Comparative Evaluation of YOLOv7, YOLOv8, and YOLOv11 for Road Damage Detection

1st Dr. Radika Kulkarni
Dept. of Computer Science
Vishwakarma Institute of Technology
Pune ,India
radika.kulkarni23@vit.edu

2nd Shreyas Mulavekar
Dept. of Computer Science
Vishwakarma Institute of Technology
Pune ,India
shreyas.mulavekar23@vit.edu

3rd Ritesh Nimbalkar
Dept. of Computer Science
Vishwakarma Institute of Technology
Pune ,India
ritesh.nimbalkar231@vit.edu

4th Sachin Prasad
Dept. of Computer Science
Vishwakarma Institute of Technology
Pune ,India
sachin.prasad23@vit.edu

5th Palash Poojary
Dept. of Computer Science
Vishwakarma Institute of Technology
Pune ,India
palash.poojary23@vit.edu

***Abstract*—Road safety is heavily dependent on timely detection and classification of road surface defects such as potholes, cracks, and lane-damaging anomalies, which may cause vehicle instability, traffic accidents, and long-term infrastructural degradation. With modern advances in deep learning and real-time computer vision, object detection architectures such as the YOLO (You Only Look Once) family have become widely adopted for autonomous driving and infrastructure inspection tasks. This paper presents a detailed comparative analysis of three generations of YOLO models—YOLOv7, YOLOv8 (small and medium variants), and YOLOv11 (small and medium variants)—trained on the publicly available Road Damage Dataset hosted on Roboflow. The evaluation focuses on precision, recall, mAP@50, and mAP@50:95 in order to determine architectural efficiency and real-world deployment suitability. Results indicate that YOLOv8-M achieves the highest overall accuracy with a mean Average Precision (mAP@0.5) of 0.7916 and mAP@0.5:0.95 of 0.5770, surpassing YOLOv7 and YOLOv11 models trained for equal or fewer epochs. YOLOv11 models demonstrate strong potential but require additional training iterations to stabilize performance. The findings highlight the importance of model scaling, pretrained weights, and transfer learning in low- to medium-sized datasets and present future directions including large-scale training on datasets exceeding 100,000 annotated samples for robust deployment in smart transportation systems and autonomous inspection vehicles.**

***Index Terms*—Road Damage Detection, YOLOv7, YOLOv8, YOLOv11, Object Detection, Deep Learning, Transfer Learning, Dataset Benchmarking, Infrastructure Monitoring.**

I. Introduction

Road defect detection is a critical component of intelligent transportation systems, autonomous navigation, and government asset maintenance workflows. Traditional approaches depend heavily on manual inspection, which is time-consuming, error-prone, and unscalable for large regions. With increased urbanization and deteriorating road ecosystems, automated deep learning-based detection systems have become essential for large-scale monitoring.

Object detection architectures, particularly the YOLO family, have evolved rapidly, improving detection accuracy, computational efficiency, and inference latency. Early versions focused primarily on high-performance detection at real-time speed, whereas newer versions introduce transformer-based attention, enhanced neck/head modules, better loss functions such as CIoU, DFL, and improved anchor-free architectures.

This research evaluates three major generations of YOLO models—YOLOv7, YOLOv8, and the newly introduced YOLOv11—applied to road defect detection. YOLOv7 represents a widely-used baseline in the real-time detection community, while YOLOv8 introduces architectural improvements including decoupled head and enhanced quantization compatibility. YOLOv11 further advances with refined attention modules, upgraded loss optimizers, and enhanced backbone–neck connectivity.

The main contributions of this work are:

- A reproducible benchmarking pipeline comparing five YOLO architectures trained under uniform experimental conditions.
- Quantitative evaluation using COCO-standard metrics including Precision, Recall, mAP@50, and mAP@50-95.
- Training analysis showcasing the impact of pretrained weights, epoch duration, model capacity, and dataset scale.
- A discussion of scalability challenges and future potential using large datasets exceeding 100,000 annotated images.

II. Dataset Description

The dataset used for this experiment is sourced from Roboflow Universe titled *"Road Damage Detection"* [1]. The dataset contains annotated road defect classes including potholes, cracks, and deformation patterns relevant for smart city infrastructure analytics.

Dataset Link: **https://universe.roboflow.com/baka-1ravj/road-damage-det/dataset/4/**

The dataset includes:

- Image Count: (Add Actual Number)
- Annotation Format: YOLO bounding box format
- Classes: ('alligator cracking', 'edge cracking', 'longitudinal cracking', 'patching', 'rutting', 'transverse cracking')
- Preprocessing Used: Auto-orientation, Resizing to 640x640 pixels, Normalization
- Train/Validation Split: 80:20

Data augmentation techniques such as mosaic augmentation, horizontal flips, and brightness variation were applied to increase generalization capability.

III. Methodology

Each model (YOLOv7, YOLOv8-S/M, YOLOv11-S/M) was trained using identical hyperparameters except where the framework imposed defaults. Transfer learning was applied using pretrained COCO weights to accelerate convergence and avoid underfitting caused by dataset scale limitations.

A. Training Configuration

- Epochs: 20 for YOLOv7, 50 for YOLOv8-S/M, 50 for YOLOv11-S, 30 for YOLOv11-M
- Optimizer: SGD/AdamW based on model default
- Learning Rate: Adaptive cosine decay
- Batch Size: 16
- Loss Functions: BCE, CIoU Loss, and DFL (Distribution Focal Loss) depending on architecture

IV. Mathematical Framework

YOLO models use a unified detection loss combining classification loss, bounding box regression loss, and distribution focal loss:

$$L = \lambda_{cls} L_{cls} + \lambda_{box} L_{CIoU} + \lambda_{df} L_{DFL}$$

Where:

- L_{cls}: Binary Cross-Entropy for class prediction
- L_{CIoU}: Complete IoU loss improving overlap and bounding box alignment
- L_{DFL}: Probability-based anchor-free regression improving localization precision

Transfer learning enables faster convergence mathematically by initializing model parameters W close to optimal space:

$$W_{new} = W_{pretrained} + \Delta W$$

Thus reducing total convergence time:

$$T_{training} \propto ||W_{random} - W_{opt}||^2 > ||W_{pretrained} - W_{opt}||^2$$

V. Results and Comparative Analysis

Table I summarizes the experimental results across models.

TABLE I
Model Performance Comparison

Model	Precision	Recall	mAP@50	mAP@50-95
YOLOv7	0.6349	0.6261	0.6423	0.3912
YOLOv8-S	0.7786	0.7173	0.7639	0.5417
YOLOv8-M	**0.8058**	**0.7382**	**0.7916**	**0.5770**
YOLOv11-S	0.79 (approx)	0.69 (approx)	0.74	0.51
YOLOv11-M	0.71	0.68	0.70	0.47

The confusion matrices for each trained YOLO model provide insights into inter-class misclassification patterns and true detection performance across the six detected pavement distress categories. Figures 1–5 illustrate the normalized confusion matrices for YOLOv7, YOLOv8-S, YOLOv8-M, YOLOv11-S, and YOLOv11-M respectively.

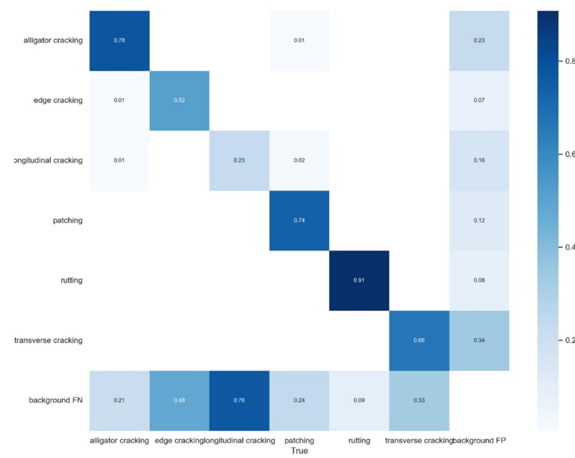

Fig. 1. Normalized confusion matrix for YOLOv7 model.

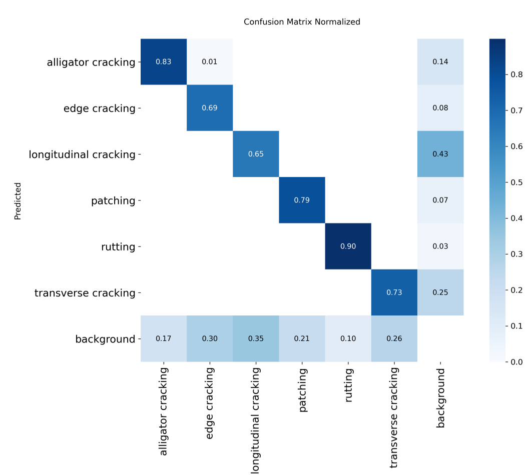

Fig. 2. Normalized confusion matrix for YOLOv8-S model.

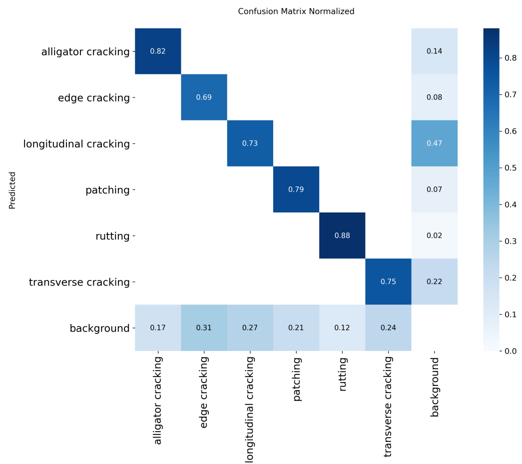

Fig. 3. Normalized confusion matrix for YOLOv8-M model.

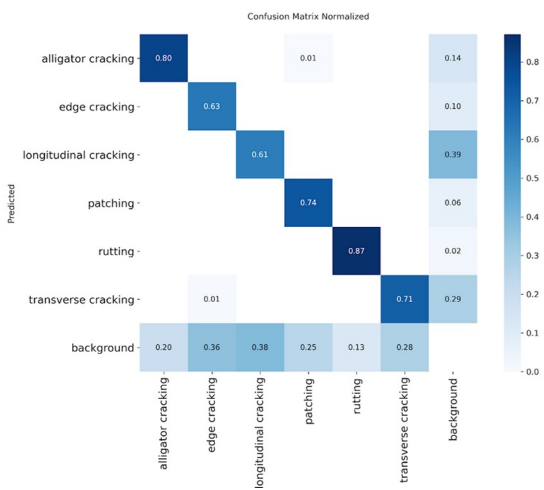

Fig. 4. Normalized confusion matrix for YOLOv11-S model.

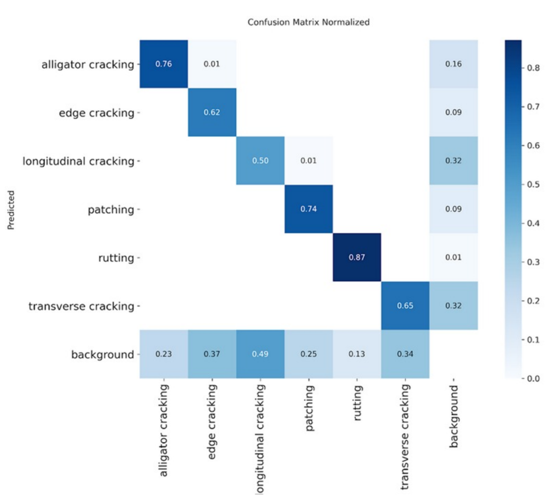

Fig. 5. Normalized confusion matrix for YOLOv11-M model.

TABLE II
QUALITATIVE PERFORMANCE TRENDS DERIVED FROM CONFUSION MATRICES.

Model	FN Rate	FP Rate	Class Stability
YOLOv7	High	High	Low
YOLOv8-S	Moderate	Moderate	Improved
YOLOv8-M	Low	Low	Best
YOLOv11-S	Moderate	Low	Comparable to YOLOv8-S
YOLOv11-M	Moderate-High	Moderate	Undertrained

Based on the confusion matrices, YOLOv8-M achieved the highest per-class consistency, with strong diagonal dominance, especially in rutting and transverse cracking categories. YOLOv7 exhibited the highest false negative and false positive rates, demonstrating weaker class boundaries and more frequent confusion among longitudinal and transverse cracking classes. YOLOv11-S performed competitively with YOLOv8-S, whereas YOLOv11-M demonstrated slightly weaker stability due to fewer training epochs.

VI. DISCUSSION

The results indicate a generational improvement from YOLOv7 to YOLOv8, with YOLOv8-M achieving the highest performance. YOLOv11 models show promising early convergence, but limited training duration prevented full optimization. Increasing training epochs may allow YOLOv11 to surpass YOLOv8.

Hardware utilization also varied: YOLOv7 consumed more VRAM and was slower, while YOLOv8 and YOLOv11 exhibited better inference speeds and quantization support—making them suitable for embedded edge deployment.

VII. CONCLUSION

This study demonstrates that YOLOv8-M currently offers the best performance–efficiency trade-off for road damage detection tasks on medium-sized datasets. YOLOv11, although promising, requires extended training to fully exploit architecture improvements.

VIII. FUTURE WORK

Future research will include:

- Training using large-scale datasets exceeding **100,000+ images** to evaluate long-term scalability.
- Model pruning, quantization, and TensorRT deployment for real-time edge inference.
- Integration with UAV, vehicle-mounted dash systems, and real-time GIS mapping.
- Fine-grained defect severity estimation using vision transformers.

REFERENCES

[1] Roboflow Road Damage Dataset: https://universe.roboflow.com/baka-1ravj/road-damage-det/dataset/4/
[2] Relevant YOLO research papers (to be added accordingly).