training and evaluation process.

### Dataset Preparation

The dataset serves as the backbone for training and evaluating deep learning models. This study utilized a dataset of 400 images of concrete cracks and spalls, selected from three publicly available datasets [41–43] and labeled for training. To enhance the generalizability and robustness of the

models, data augmentation techniques were employed, increasing the dataset size to 10,995 images. The dataset with masks is available at https://www. kaggle.com/datasets/stmlen/cconcrack. Figure 2 shows examples of images from dataset and Figure 3 demonstrates some labeled data samples. This image shows various labeled data samples used for training the model, highlighting different types of damage including cracks and spalls.

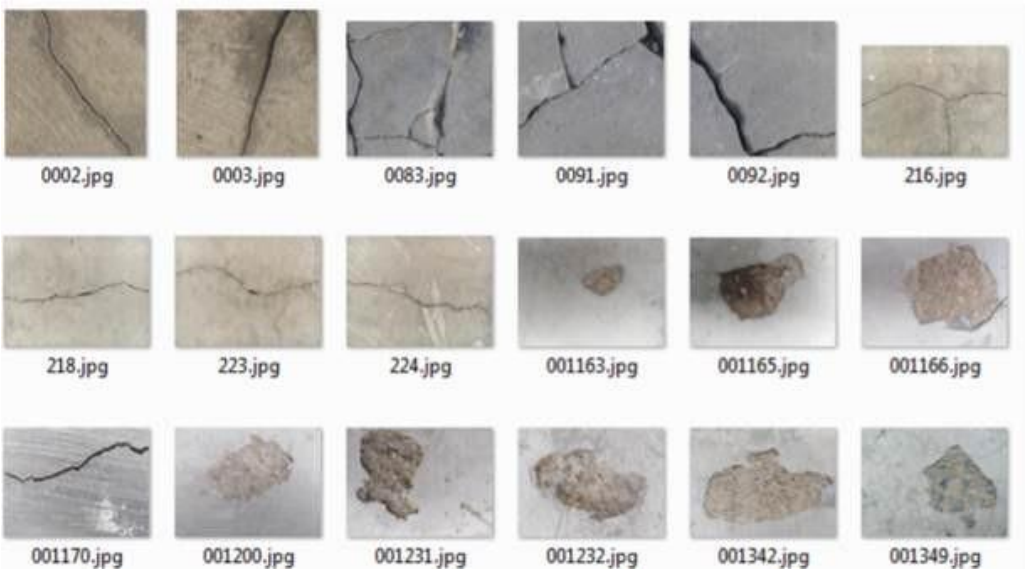


*Figure 2 Examples of Images from the Dataset*

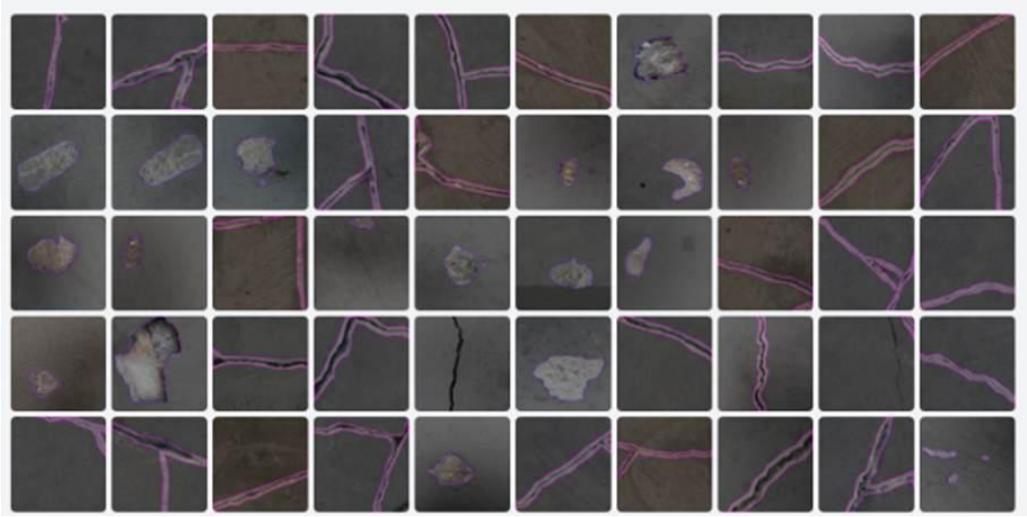

*Figure 3 Labeled Data Samples*

Data augmentation plays a pivotal role in enhancing the model's ability to generalize across varying real-world conditions. In this study, several augmentation strategies were employed. Geometric transformations such as rotation, flipping (both horizontal and vertical), and cropping were used to simulate different perspectives of structural damage. Color adjustments—including variations in brightness, contrast, and hue—were applied to account for changes in lighting conditions. Zoom and scaling operations were incorporated to mimic both close-up and distant views of defects. Additionally, displacement along the X and Y axes was introduced to improve the model's

robustness by simulating shifts in object positioning. Collectively, these techniques produced a diverse and representative dataset, significantly improving the model's performance across a wide range of environmental and operational scenarios. Precise annotations are critical for the success of supervised learning models.

Images were labeled at the instance level using the Roboflow platform, enabling the identification of individual damage types, such as cracks and spalls. Each annotation included bounding boxes and segmentation masks, ensuring compatibility with both YOLO-v7 instance segmentation and Mask R-CNN architectures.

## Algorithms and Implementation

Two state-of-the-art deep learning models, YOLO-v7 instance segmentation and Mask R-CNN, were implemented and fine-tuned on the proposed dataset of concrete crack to evaluate their performance in structural damage detection.

### *Mask R-CNN*

Mask R-CNN is a two-stage instance segmentation model known for its ability to produce high-resolution segmentation masks. Its architecture comprises several key components. The backbone network, typically a ResNet or ResNeXt, is responsible for extracting hierarchical features from the input image. These features are then passed to the Region Proposal Network (RPN), which identifies regions of interest (ROIs) that are likely to contain objects. To ensure accurate segmentation, ROI Align is used to preserve spatial alignment when processing the ROIs. Finally, the mask head generates pixel-level segmentation masks for each detected object. This two-stage approach enables Mask R-CNN to perform precise damage localization, making it particularly effective in scenarios where high segmentation accuracy is critical. Figure 4 shows Mask R-CNN model architecture.

### *YOLO-v7*

YOLO-v7 instance segmentation represents the latest evolution in the YOLO family, renowned for its real-time object detection capabilities. Unlike Mask R-CNN, YOLO-v7 operates as a single-stage model, combining object detection and instance segmentation within a unified architecture. Among its key features, the Efficient Layer Aggregation Network (E-ELAN) enhances feature extraction by optimizing the flow of information across network layers. Re-parameterized convolution layers further contribute to computational efficiency during both training and inference. The model's single-stage detection mechanism removes the need for a separate region proposal stage, resulting in significantly faster processing. Thanks to this streamlined design, YOLO-v7 is capable of handling high frame-rate image processing, making it particularly well-suited for real-time damage detection applications. Figure 5 shows YOLO model architecture.

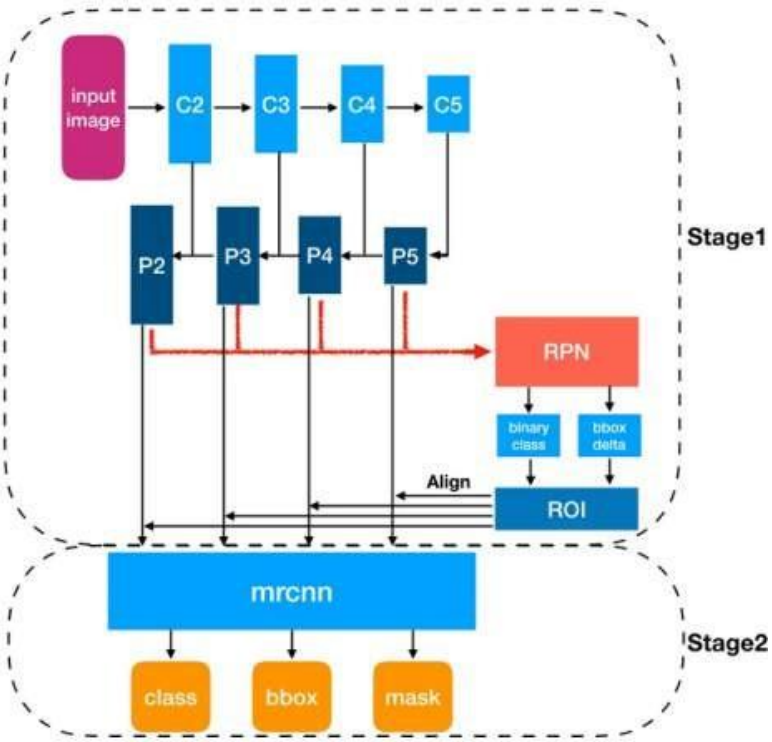


*Figure 4 Mask R-CNN Model Architecture [45]*

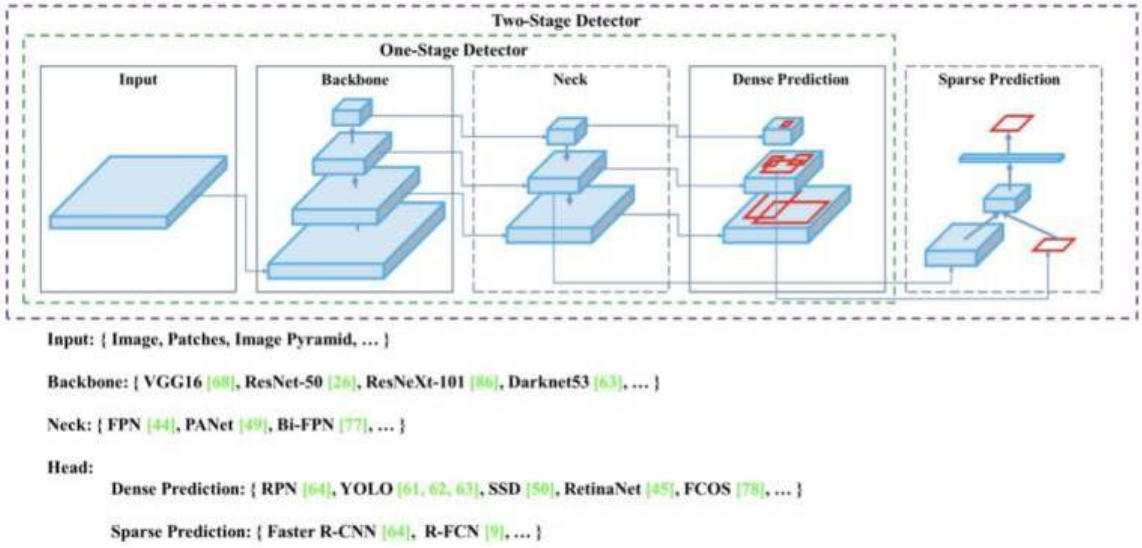


*Figure 5 YOLO Model Architecture [46]*

## Training and Evaluation

PyTorch framework used for training, utilizing GPU acceleration to ensure computational efficiency. The training process included extensive hyperparameter tuning to optimize the balance between model complexity and performance. Key hyperparameters adjusted during this process included the learning rate, which controlled the speed of convergence during training, and the batch size, which influenced training stability by determining how many images were processed simultaneously. The Stochastic Gradient Descent (SGD) algorithm was employed as the optimizer to minimize the loss function effectively. Additionally, early stopping was implemented to prevent

overfitting by terminating the training process when validation performance no longer improved. This tuning strategy contributed significantly to achieving robust and generalizable model performance.

**Evaluation Metrics**

Performance evaluation of the models was conducted using four key metrics. Precision measured the proportion of true positive predictions out of all predicted positives, reflecting the model's accuracy in identifying relevant instances. Recall quantified the proportion of true positives among all actual positives, indicating the model's ability to detect all relevant instances. Mean Average Precision at a 50% Intersection over Union (mAP50) served as a comprehensive indicator of detection accuracy, evaluating how well the predicted bounding boxes matched the ground truth. Finally, Frames Per Second was used to assess the models' inference speed, providing insight into their suitability for real-time applications.

1. *Precision:* Measures the proportion of correctly identified damage instances out of all predicted instances (Equation (1-3)).

$$P = \frac{TP}{TP + FP} \tag{1-3}$$

2. *Recall:* Evaluates the proportion of actual damage instances correctly detected by the model (Equation (2-3)).

$$R = \frac{TP}{TP + FN} \tag{2-3}$$

3. *Mean Average Precision (mAP50):* Provides a single metric summarizing detection accuracy at a 50% Intersection over Union (IoU) threshold (Equation (3-3)).

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{3-3}$$

4. *Frames Per Second (FPS):* Quantifies the real-time applicability of the models by measuring the number of images processed per second.

These metrics collectively highlight the models' detection accuracy, coverage, and real-time efficiency, offering a nuanced comparison. Precision emphasizes a model's reliability in avoiding false positives, while recall highlights its ability to detect all instances of damage, including subtle ones. The mAP50 metric provides a balanced evaluation of precision and recall across varying IoU thresholds, and FPS assesses the practicality of each model for real-time applications.

Both YOLO-v7 instance segmentation and Mask R-CNN were fine-tuned on the proposed dataset of concrete damages. Also transfer learning with pre-trained weights from the COCO dataset were adapted, accelerating convergence and improving accuracy. Training involved extensive data augmentation to simulate real-world conditions. Each batch of images included variations in lighting, orientation, and scale, ensuring the models could generalize to unseen scenarios. Training and evaluation were conducted in high-performance computing environments, including GPUs

available on Kaggle and Google Colab. These platforms facilitated efficient model training and hyperparameter optimization.

Mask R-CNN, due to its two-stage architecture, demands substantial computational resources for both training and inference. While the model is resource-intensive, this investment results in high-quality, pixel-level segmentation masks that are particularly valuable for detailed damage analysis. Its precise localization capabilities make it well-suited for post-event evaluations or applications where segmentation accuracy is prioritized over speed.

In contrast, YOLO-v7 benefits from a streamlined, single-stage architecture that significantly reduces training and inference time. This efficiency enables the model to perform real-time detection, making it ideal for field applications requiring rapid assessment. Architectural enhancements such as Efficient Layer Aggregation and re-parameterized convolutions contribute to its robust ability to detect multiple instances of damage quickly and accurately.

Following training, models were evaluated on the test dataset to assess their performance comprehensively. This evaluation focused on comparing key metrics such as accuracy, precision, recall, and Frames Per Second. Validation procedures included testing on unseen data to examine the models' generalizability beyond the training set, as well as analyzing failure cases to identify common misclassification patterns and pinpoint areas needing improvement.

### Challenges

The study encountered several challenges that impacted model performance and generalizability. First, limited dataset diversity restricted the models' ability to generalize across different types of structural damage and environmental scenarios. Additionally, varying lighting conditions in the image data introduced noise, complicating the training process and potentially reducing detection accuracy. Finally, computational resource constraints limited the capacity to train models on larger, more representative datasets, thereby affecting the depth and breadth of model learning.

To address the challenges posed by limited data and computational resources, several strategies were implemented. Data augmentation was extensively used to simulate a wide range of real-world conditions, enhancing the models' ability to generalize. Transfer learning further contributed to efficient model development by leveraging pretrained weights, significantly accelerating convergence and reducing the dependence on large labeled datasets. The methodology outlined in this study combines robust dataset preparation, advanced model architectures and training processes. By leveraging the complementary strengths of YOLO-v7 instance segmentation and Mask R-CNN, this research provides a comprehensive evaluation of their applicability to structural damage detection.

## Results

This section evaluates the comparative performance of proposed fine-tuned YOLO-v7 instance segmentation and Mask R-CNN for detecting and analyzing damage in concrete structures, highlighting the strengths, weaknesses, and trade-offs of each model based on various metrics.The performance of the models on the test dataset revealed distinct strengths and limitations. Mask R-CNN achieved a high mAP50 score of 92.1%, demonstrating strong effectiveness in accurately

detecting and segmenting damage. However, its detection speed was limited to 18 Frames Per Second, which restricts its suitability for real-time applications. Its two-stage architecture provided detailed instance segmentation, making it ideal for tasks that require precise analysis of damage extent and severity. Nonetheless, the relatively slower inference time was a key weakness, rendering it less optimal for scenarios demanding rapid detection.

YOLO-v7 Semantic Segmentation outperformed Mask R-CNN with a mAP50 of 96.1%, thanks to advanced architectural features such as E-ELAN and re-parameterized convolution layers. It processed images at a rapid rate of 40 FPS, making it highly suitable for real-time applications. Its single-stage detection framework effectively balanced speed and accuracy, enabling efficient detection of multiple damage instances in dynamic environments. However, despite its high precision, YOLO-v7 occasionally sacrifices segmentation detail compared to the more meticulous output of Mask R-CNN. Figure 6 and Figure 7 present the YOLO-v7 training results with 200 iterations and comparison of the accuracy of the proposed fine-tuned YOLO-v7 and Mask R-CNN models, respectively. Table 1 shows YOLO-v7 instance segmentation model accuracy results for two classes. Table 2 demonstrates comparison of Instance Segmentation Models, Mask R-CNN and YOLO-v7 and Table 3 compares the speed of the investigated models.
The findings underscore the inherent trade-offs between speed and precision in the two models. YOLO-v7's streamlined architecture and real-time processing capabilities make it highly suitable for field applications such as autonomous inspections and continuous monitoring systems, where its ability to handle multiple images per second ensures rapid response times crucial for infrastructure maintenance. In contrast, Mask R-CNN's two-stage process provides high-resolution segmentation masks, offering detailed analysis of damage extent and severity, which is particularly valuable in scenarios like post-event damage assessment where detection speed is less critical than precision. YOLO-v7's performance benefits greatly from architectural innovations such as Efficient Layer Aggregation Networks (E-ELAN) and re-parameterized convolution layers, which enhance feature extraction and inference efficiency, leading to superior accuracy and speed. Meanwhile, Mask R-CNN's versatility stems from its integration of Region Proposal Networks (RPN) and ROI Align mechanisms, which ensure precise spatial alignment of proposals and enable accurate segmentation of complex damage patterns.

Figure 8 shows the accuracy of the proposed Fine-tuned YOLO-v7 model tested with a random photo taken from the Internet and Figure 9 demonstrates the accuracy of the model tested with a random video taken from the Internet.

## Limitations and Recommendations

While this study provides valuable insights, it also reveals areas for further exploration and improvement. Key recommendations for future research include:

### *Dataset Expansion and Diversity*

The dataset used in this study, although augmented to improve robustness, primarily focused on specific types of concrete damage (cracks and spalls). Expanding the dataset to include a wider variety of damage types and structural conditions, such as corrosion, scaling, and delamination, would enhance model generalizability. Incorporating diverse environmental scenarios (e.g., varying lighting conditions, weather impacts, and occlusions) would improve real-world performance.

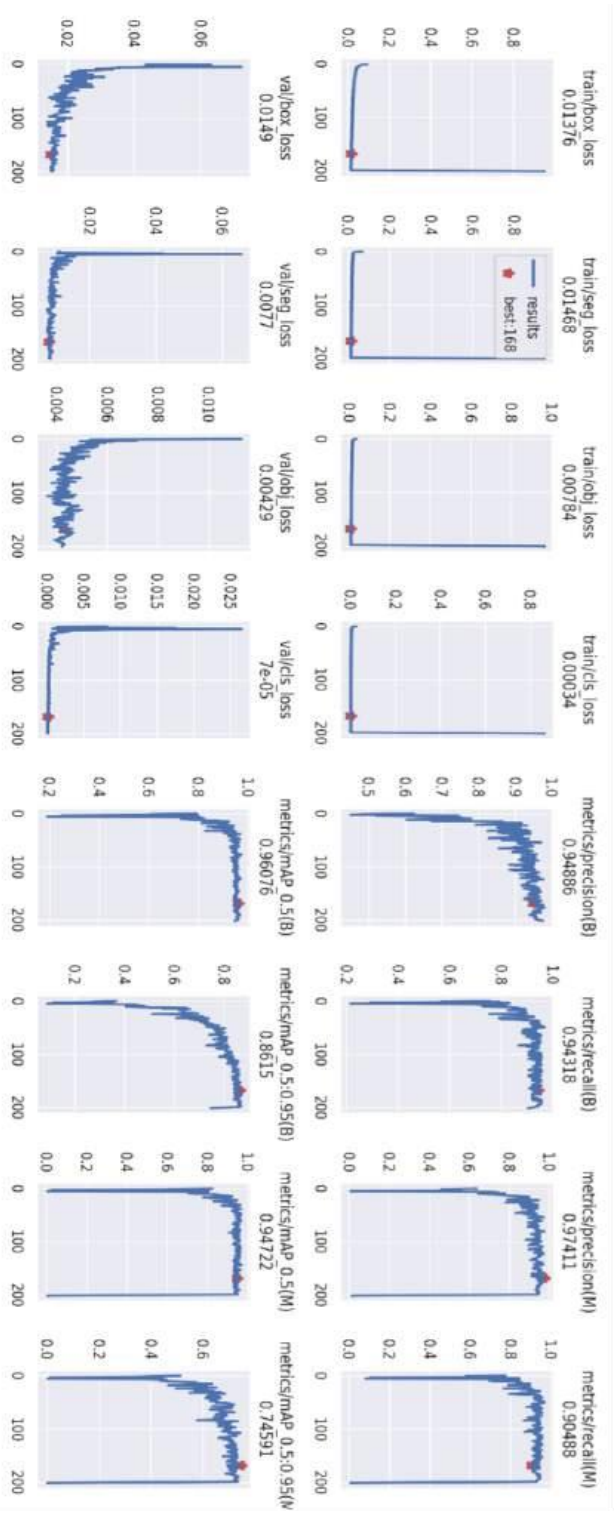


*Figure 6 YOLO-v7 Model training results with 200 iterations*

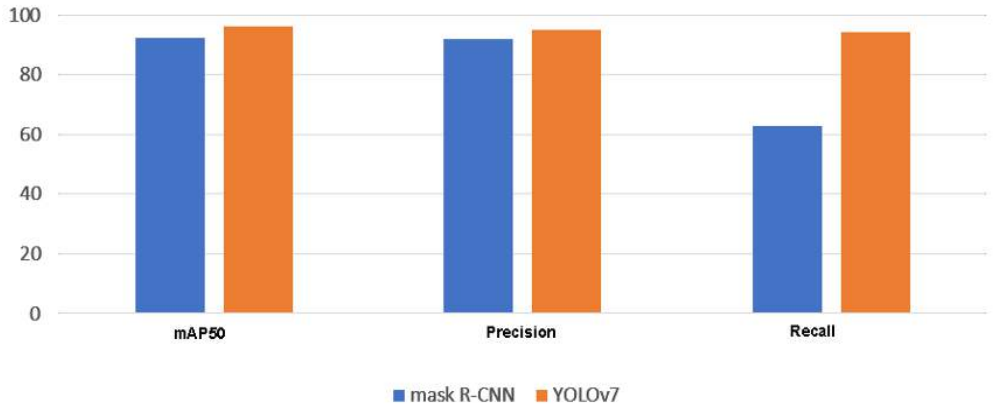


*Figure 7 Comparison of the accuracy of the two models*

Table 1 YOLO-v7 instance segmentation model results for two classes

| Class | Precision | Recall | mAP50 |
|---|---|---|---|
| **Crack** | 94.5 | 88.6 | **92.7** |
| **Spall** | **95.2** | **99.9** | **99.5** |

Table 2 Comparison of Two Instance Segmentation Models

| Model | mAP50 | Precision | Recall |
|---|---|---|---|
| **Mask R-CNN** | 92.1 | 92.0 | **62.8** |
| **YOLO-v7** | **96.1** | **94.9** | **94.3** |

Table 3 Comparison of the Speed of Two Instance Segmentation Models

| Model | Time (msec) | Speed (fps) |
|---|---|---|
| **Mask R-CNN** | 55 | **18** |
| **YOLO-v7** | **24.9** | **40** |

*Hybrid Model Development*

Combining the strengths of YOLO-v7's real-time capabilities with Mask R-CNN's detailed segmentation accuracy could result in a hybrid model optimized for both speed and precision. Future research could explore innovative architectures that integrate these complementary strengths.

*Application to Other Infrastructure Types*

While this study focuses on concrete structures, extending the models to other types of infrastructure, such as steel bridges, pavements, and wooden structures, would broaden their applicability. Customizing models for these contexts would require additional training datasets and specialized adaptations.

*Real-Time Implementation and Deployment*

Developing lightweight versions of these models for deployment on portable devices and drones could facilitate on-site inspections. This would require optimization techniques to reduce computational demands without sacrificing performance.

*Interdisciplinary Approaches*

Collaborating with experts in materials science, civil engineering, and computer science could lead to more holistic SHM systems. For example, integrating sensor data with visual inspections could provide a more comprehensive understanding of structural health.

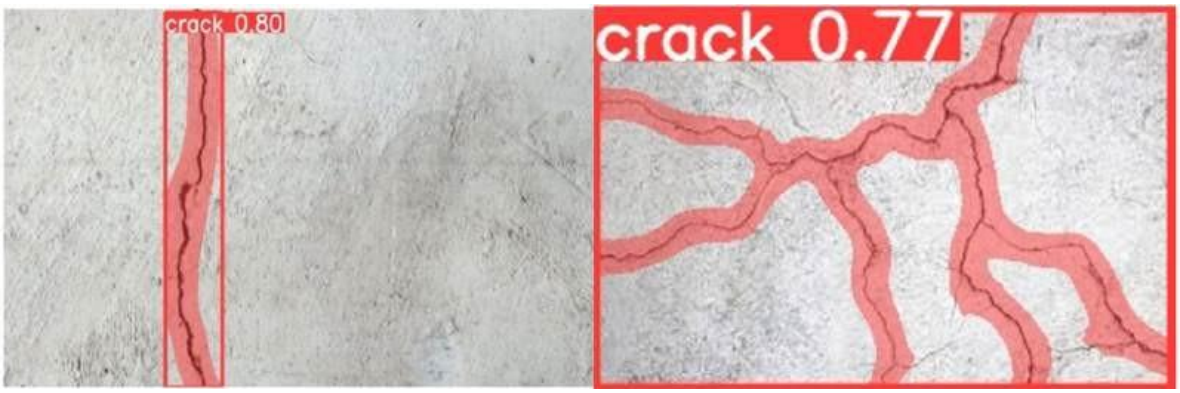


*Figure 8 Testing the accuracy of the proposed fine-tuned YOLO-v7 model with two random photos taken from the Internet*

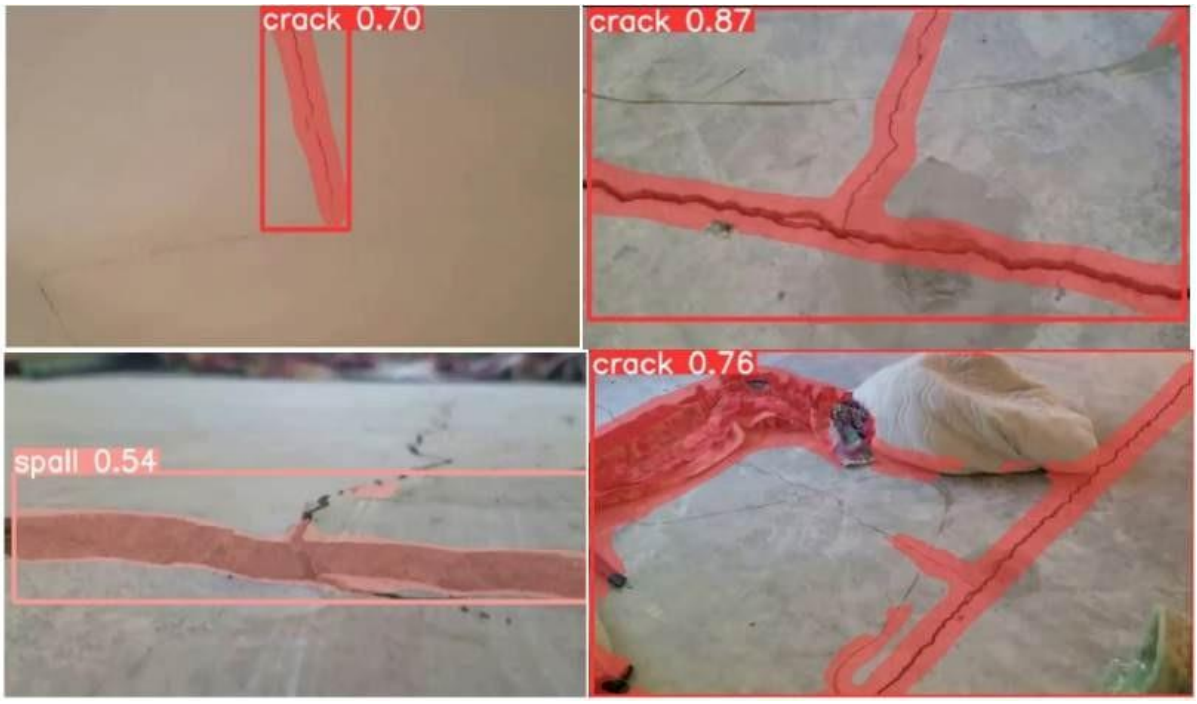


*Figure 9 Testing the accuracy of the proposed fine-tuned YOLO-v7 model using a random video taken from the Internet*

## Conclusion

This study demonstrates the significant potential of deep learning in transforming structural health monitoring through automated damage detection in concrete infrastructure. Although convolutional neural networks (CNNs) have demonstrated considerable success in damage detection through image classification and object recognition, the application of instance segmentation in this domain has been comparatively limited. To address this gap, three distinct datasets were integrated and annotated at the instance level to train and evaluate two segmentation models—Mask R-CNN and YOLOv7—based on their accuracy and computational efficiency. By systematically evaluating

proposed fine-tuned YOLO-v7 instance segmentation and Mask R-CNN, this paper reveals the distinct strengths of each model in addressing different operational demands. YOLO-v7, with its impressive real-time performance (40 FPS) and high detection accuracy (mAP50 of 96.1%), is well-suited for continuous, in-field monitoring where rapid decision-making is essential. In contrast, Mask R-CNN, while operating at a slower frame rate (18 FPS), offers high-precision segmentation (mAP50 of 92.1%), making it ideal for offline analysis and detailed post-event damage assessment. These findings highlight the importance of aligning model selection with the specific performance requirements of SHM applications. By leveraging the complementary advantages of these models, infrastructure monitoring systems can be tailored to balance speed, accuracy, and contextual needs.

This research underscores the broader value of AI-driven approaches in replacing or enhancing traditional manual inspections, offering scalable, objective, and efficient alternatives for infrastructure management. The outcomes pave the way for future exploration into hybrid models, multi-class damage classification, and applications beyond concrete—such as steel or composite structures. Ultimately, this work lays a foundation for developing smarter, safer, and more responsive SHM solutions that integrate advanced computer vision technologies into the core of civil infrastructure maintenance.

## Acknowledgments

The authors acknowledge the contributions of the referenced datasets and the Sharif University of Technology's Department of Aerospace Engineering for supporting this research.

## Data Availability Statement

The dataset with masks is available at https://www.kaggle.com/datasets/stmlen/cconcrack.

## Conflict of interest

The authors declared no conflict of interest.

## References

[1] Gatti, M.: Structural health monitoring of an operational bridge: A case study. Engineering Structures 195, 200–209 (2019)

[2] Kim, H., Ahn, E., Shin, M., Sim, S.-H.: Crack and noncrack classification from concrete surface images using machine learning. Structural Health Monitoring 18, 725–738 (2019)

[3] Samadzad, A., Cathey, S., Whelan, M., Braxtan, N., Chen, S.: Finite element analysis of over-height vehicle collisions on prestressed girder bridges. In: Bridge Maintenance, Safety, Management, Digitalization and Sustainability, pp. 1801– 1808. CRC Press, ??? (2024)

[4] Soltanmohammadi, E., Hikmet, N.: Optimizing healthcare big data processing with containerized pyspark and parallel computing: A study on etl pipeline efficiency. Journal of Data Analysis and Information Processing 12, 544–565 (2024) https://doi.org/10.4236/jdaip.2024.124029

[5] Safari, S., DuBose, T., Head, M.H., Shenton, H.W., Tatar, J., Chajes, M.J., Hastings, J.N.: Diagnostic load testing and assessment of a corroded corrugated metal pipe culvert before rehabilitation. Structure and Infrastructure Engineering 20(7–8), 1149–1158 (2023)

[6] Safari, S., Head, M., DuBose, T.: Vision-based measurements for monitoring a rehabilitated corrugated metal pipe culvert. In: Proceedings of the ASCE International Conference, pp. 416–434 (2023). https://doi.org/10.1061/9780784484777.037

[7] Mohammadagha, M., Najafi, M., Kaushal, V., Jibreen, A.M.A.: Machine Learning Models for Reinforced Concrete Pipes Condition Prediction: The State-of-the-Art Using Artificial Neural Networks and Multiple Linear Regression in a Wisconsin Case Study. https://arxiv.org/abs/2502.00363. arXiv preprint arXiv:2502.00363 (2025)

[8] Farrar, C.R., Worden, K.: Structural Health Monitoring: a Machine Learning Perspective. John Wiley & Sons, ??? (2012)

[9] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)

[10] He, K., Gkioxari, G., Doll´ar, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)

[11] Silva, W.R.L.d., Lucena, D.S.d.: Concrete cracks detection based on deep learning image classification. In: Proceedings, vol. 2, p. 489 (2018)

[12] Zhu, X., Law, M.T., Sohn, H.: Compressive-sensing data reconstruction for structural health monitoring: a machine-learning approach. Smart Materials and Structures 29(7), 075003 (2020) https://doi.org/10.1088/1361-665X/ab88db

[13] Khajehabdollahi, A., Kim, H., Hoskere, V., Narazaki, Y., Spencer Jr, B.F.: A data-based structural health monitoring approach for damage detection in steel bridges using experimental data. Engineering Structures 234, 111973 (2021) https://doi.org/10.1016/j.engstruct.2021.111973

[14] Sun, H., Zhou, H., Wang, Y., Omenzetter, P.: A data-driven approach for drive by damage detection in bridges considering the influence of temperature change. Mechanical Systems and Signal Processing 166, 108319 (2022) https://doi.org/ 10.1016/j.ymssp.2021.108319

[15] He, Y., Cao, M., Yu, Y., Chen, Z., Zhang, L.: Damage detection in railway bridges using traffic-induced dynamic responses. Structural Control and Health Monitoring 28(6), 2718 (2021) https://doi.org/10.1002/stc.2718

[16] Feng, Y., Narazaki, Y., Spencer Jr, B.F.: Synthetic data augmentation for pixel-wise steel fatigue crack identification using fully convolutional networks. Automation in Construction 140, 104282 (2022) https://doi.org/10.1016/j.autcon.2022.104282

[17] Gao, Y., Mosalam, K.M.: Deep transfer learning for image-based structural damage recognition. Computer-Aided Civil and Infrastructure Engineering 33(9), 748–768 (2018) https://doi.org/10.1111/mice.12336

[18] Li, H., Zhao, X., Li, J., Xie, X., Guo, T.: Automated pixel-level crack detection and measurement using deep convolutional neural networks. Computer-Aided Civil and Infrastructure Engineering 36(5), 464–480 (2021) https://doi.org/10. 1111/mice.12639

[19] Feng, Y., Narazaki, Y., Spencer Jr, B.F.: Model-assisted crack detection with uncertainty quantification using fully convolutional networks. Computer-Aided Civil and Infrastructure Engineering 36(10), 1236–1254 (2021) https://doi.org/ 10.1111/mice.12695

[20] Gong, C., Ma, J., Li, J., Zhang, S.: Steel crack detection based on deep learning and image processing. Structural Health Monitoring 22(3), 1132–1147 (2023) https://doi.org/10.1177/14759217221120498

[21] Kim, H., Spencer Jr, B.F.: Automated vision-based bridge inspection using deep learning with robotic platforms. Journal of Structural Engineering 147(2), 04020345 (2021) https://doi.org/10.1061/(ASCE)ST.1943-541X.0002886

[22] Min, K., Lee, J., Cho, S.: Deep learning-based real-time autonomous crack evaluation system for infrastructure inspection. Automation in Construction 119, 103331 (2020) https://doi.org/10.1016/j.autcon.2020.103331

[23] Bang, S., Kim, H., Kim, H., Cho, S.: Encoder–decoder network for pixel-wise crack detection using vgg16 and resnet34. Computer-Aided Civil and Infrastructure Engineering 36(6), 607–623 (2021) https://doi.org/10.1111/mice.12654

[24] Zou, Q., Wang, S., Wang, Y., Qi, X.: Deepcrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing 338, 139–153 (2019) https://doi.org/10.1016/j.neucom.2019.01.035

[25] Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. https://arxiv.org/abs/2004.10934. arXiv:2004.10934 (2020)

[26] Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., Fieguth, P.: A review on computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure. Advanced Engineering Informatics 29(2), 196–210 (2015) https://doi.org/10.1016/j.aei.2015.01.008

[27] Jiang, H., Wu, Z., Li, Q., Wang, Y.: Vision-based detection of bridge surface cracks using a hybrid deep learning model. Engineering Structures 226, 111397 (2021) https://doi.org/10.1016/j.engstruct.2020.111397

[28] Wang, D., Lin, Y., Cao, M., Zhang, Z.: Bridge crack detection using attention-based deep learning with synthetic images. IEEE Transactions on Intelligent Transportation Systems 23(11), 19951–19961 (2022) https://doi.org/10.1109/ TITS.2022.3141324

[29] Li, S., Li, H., Wang, S., Deng, Y.: Improved u-net model for concrete crack detection using multi-scale feature fusion and attention mechanism. Automation in Construction 132, 103941 (2021) https://doi.org/10.1016/j.autcon.2021.103941

[30] Zhang, J., Yang, J., Ma, Y., Zhang, Y.: Concrete crack detection based on faster r-cnn with fpn and iou-balanced loss. Sensors 20(15), 4212 (2020) https://doi. org/10.3390/s20154212

[31] Fan, Y., Wang, X., Li, Y., Guo, Y.: Crack detection based on deep convolutional neural network using digital image processing. Measurement 177, 109317 (2021) https://doi.org/10.1016/j.measurement.2021.109317

[32] Kim, Y., Son, H., Kim, C.: Automated vision-based concrete crack detection using a deep learning algorithm. Sensors 20(17), 5104 (2020) https://doi.org/10.3390/ s20175104

[33] Zhang, L., Zhang, L., Lu, W., Zhang, Y.: Crack detection in concrete using convolutional neural networks and data augmentation. IEEE Access 8, 144560–144570 (2020) https://doi.org/10.1109/ACCESS.2020.3014066

[34] Ellenberg, A., Kontsos, A., Moon, F., Bartoli, I.: Bridge deck delamination detection with ground-penetrating radar and infrared thermography using data fusion and machine learning. NDT E International 91, 45–56 (2017) https://doi.org/10.1016/j.ndteint.2017.06.002

[35] Dorafshan, S., Thomas, R.J., Maguire, M.: Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Construction and Building Materials 186, 1031–1045 (2018) https://doi.org/10. 1016/j.conbuildmat.2018.07.111

[36] Yeum, C.M., Dyke, S.J.: Vision-based automated crack detection for bridge inspection. Computer-Aided Civil and Infrastructure Engineering 30(10), 759– 770 (2015) https://doi.org/10.1111/mice.12120

[37] Prasanna, P., Dana, K., Gucunski, N., Basily, B.B., La, H.M., Lim, R.S., Parvardeh, H.: Automated crack detection on concrete bridges. IEEE Transactions on Automation Science and Engineering 13(2), 591–599 (2016) https://doi.org/ 10.1109/TASE.2015.2500179

[38] Yokoyama, K., Tohyama, T., Ohno, K.: Concrete crack detection using machine learning with digital image processing. Journal of Advanced Concrete Technology 15(10), 476–486 (2017) https://doi.org/10.3151/jact.15.476

[39] Atha, D.J., Jahanshahi, M.R.: Evaluation of deep learning approaches based on convolutional neural networks for crack detection. Computers in Civil Engineering 32(3), 04016054 (2017) https://doi.org/10.1061/(ASCE)CP.1943-5487.0000650

[40] Xu, Y., Xu, L., Zhao, X., Zhu, H., Guo, Y.: Automatic crack detection and segmentation using u-net based fully convolutional networks. Computer-Aided Civil and Infrastructure Engineering 36(4), 415–430 (2021) https://doi.org/10.1111/ mice.12585

[41] Fırat O¨ zgenel: Concrete crack segmentation dataset. Mendeley Data 1 (2019)

[42] Zhang, C., Chang, C.C., Jamshidi, M.: Simultaneous pixel-level concrete defect detection and grouping using a fully convolutional model. Structural Health Monitoring 20(4), 2199–2215 (2021)

[43] Shawn, Y.: FCN for crack recognition. https://github.com/OnionDoctor/FCN for crack recognition. Accessed March 13, 2018 (2018)

[44] Pi Ko, Samuel A. Prieto, and Borja Garcia de Soto. Developing a free and open-source automated building exterior crack inspection software for construction and facility managers. https://arxiv.org/abs/2206.09742, 2022. arXiv preprint arXiv:2206.09742.

[45] Xiang Zhang. Simple understanding of mask r-cnn. https://alittlepain833.medium.com/simple-understanding-of-mask-rcnn-134b5b330e95, April 2018b.

[46] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Scaled-yolov4: Scaling cross stage partial network. arXiv preprint arXiv:2011.08036, November 2020.

[47] Ataei, S. (2023). Detection and Evaluation of Damage in Concrete Structures Using Data-Driven Methods (Order No. 32166062). Available from ProQuest One Academic. (3237858202). https://www2.lib.ku.edu/login?url=https://www.proquest.com/dissertations-theses/detection-evaluation-damage-concrete-structures/docview/3237858202/se-2

[48] Ataei, S. T., Zadeh, P. M., & Ataei, S. (2025). Vision-based autonomous structural damage detection using data-driven methods. arXiv preprint arXiv:2501.16662.