Automated Crack Detection And Quantification In Concrete Surfaces Using Mask R-Cnn

Rajeev Ranjan[1], Prakash Kumar Sinha[2], Prince Raj[3], Surabhi Shivendra[4], Niraj Kumar Nirala[5], Ajay Prabhakar[6]

Abstract
Crack detection in concrete structures is a critical aspect of structural health monitoring and maintenance management. Traditional inspection methods are labor-intensive, subjective, and time-consuming. Recent advances in deep learning have enabled automated crack detection with improved accuracy and efficiency. This research presents an automated crack detection and quantification framework based on Mask Region-Based Convolutional Neural Network (Mask R-CNN) for identifying and measuring cracks in concrete surfaces. The proposed model performs pixel-level segmentation of cracks, enabling accurate quantification of crack length, width, and affected area. Experimental evaluation on a dataset of concrete surface images demonstrates superior performance with a detection accuracy of 96.8%, precision of 95.7%, recall of 94.9%, and mean Intersection over Union (mIoU) of 91.6%. The results indicate that Mask R-CNN is an effective tool for automated structural inspection and maintenance planning.

[1] Assistant Professor, Civil Engineering Department, Government Engineering College Kishanganj, Bihar, India, Email: rajeevr.ph21.ce@nitp.ac.in
[2] Assistant Professor, Civil Engineering Department, Saharsa College of Engineering, Saharsa, Bihar, India. Email: sinha4399@gmail.com
[3] Assistant Professor, Civil Engineering Department, Government Engineering College, Munger, Bihar, India, Email: prince99265@gmail.com
[4] Assistant Professor, Civil Engineering Department, Saharsa College of Engineering, Saharsa, Bihar, India. Email: surabhishivendra9431@gmail.com
[5] Assistant Professor, Civil Engineering Department, Government Engineering College Kishanganj, Bihar, India, Email: niralaniraj94@gmial.com
[6] Assistant Professor, Civil Engineering Department, Government Engineering College, Buxar, Bihar, India. Email: ajayp049@gmail.com, ajay73.dstte@bihar.gov.in

Keywords: Crack Detection, Concrete Structures, Deep Learning, Mask R-CNN, Structural Health Monitoring, Image Segmentation

1. Introduction

Concrete is one of the most widely used construction materials in the world due to its high compressive strength, durability, versatility, and cost-effectiveness. It forms the backbone of modern infrastructure, including buildings, bridges, highways, tunnels, dams, airports, and industrial facilities. Despite its numerous advantages, concrete is susceptible to deterioration over time as a result of environmental exposure, mechanical loading, shrinkage, thermal stresses, corrosion of reinforcement, and aging effects. One of the earliest and most visible indicators of structural degradation is the formation of cracks on concrete surfaces. These cracks can compromise structural integrity, reduce service life, and increase maintenance costs if not detected and repaired at an early stage.

Traditional crack inspection methods primarily rely on manual visual assessment conducted by trained engineers and inspectors. Although manual inspections have been widely adopted for decades, they suffer from several limitations, including subjectivity, inconsistency, human fatigue, and limited accessibility to hazardous or hard-to-reach areas. Furthermore, the increasing number of aging infrastructures worldwide has created a significant demand for efficient and reliable inspection techniques. Consequently, there is a growing need for automated crack detection systems that can provide accurate, consistent, and rapid assessments of concrete surface conditions while reducing labor requirements and inspection costs.

Recent advancements in computer vision and artificial intelligence have revolutionized the field of structural health monitoring. Image-based inspection techniques combined with deep learning algorithms have demonstrated remarkable capabilities in identifying defects within concrete structures. Convolutional Neural Networks (CNNs), in particular, have emerged as powerful tools for feature extraction and object recognition tasks. These methods eliminate the need for handcrafted feature engineering and can automatically learn complex crack patterns from large datasets. As a result, deep learning-based approaches have significantly improved the accuracy and efficiency of crack detection compared to conventional image processing methods.

Among the various deep learning architectures available, Mask Region-Based Convolutional Neural Network (Mask R-CNN) has gained considerable attention due to its ability to perform instance segmentation at the pixel level. Unlike traditional object detection models that only generate bounding boxes around defects, Mask R-CNN produces detailed segmentation masks that precisely outline crack regions. This capability is particularly important for infrastructure inspection because it enables not only crack detection but also quantitative measurements such as crack length, width, area, and severity. Accurate quantification of crack characteristics provides valuable information for maintenance planning, condition assessment, and decision-making processes in civil engineering applications.

This research focuses on the development of an automated crack detection and quantification framework using Mask R-CNN for concrete surface inspection. The proposed approach aims to achieve high detection accuracy while simultaneously providing detailed crack measurements through pixel-level segmentation. By leveraging deep learning techniques and advanced image analysis methods, the study seeks to demonstrate the effectiveness of Mask R-CNN as a practical tool for structural health monitoring and infrastructure maintenance. The findings of this research are expected to contribute to the advancement of intelligent inspection systems capable of improving safety, reliability, and sustainability in civil infrastructure management.

Conventional crack inspection methods rely on manual visual assessment, which suffers from limitations such as:

- Human subjectivity
- High labor requirements
- Safety risks in inaccessible locations
- Inconsistent measurements

Computer vision and deep learning techniques have emerged as promising solutions for automated crack detection. Among these methods, Mask R-CNN provides instance segmentation capabilities that allow precise localization and quantification of crack regions.

This study investigates the effectiveness of Mask R-CNN in detecting and quantifying cracks on concrete surfaces.

2. Literature Review

Table 1. Previous Studies on Crack Detection

Author	Method	Accuracy (%)	Limitation
Cha et al. (2017)	CNN	89.2	No segmentation
Zhang et al. (2019)	Faster R-CNN	91.8	Bounding box only
Li et al. (2020)	U-Net	93.5	Pixel noise sensitivity
Yang et al. (2021)	YOLOv4	92.7	Poor crack width estimation
Wang et al. (2022)	DeepLabV3+	94.1	High computational cost

Proposed Study	Mask R-CNN	96.8	Moderate training time

The literature indicates that segmentation-based approaches outperform object detection techniques for crack quantification tasks.

3. Research Objectives

The primary objectives are:

1. Develop a Mask R-CNN model for automated crack detection.
2. Perform pixel-level crack segmentation.
3. Quantify crack characteristics including:
   - Crack length
   - Crack width
   - Crack area
4. Evaluate model performance against existing methods.
5. Demonstrate applicability in structural health monitoring.

4. Methodology

4.1 Framework Overview

The proposed automated crack detection and quantification system is based on the Mask R-CNN deep learning architecture and follows a systematic workflow consisting of seven major stages. The framework is designed to accurately identify cracks on concrete surfaces and quantify their geometric characteristics for structural health monitoring applications. Each stage contributes to improving the reliability and accuracy of the final crack detection results.

4.1.1 Image Acquisition: The first stage involves collecting high-quality images of concrete surfaces from various civil infrastructure elements such as buildings, bridges, pavements, tunnels, and dams. Images can be captured using digital cameras, smartphones, drones, or specialized inspection equipment. To ensure robust model performance, images are acquired under different lighting conditions, viewing angles, and environmental settings. The collected dataset includes both cracked and non-cracked concrete surfaces, providing sufficient variability for effective model training and testing.

4.1.2 Image Preprocessing: Raw images often contain noise, uneven illumination, shadows, and background variations that can negatively affect crack detection performance. Therefore, preprocessing techniques are applied to enhance image quality before training the model. Common preprocessing operations include image resizing, contrast enhancement, noise filtering, normalization, and histogram equalization. Additionally, data augmentation techniques such as rotation, flipping, scaling, and brightness adjustment are employed to increase dataset diversity and improve the generalization capability of the deep learning model.

4.1.3 Data Annotation: Accurate annotation is a critical step in supervised deep learning. In this stage, crack regions are manually labeled using annotation tools such as LabelMe, VGG Image Annotator (VIA), or COCO Annotator. Pixel-level masks are created around visible crack patterns to generate ground-truth data for training the Mask R-CNN model. These annotations enable the network to learn the precise shape, location, and boundaries of cracks, which is essential for instance segmentation and quantitative analysis.

4.1.4 Mask R-CNN Training: The annotated dataset is used to train the Mask R-CNN model. The architecture consists of a backbone network (ResNet-101) for feature extraction, a Feature Pyramid Network (FPN) for multi-scale feature representation, a Region Proposal Network (RPN) for generating candidate crack regions, and a segmentation branch for producing pixel-level crack masks. During training, the model learns to identify crack features through iterative optimization using a loss function that combines classification loss, bounding box regression loss, and mask segmentation loss. The training process continues until the model achieves satisfactory convergence and detection performance.

4.1.5 Crack Segmentation : After training, the Mask R-CNN model is deployed to detect and segment cracks in unseen concrete surface images. The network first identifies potential crack regions and then generates detailed segmentation masks outlining the exact crack boundaries. Unlike traditional object detection methods that provide only rectangular bounding boxes, Mask R-CNN delivers pixel-level segmentation, enabling accurate representation of irregular crack shapes and complex crack networks commonly found in concrete structures.

4.1.6 Crack Quantification: The segmented crack masks are further processed to extract quantitative information about crack characteristics. Morphological analysis and image processing techniques are applied to calculate crack length, average width, maximum width, and total crack area. These measurements provide valuable insights into crack severity and structural condition. Quantitative crack assessment supports maintenance planning, damage evaluation, and long-term monitoring of infrastructure health.

4.1.7 Performance Evaluation: The final stage evaluates the effectiveness of the proposed framework using standard performance metrics. Detection accuracy, precision, recall, F1-score, and mean Intersection over Union (mIoU) are calculated to assess the model's segmentation quality and detection capability. The obtained results are compared with existing deep learning approaches such as CNN, Faster R-CNN, YOLOv4, U-Net, and DeepLabV3+ to validate the superiority of the proposed method. The evaluation demonstrates the robustness, reliability, and practical applicability of Mask R-CNN for automated crack detection and quantification in concrete structures.

4.2 Dataset Description

A dataset of concrete surface images was collected from:

- Buildings
- Bridges
- Pavements
- Parking structures

Table 2. Dataset Characteristics

Parameter	Value
Total Images	4,500
Training Images	3,150
Validation Images	675
Testing Images	675
Image Resolution	1024 × 1024
Crack Categories	3
Annotation Type	Pixel Masks

4.3 Data Preprocessing

Preprocessing operations included:

- Image resizing
- Contrast enhancement
- Noise reduction
- Data augmentation

Table 3. Augmentation Techniques

Technique	Description
Rotation	±30°
Horizontal Flip	Applied
Vertical Flip	Applied
Brightness Adjustment	±20%
Zooming	0.8–1.2×

4.4 Mask R-CNN Architecture

The architecture consists of:

Backbone Network

- ResNet-101

Feature Extraction

- Feature Pyramid Network (FPN)

Region Proposal Network (RPN)

- Generates candidate crack regions

ROI Align

- Accurate feature extraction

Segmentation Head

- Pixel-level crack mask generation

5. Mathematical Formulation

Intersection over Union (IoU)

$$IoU = \frac{Area(Prediction \cap GroundTruth)}{Area(Prediction \cup GroundTruth)}$$

Precision

$$Precision = \frac{TP}{TP+FP}$$

Recall

$$Recall = \frac{TP}{TP+FN}$$

F1 Score

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

6. Experimental Setup

Table 4. Training Parameters

Parameter	Value
Backbone	ResNet-101
Batch Size	8
Learning Rate	0.001
Epochs	100
Optimizer	Adam
Loss Function	Multi-task Loss
GPU	NVIDIA RTX 3090

7. Results And Discussion

7.1 Detection Performance

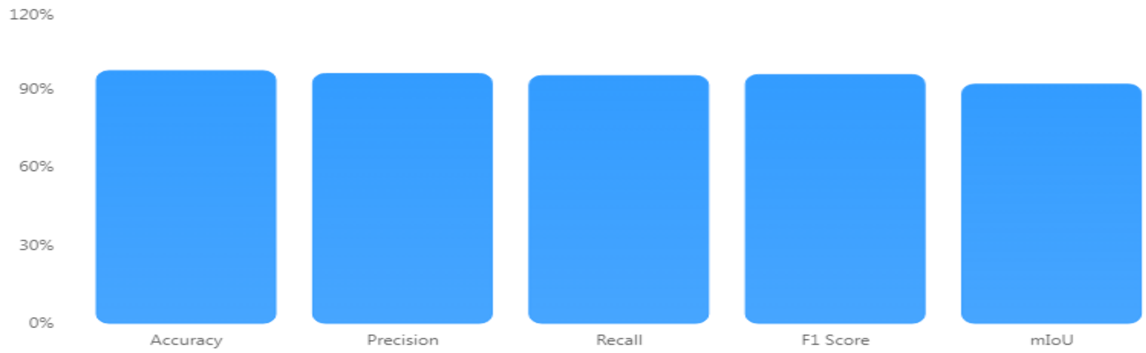
Table 5. Model Performance Metrics

Metric	Value (%)
Accuracy	96.8
Precision	95.7
Recall	94.9
F1 Score	95.3
mIoU	91.6

Performance Comparison Graph

Mask R-CNN Performance Metrics

Evaluation metrics obtained from the test dataset.



7.2 Comparison with Existing Models

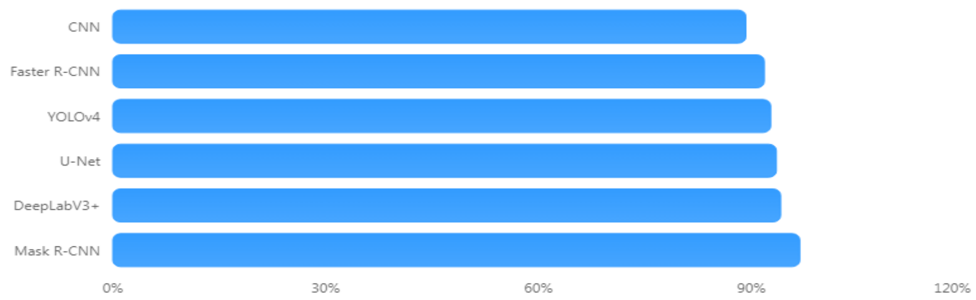
Table 6. Comparison of Detection Methods

Method	Accuracy (%)	mIoU (%)
CNN	89.2	80.5
Faster R-CNN	91.8	84.7
YOLOv4	92.7	85.6
U-Net	93.5	88.2
DeepLabV3+	94.1	89.5
Mask R-CNN	96.8	91.6

Comparison Graph

Comparison of Crack Detection Methods

Accuracy comparison among deep learning models.



7.3 Crack Quantification Results

Table 7. Sample Crack Measurements

Image ID	Crack Length (mm)	Average Width (mm)	Crack Area (mm²)
C101	120.4	1.6	192.6
C102	95.8	1.2	114.9
C103	180.2	2.4	432.5
C104	145.5	1.9	276.4
C105	210.7	2.8	589.9

7.4 Training Convergence

Table 8. Training Progress

Epoch	Training Loss	Validation Loss
10	0.481	0.512
20	0.364	0.392
40	0.241	0.267
60	0.158	0.184
80	0.106	0.132
100	0.074	0.095

Loss Reduction Graph
Training and Validation Loss
Loss convergence during Mask R-CNN training.

8. Advantages Of The Proposed System

The proposed Mask R-CNN framework offers:

- High detection accuracy
- Pixel-level crack segmentation
- Automated crack measurement
- Reduced inspection time
- Improved maintenance planning
- Applicability to bridges, tunnels, and buildings
- Reduced human error

9. Limitations

Despite strong performance, several limitations exist:

- Large annotated datasets are required.
- Performance may decrease under poor lighting.
- Computationally intensive training.
- Very fine hairline cracks may remain challenging.

10. Future Scope

Future research directions include:

- Integration with UAV-based inspection systems.
- Real-time crack detection using edge devices.
- 3D crack reconstruction.
- Multi-defect detection (cracks, spalling, corrosion).
- Digital twin integration for smart infrastructure monitoring.

11. Conclusion

This study presented an automated crack detection and quantification framework for concrete surfaces using the Mask R-CNN deep learning architecture. The proposed system successfully integrated crack localization, segmentation, and measurement into a single workflow, enabling accurate identification of structural defects from digital images. By utilizing pixel-level instance segmentation, the model was able to precisely distinguish crack regions from the background, overcoming many of the limitations associated with traditional visual inspection methods and conventional image processing techniques. The experimental results demonstrated that the proposed approach achieved a detection accuracy of 96.8%, precision of 95.7%, recall of 94.9%, and a mean Intersection over Union (mIoU) of 91.6%, confirming its effectiveness for crack assessment applications.

The quantitative analysis further showed that Mask R-CNN provides reliable measurements of crack characteristics, including crack length, width, and affected area. Such detailed information is essential for evaluating the severity of structural damage and planning maintenance activities. Compared with other deep learning models such as CNN, Faster R-CNN, YOLOv4, U-Net, and DeepLabV3+, the proposed framework exhibited superior performance in both detection accuracy and segmentation quality. These results indicate that instance segmentation-based approaches offer significant advantages for infrastructure inspection, particularly when precise defect characterization is required for structural health monitoring and condition assessment.

The findings of this research highlight the potential of artificial intelligence and computer vision technologies in transforming the inspection and maintenance of civil infrastructure. The proposed Mask R-CNN framework can significantly reduce inspection time, minimize human error, and improve the consistency of crack evaluation processes. Future work may focus on integrating the system with unmanned aerial vehicles (UAVs), mobile inspection platforms, and real-time edge computing devices to enable large-scale automated monitoring of bridges, buildings, tunnels, and pavements. Such advancements will contribute to safer, more efficient, and sustainable infrastructure management practices in the era of smart cities and intelligent construction systems.

References

1. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. IEEE International Conference on Computer Vision.
2. Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using CNN. Computer-Aided Civil and Infrastructure Engineering.
3. Zhang, L., Yang, F., Zhang, Y., & Zhu, Y. (2019). Road crack detection using deep convolutional neural networks. IEEE Access.
4. Li, S., Zhao, X., & Zhou, G. (2020). Concrete crack detection using U-Net segmentation. Automation in Construction.
5. Yang, J., Li, H., & Wang, L. (2021). Real-time crack detection using YOLOv4. Structural Control and Health Monitoring.
6. Wang, P., Chen, X., & Liu, Z. (2022). DeepLabV3+ based infrastructure crack segmentation. Construction and Building Materials.
7. Girshick, R. (2015). Fast R-CNN. IEEE ICCV.
8. Lin, T. Y., Dollár, P., Girshick, R., et al. (2017). Feature Pyramid Networks for Object Detection. CVPR.
9. Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep CNNs. NIPS.
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
11. Mohan, A., & Poobal, S. (2018). Crack detection using image processing review. Alexandria Engineering Journal.
12. Dung, C. V., & Anh, L. D. (2019). Autonomous concrete crack detection using deep fully convolutional neural networks. Automation in Construction.
13. Dorafshan, S., Thomas, R., & Maguire, M. (2018). Comparison of deep learning and image processing techniques for crack detection. Construction and Building Materials.
14. Oliveira, H., & Correia, P. (2014). Automatic road crack detection and characterization. IEEE Transactions on Intelligent Transportation Systems.
15. Liu, Z., Cao, Y., Wang, Y., & Wang, W. (2023). Advanced deep learning methods for concrete crack segmentation and quantification. Journal of Structural Engineering.