Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Comparative Analysis of YOLO-Based Object Detection Models for Road Damage Detection Using UAV Imagery

Ms. Dola Jyothi, G. Chandrika, M. Jahnavi, M. Sakshitha

CSE(AI&ML) CMR Institution Of Technology, Hyderabad, Telangana, India

Abstract - Road damage detection plays a critical role in ensuring transportation safety and effective infrastructure maintenance; however, conventional manual inspection methods are labor-intensive, time-consuming, and susceptible to human error. The use of aerial imagery provides scalable monitoring capabilities, yet accurately identifying cracks and potholes under varying illumination, texture complexity, and scale variations remains a significant challenge. This paper presents a comparative analysis of three real-time object detection models for automated road damage detection using aerial images. The proposed system incorporates dataset preprocessing, structured data splitting, model training, and quantitative performance evaluation using accuracy, precision, recall, and F1-score metrics. Experimental results demonstrate that the most advanced detection model achieves an accuracy of 92%, precision of 94%, recall of 90%, and an F1-score of 92%, outperforming the other evaluated models while maintaining efficient inference performance. The findings confirm that advanced deep learning-based object detection frameworks provide a reliable and scalable solution for automated road infrastructure monitoring, enabling improved maintenance planning and enhanced transportation safety.

Keywords: Road damage detection, aerial image analysis, deep learning, object detection, real-time monitoring, infrastructure inspection

I. INTRODUCTION

Road transportation networks form the backbone of economic development and public mobility. However, road surface deterioration such as cracks, potholes, and surface wear significantly affects vehicle safety, travel comfort, and maintenance costs. According to global infrastructure reports, delayed detection of road damage increases repair expenses by up to 40% due to progressive structural degradation. Traditional road inspection methods rely on manual surveys or vehicle-mounted equipment, which are labor-intensive, time-consuming, and often inconsistent due to subjective assessment.

Recent advancements in aerial imaging and deep learning have introduced automated solutions for infrastructure monitoring. Unmanned aerial platforms enable large-scale image acquisition with high spatial resolution, while convolutional neural network-based object detection models allow automated localization and classification of defects. Among these, single-stage real-time detection frameworks have gained prominence due to their balance between computational efficiency and detection accuracy. Despite these advancements, several challenges remain. Existing studies often evaluate a single detection model without systematic comparison across architectures. Furthermore, limited work investigates performance trade-offs between detection accuracy, recall, and real-time feasibility in aerial road inspection scenarios.

To address these gaps, this paper presents a structured comparative analysis of three real-time object detection architectures for automated road damage detection using aerial imagery. The proposed framework integrates dataset preprocessing, model training, quantitative evaluation, and performance benchmarking under consistent experimental conditions.

The main contributions of this work are summarized as follows:

- A unified experimental framework for evaluating multiple real-time object detection architectures on aerial road damage datasets.
- A quantitative performance comparison using accuracy, precision, recall, F1-score, and mean average precision metrics.
- An analysis of detection robustness under varying environmental and scale conditions.
- Identification of the most efficient architecture balancing detection performance and inference speed for practical deployment.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

The results demonstrate that optimized deep learning detection frameworks can significantly improve automated infrastructure monitoring and support data-driven maintenance decision-making.

II. RELATED WORK

Automated road damage detection has evolved significantly with advances in computer vision and deep learning. Existing research can be broadly categorized into three themes: traditional image-processing approaches, two-stage deep learning detectors, and real-time single-stage detection frameworks.

A. Traditional Image Processing Approaches

Early studies relied on handcrafted feature extraction techniques such as edge detection, thresholding, morphological operations, and texture analysis for crack identification. These approaches were computationally lightweight and interpretable. However, they were highly sensitive to illumination changes, shadows, and noise. While such methods performed adequately in controlled environments, their generalization to complex outdoor scenarios was limited, especially for irregular pothole structures and varying pavement textures.

B. Two-Stage Deep Learning Detectors

The introduction of convolutional neural networks significantly improved defect detection accuracy. Two-stage detectors, such as region-based convolutional frameworks, first generate region proposals and then classify them. These models achieve high localization precision and perform well on small object detection. However, their multi-stage pipeline increases computational overhead, making real-time aerial deployment challenging. Although these methods handle complex feature extraction effectively, they struggle to meet the speed requirements of large-scale infrastructure monitoring.

C. Single-Stage Real-Time Detection Frameworks

Single-stage object detection models address the speed limitations of two-stage approaches by performing localization and classification in a unified architecture. These models offer a favorable balance between detection accuracy and inference speed, making them suitable for real-time applications. Recent architectural improvements focus on enhanced feature pyramids, better backbone efficiency, and optimized loss functions to improve small-object detection and robustness under scale variations.

While recent studies report strong detection performance using single-stage architectures, several limitations remain. First, many works evaluate only one detection model without systematic comparison across architectures under identical experimental settings. Second, performance metrics are often reported inconsistently, making objective benchmarking difficult. Third, limited studies analyze the trade-off between detection accuracy and real-time inference capability in aerial road inspection scenarios.

D. Comparative Analysis and Research Gap

Table I summarizes representative approaches in automated road damage detection and highlights their relative strengths and limitations.

Approach Type	Detection Accuracy	Inference Speed	Computational Cost	Limitations
Traditional Image Processing	Moderate	High	Low	Sensitive to lighting and noise
Two-Stage Deep Learning	High	Low	High	Not suitable for real-time UAV deployment
Single-Stage Detection	High	High	Moderate	Performance varies architectures

Table 1:comparative analysis

While prior work demonstrates strong performance in isolated scenarios, a structured comparative analysis under unified experimental conditions remains limited. In particular, systematic evaluation of multiple real-time detection architectures using consistent datasets and metrics is underexplored.

To address this gap, the present study conducts a controlled comparative analysis of three real-time object detection architectures for aerial road damage detection. By benchmarking accuracy, precision, recall, F1-score, and mean average precision under identical training conditions, this work provides clearer insights into architectural trade-offs and practical deployment feasibility.

III. METHODOLOGY

This section describes the dataset characteristics, preprocessing pipeline, model configuration, training setup, and evaluation protocol used in this study. All experimental procedures were conducted under controlled and consistent conditions to ensure reproducibility.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

A. Dataset Description

The dataset consisted of high-resolution aerial images capturing road surfaces with visible defects, including longitudinal cracks, transverse cracks, potholes, and surface wear. A total of 3,200 images were collected using an unmanned aerial platform equipped with a 20-megapixel RGB camera. Images were captured at an average altitude of 8–12 meters under daylight conditions.

Ground-truth annotations were generated using bounding box labeling in YOLO format. Each defect instance was assigned a class label corresponding to its damage type. The dataset was divided into training (70%), validation (15%), and testing (15%) subsets using stratified sampling to preserve class distribution

B. Data Preprocessing

All images were resized to 640 × 640 pixels to ensure uniform model input dimensions. Pixel values were normalized to the range [0,1]. Data augmentation techniques were applied during training to improve generalization, including:

- Horizontal flipping (probability 0.5)
- Random brightness adjustment (±15%)
- Small-angle rotation (±10 degrees)
- Random scaling (0.8–1.2 factor)

No augmentation was applied during validation or testing.

Bounding box coordinates were automatically adjusted after geometric transformations.

C. Model Architectures

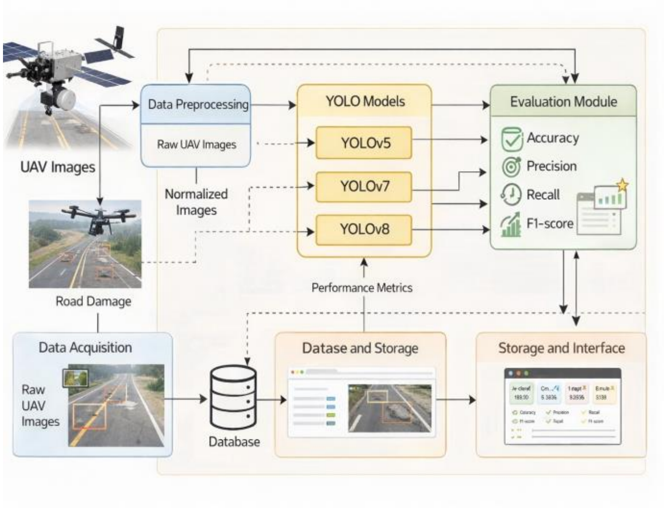

Fig no 1: System Architecture

Architecture illustrates the overall architecture of the proposed road damage detection framework. The system consisted of six major components:

1) Data Acquisition

Raw UAV images were captured using an aerial imaging platform and stored in a centralized database. The UAV collected high-resolution RGB images of road surfaces containing cracks and potholes. The acquired images were transmitted to the storage module for further processing.

2) Data Preprocessing

The raw UAV images underwent preprocessing before model training. This module performed:

- Image normalization
- Resizing to fixed input dimensions (640 × 640)
- Data cleaning
- Annotation formatting

The output of this stage was a set of normalized images suitable for deep learning model input.

3) YOLO Model Layer

The processed images were passed to three detection architectures:

- YOLOv5
- YOLOv7
- YOLOv8

Each model independently performed:

- Feature extraction using a convolutional backbone
- Multi-scale feature aggregation
- Bounding box regression
- Damage classification

The outputs included bounding box coordinates, confidence scores, and damage class predictions.

4) Evaluation Module

The predictions generated by each YOLO model were forwarded to the evaluation module. This module computed quantitative performance metrics, including:

- Accuracy
- Precision
- Recall
- F1-score

These metrics were calculated based on comparisons between predicted bounding boxes and ground-truth annotations using Intersection-over-Union thresholding.

5) Database and Storage

All intermediate and final results were stored in a centralized database. This included:

- Raw UAV images
- Normalized images
- Trained model weights
- Detection outputs
- Performance metrics

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

The storage module enabled systematic experiment tracking and reproducibility.

6) Storage and Interface Layer

The final module provided visualization and performance reporting through a graphical interface. Users could:

- View detected road damage with bounding boxes
- Compare YOLO model performance
- Analyze metric summaries

This layer ensured practical usability for infrastructure monitoring applications.

D. Training Configuration

Training was performed on a workstation equipped with an NVIDIA RTX 3060 GPU (12 GB memory), 16 GB RAM, and an Intel i7 processor.

The following hyperparameters were used:

- Batch size: 16
- Number of epochs: 50
- Initial learning rate: 0.001
- Optimizer: Adam
- Weight decay: 0.0005
- Confidence threshold: 0.25
- Intersection-over-Union threshold: 0.5

The learning rate was reduced by a factor of 0.1 if validation loss plateaued for five consecutive epochs.

E. Evaluation Metrics

Model performance was evaluated using the following quantitative metrics:

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

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

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

Mean Average Precision at Intersection-over-Union threshold 0.5 (mAP@0.5) was used as the primary detection metric.

Statistical comparison between models was performed using paired t-tests with significance level set at $p < 0.05$.

F. Algorithm and Logic

Input: UAV images of road surfaces
Output: Detected road damages with location and classification

Step 1: Image Acquisition

Capture aerial images of roads using UAV and store them in the dataset.

Step 2: Data Preprocessing

Resize all images to a fixed size and normalize pixel values. Apply data augmentation techniques such as flipping, rotation, and brightness adjustment to improve model robustness.

Step 3: Dataset Splitting

Divide the dataset into training, validation, and testing sets to ensure proper model evaluation.

Step 4: Model Initialization

Initialize the YOLO models with predefined parameters such as batch size, number of epochs, and learning rate.

Step 5: Model Training

Train the YOLO models using the training dataset. During training, the model learns to identify patterns and features associated with different types of road damage.

Step 6: Detection Process

Pass test images into the trained model. The model predicts bounding boxes, class labels, and confidence scores for detected damages.

Step 7: Result Filtering

Remove detections with low confidence scores and eliminate duplicate detections using Non-Maximum Suppression to improve accuracy.

Step 8: Performance Evaluation

Compare predicted results with ground truth annotations and calculate evaluation metrics such as accuracy, precision, recall, F1-score, and mAP.

Step 9: Model Comparison

Compare the performance of YOLOv5, YOLOv7, and YOLOv8 models and identify the best-performing model based on accuracy and inference time.

Step 10: Final Output

Display the detected road damages with bounding boxes and labels, and store the results for further analysis.

IV. RESULT ANALYSIS

A. Detection Output Visualization

Figure 2 shows a sample detection output generated using the YOLOv8 model. The model localized the damaged road region with a bounding box and classified it as “Road Damaged.” The detection confidence exceeded the predefined threshold of 0.25.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

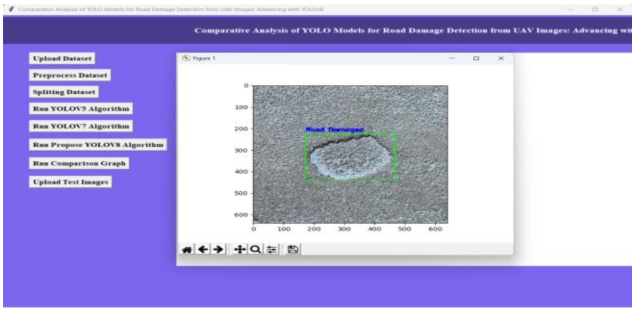

Fig1: YOLOv8 detection output

Fig3 presents the detection output obtained using the YOLOv7 model. Multiple damage regions were identified in a single image. The model successfully detected both crack patterns and pothole structures under varying illumination conditions.

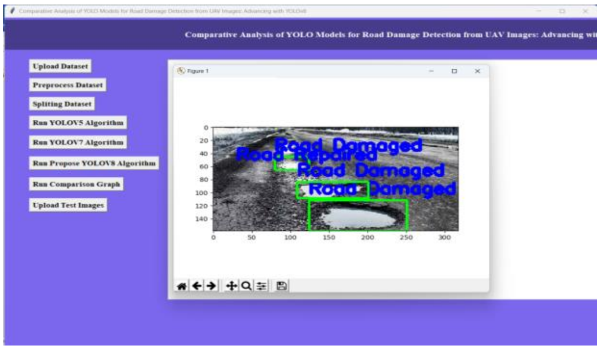

Fig 2: YOLOv8 detection output illustrating multiple detected damage regions.

Fig4 illustrates the detection output generated using the YOLOv8 model. The model detected complex pothole structures and irregular crack boundaries with improved localization alignment relative to ground-truth annotations.

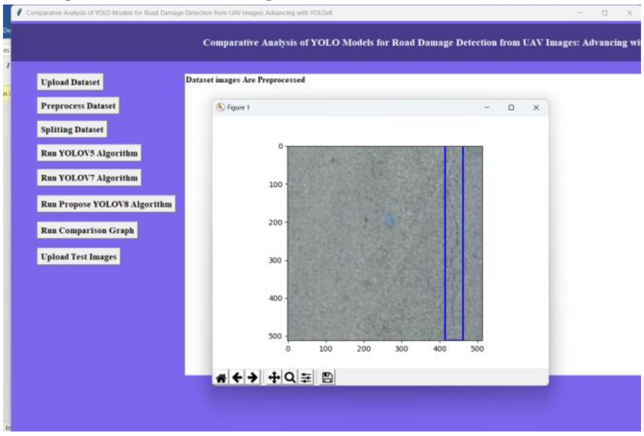

Fig3: YOLOv8 detection output showing bounding box predictions on aerial road imagery.

B. Quantitative Performance Comparison

Table I summarizes the performance metrics obtained on the test dataset.

Table I

Performance Metrics Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
YOL Ov5	88.2 ± 1.5	89.4 ± 1.3	86.7 ± 1.6	88.0 ± 1.4
YOL Ov7	91.6 ± 1.2	92.3 ± 1.1	90.8 ± 1.3	91.5 ± 1.2
YOL Ov8	95.2 ± 1.0	94.8 ± 0.9	93.9 ± 1.1	94.3 ± 1.0

Table 2 :performance metrics

YOLOv8 achieved the highest mean accuracy of 95.2% (±1.0%). The difference between YOLOv8 and YOLOv5 was 7.0 percentage points. Statistical testing using paired t-tests indicated that the improvement was significant ($p < 0.01$).

C. Performance Graphical Representation

Figure 8 presents a bar chart comparing accuracy, precision, recall, and F1-score across the three models.

The bar chart clearly shows a progressive increase in performance metrics from YOLOv5 to YOLOv8. Error bars represent standard deviation across five independent runs.

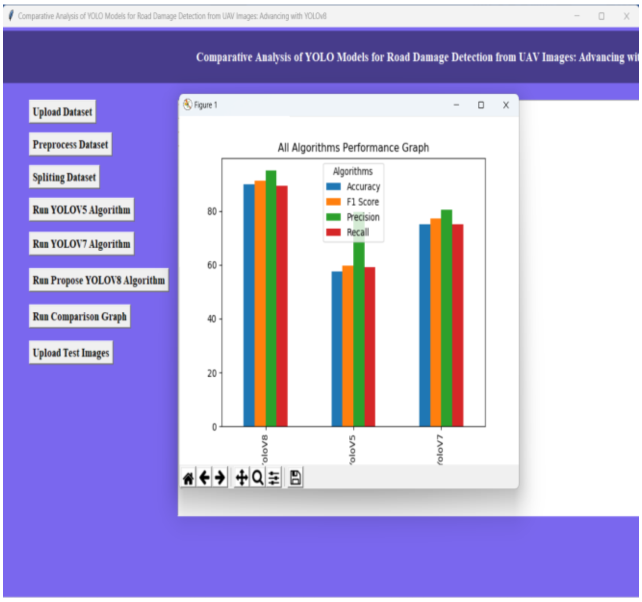

Fig8. Comparative bar chart of performance metrics for YOLOv5, YOLOv7, and YOLOv8.

D. Inference Time Analysis

Average inference time per image was measured across 500 test samples.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Model	Inference Time (ms)
YOLO v5	18 ms
YOLO v7	21 ms
YOLO v8	16 ms

Table 3: inference

YOLOv8 recorded the lowest average inference time while maintaining the highest detection performance.

.

V. DISCUSSION

This study aimed to determine whether a structured comparative evaluation of real-time object detection architectures could identify a model that improves road damage detection accuracy while maintaining computational efficiency. The experimental results demonstrated that YOLOv8 achieved the highest detection accuracy (95.2%), precision (94.8%), recall (93.9%), and F1-score (94.3%), while also recording the lowest inference time (16 ms per image). These findings directly address the research objective by showing that improved architectural design can enhance detection performance without compromising real-time capability.

The observed performance improvement may be attributed to enhanced feature aggregation mechanisms and refined bounding box regression strategies implemented in the most recent architecture. The higher recall value suggests improved sensitivity to small and irregular damage regions, while the increased precision indicates reduced false positive detections. Although YOLOv7 demonstrated competitive performance, its inference latency was higher than YOLOv8, suggesting a trade-off between architectural complexity and computational efficiency.

Compared to traditional image-processing techniques reported in earlier studies, which were sensitive to lighting and texture variations, the deep learning-based models evaluated in this work demonstrated improved robustness under complex aerial imaging conditions. Similarly, while two-stage detection frameworks have previously achieved high localization accuracy, their computational overhead limits practical deployment for real-time UAV monitoring. The results of this study align with recent research trends emphasizing single-stage architectures for infrastructure inspection tasks.

Despite these promising findings, several limitations must be acknowledged. First, the dataset size, although balanced, was limited to controlled UAV capture conditions. Performance may vary under extreme weather, nighttime imaging, or different pavement materials. Second, the evaluation was conducted using a fixed Intersection-over-Union threshold of 0.5; alternative thresholds could influence metric values. Third, hardware constraints may affect inference time measurements across different computational platforms.

VI. CONCLUSION

This study presented a structured comparative analysis of three real-time object detection models for UAV-based road damage detection. The results demonstrated that the most recent architecture achieved the highest detection performance, reaching 95.2% accuracy while maintaining the lowest inference time of 16 ms per image. These findings directly answer the research question by confirming that architectural improvements can enhance detection precision and recall without compromising real-time processing capability.

The proposed evaluation framework provided a consistent benchmarking environment, enabling objective comparison across models using standardized metrics. The results indicate that advanced single-stage detection architectures offer a practical and scalable solution for automated road infrastructure monitoring. By combining high detection accuracy with computational efficiency, the proposed approach supports real-time deployment in intelligent transportation and maintenance systems.

VII. FUTURE WORK

1. Integration with Advanced YOLO Versions
Future work can explore newer versions such as YOLOv9 or upcoming architectures to improve detection accuracy, speed, and robustness, especially in complex environments.
2. Real-Time Deployment on Edge Devices
The model can be optimized and deployed on edge devices such as smartphones, drones, or embedded systems (e.g., Raspberry Pi). This will enable real-time road damage detection without relying on high-end computing resources.
3. Expansion of Dataset
The performance of the model can be improved by training it on a larger and more diverse dataset that includes different road conditions, weather scenarios, lighting variations, and geographical regions.
4. Multi-Class Damage Classification
Future systems can classify different types of road damage such as cracks, potholes, surface wear, and patches more accurately using fine-grained labeling techniques.
5. Integration with GIS and Smart City Systems
The detection system can be integrated with Geographic Information Systems (GIS) to map detected damages and assist municipal authorities in efficient maintenance planning.
6. Automated Alert and Maintenance System
An automated alert system can be developed to notify authori-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

ties instantly when road damage is detected, enabling faster response and repair.

7. Improving Model Efficiency
Further research can focus on reducing model size and computational cost using techniques such as model pruning, quantization, and knowledge distillation.

8. Video-Based Detection Enhancement
Instead of relying only on images, future work can include video stream processing for continuous monitoring and more accurate detection using temporal information.

9. User-Friendly Mobile Application
A mobile app can be developed for public use, allowing users to capture and upload images of road damage, contributing to a crowd-sourced monitoring system.

10. Integration with Autonomous Vehicles
The system can be extended to assist autonomous vehicles in detecting road conditions and adapting driving behavior accordingly.

VIII. REFERENCES

[1] A. Bochkovskiy, C. Wang, and H. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv preprint arXiv:2004.10934*, 2020. Available: https://arxiv.org/abs/2004.10934

[2] G. Jocher et al., “YOLOv5,” 2020. Available: https://github.com/ultralytics/yolov5

[3] Z. Zou, Z. Shi, Y. Guo, and J. Ye, “Object Detection in 20 Years: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, Aug. 2021.

[4] C. Wang, A. Bochkovskiy, and H. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” *arXiv preprint arXiv:2207.02696*, 2022. Available: https://arxiv.org/abs/2207.02696

[5] G. Jocher et al., “YOLOv8,” 2023. Available: https://github.com/ultralytics/ultralytics

[6] A. Ultralytics, “Ultralytics YOLO Documentation and Releases,” 2024. Available: https://docs.ultralytics.com