

Automated Detection of Road Defects for Driving Safety

Author 1 M JAYAVARDHINI Author 2 Mr.P.VALLARASU

Department of Master Of Computer Applications, [AVS Engineering college], [Salem]. [Tamilnadu]

Department of Artificial Intelligence And Data science, [AVS Engineering College], [Salem]. [Tamilnadu]

Abstract---

Road networks represent an essential factor in economic prosperity and wellbeing. Assorted damages and defects on the road surface mainly impact the transportation efficiency of road networks and driving safety. In current traditions, it may take weeks or even months before competent government departments repair such road conditions due to their lack of timely awareness. This research proposes a framework for collecting data and analysis using deep learning (DL) techniques, more specifically convolutional neural network (CNN) and YOLOv8 model for effective and efficient monitoring and assessment of road surface conditions. Data collected in four different scenarios and multiple vehicle data help to discover abnormal defects and verify automated detection. DL techniques were developed to identify bumps and potholes automatically, which are the two significant types of road defects. Cross-validation was set to validate the accuracy of the two decision-making strategies, CNN and YOLOv8. Automation technology is expected to provide qualitative and quantitative information about road defect conditions, enabling timely maintenance to improve transportation efficiency and driving safety.

Keywords---*Road defect detection; deep learning; convolutional neural network; YOLOv8; pothole detection; bump detection; computer vision; road safety; autonomous vehicles.*

I. INTRODUCTION

Road infrastructure forms the backbone of transportation systems worldwide, directly influencing economic development, trade efficiency, and public safety. The quality of road surfaces determines the safety and comfort of daily commutes for millions of road users. When road surfaces deteriorate through the formation of potholes and speed bumps, the consequences range from vehicle damage and increased fuel consumption to serious traffic accidents and fatalities. In many developing and developed nations alike, road defect detection and repair processes remain largely manual, reactive, and slow [1].

Traditional road inspection methods rely on human inspectors conducting periodic field surveys — a process that is labor-intensive, subjective, and unable to provide the timely data needed for proactive maintenance. The gap between defect occurrence and repair can span weeks or even months, during which the hazard remains unaddressed. The exponential growth of road networks has further widened this gap, making manual inspection increasingly impractical at scale [2]. The emergence of deep learning (DL) and computer vision has provided researchers and engineers with powerful tools to automate the detection and classification of road surface anomalies. Convolutional neural networks (CNNs) are particularly well-suited for image-based detection tasks due to their ability to learn hierarchical spatial features directly from raw image data without manual feature engineering. More

recently, the YOLO (You Only Look Once) family of real-time object detectors has demonstrated state-of-the-art performance in both detection speed and accuracy across diverse visual recognition domains [3].

This paper proposes a comprehensive automated road defect detection framework that integrates CNN-based classification with YOLOv8-based real-time detection to identify the two most prevalent types of road defects: potholes and speed bumps. The framework is validated across four distinct driving scenarios using multi-vehicle data collection, ensuring robustness across varied real-world conditions. The key contributions of this work are:

- (1) A multi-scenario dataset of 12,000 labeled road images collected from urban, rural, highway, and construction-zone environments.
- (2) A CNN architecture optimized for road defect classification, evaluated via five-fold cross-validation.
- (3) A fine-tuned YOLOv8 model for real-time, bounding-box-level defect detection.
- (4) An end-to-end reporting pipeline integrating GPS geo-tagging and a cloud-based maintenance dashboard.

II. RELATED WORKS

Early automated road inspection systems were built on classical image processing techniques. Thresholding, morphological operations, and edge-detection algorithms were applied to grayscale images to isolate regions of

surface damage [4]. While computationally efficient, these methods were highly sensitive to changes in lighting, pavement texture, and camera calibration, limiting their deployment in uncontrolled road environments.

The introduction of machine learning methods improved robustness considerably. Support vector machines (SVMs) trained on handcrafted feature descriptors such as histogram of oriented gradients (HOG) and local binary patterns (LBP) achieved reasonable classification accuracy on benchmark datasets. However, the reliance on manual feature engineering constrained generalization across different road types and weather conditions [5].

Deep learning approaches have since dominated the field. Maeda et al. [6] constructed a large-scale road damage dataset and benchmarked multiple CNN architectures, demonstrating that transfer learning from ImageNet significantly accelerates convergence and improves accuracy. Zhang et al. [7] proposed a lightweight detection network optimized for embedded deployment on vehicle-mounted edge devices. Fan et al. [8] explored stereo vision combined with CNNs to estimate pothole depth alongside surface detection.

The YOLOv5 and YOLOv8 architectures have attracted growing attention for pavement distress detection due to their real-time performance. Arya et al. [9] demonstrated that YOLO-based models outperform two-stage detectors in frames-per-second throughput while maintaining competitive mean average precision. Nevertheless, existing works often evaluate under limited environmental diversity, restricting generalizability. Our work addresses this gap by systematically evaluating performance across four environmental scenarios and employing five-fold cross-validation to produce statistically reliable estimates.

III. PROPOSED METHODOLOGY

The proposed framework consists of three integrated stages: (1) data acquisition and preprocessing, (2) model development, and (3) deployment with real-time

reporting. Each stage is designed to be modular, enabling independent upgrades to individual components without disrupting the overall pipeline.

A. Data Collection and Preprocessing

Road surface data were collected using vehicle-mounted RGB cameras operating at 30 frames per second alongside three-axis accelerometers. Data collection campaigns spanned four distinct environments: urban arterials with dense traffic and mixed surface quality; rural roads with variable lighting and road shoulders; highway on-ramps with high-speed conditions; and active construction zones with heterogeneous surface materials and debris. A total of 12,000 image frames were extracted and annotated by domain experts into three classes: pothole, bump, and normal surface.

Preprocessing included histogram equalization for exposure normalization, Gaussian smoothing to reduce sensor noise, and image resizing to 224×224 pixels for CNN input and 640×640 pixels for YOLOv8. Data augmentation strategies — random horizontal flipping, rotation ($\pm 15^\circ$), brightness jitter, and Gaussian noise injection — were applied to artificially expand the training set and improve model generalization across diverse road conditions.

B. CNN Classification Architecture

The CNN model was designed with five successive convolutional blocks. Each block comprised a 3×3 convolution layer with ReLU activation, batch normalization, and 2×2 max-pooling. Filter depths progressed from 32 to 256 across the five blocks to capture low-level texture features in early layers and high-level semantic patterns in deeper layers. Table III details the full architecture. A global average pooling layer followed by two fully connected layers with dropout regularization ($p=0.5$) formed the classification head, producing a softmax probability distribution over the three output classes.

Training was performed using the Adam optimizer with an initial learning rate of 1×10^{-3} , decayed by a factor of 0.1 every 15 epochs, over a total of 50 epochs. Categorical cross-entropy was used as the loss function. Five-fold cross-validation was employed to partition the 12,000-image dataset, with each fold used once as the validation set and the remaining four for training, ensuring unbiased performance estimation.

TABLE III: CNN Model Architecture Table Structure

Layer	Type	Output Shape	Filters / Units	Activation
Conv Block 1	Conv2D + BN + MaxPool	112×112×32	32	ReLU
Conv Block 2	Conv2D + BN + MaxPool	56×56×64	64	ReLU
Conv Block 3	Conv2D + BN + MaxPool	28×28×128	128	ReLU
Conv Block 4	Conv2D + BN + MaxPool	14×14×256	256	ReLU
Conv Block 5	Conv2D + BN + MaxPool	7×7×256	256	ReLU
GlobalAvgPool	Global Avg Pooling	256	—	—
FC-1	Fully Connected	512	512	ReLU
Dropout	Regularization	512	p=0.5	—
FC-2 (Output)	Fully Connected	3	3	Softmax

C. YOLOv8 Detection Model

YOLOv8 was selected for its anchor-free detection head and multi-scale feature pyramid network (FPN), which enable detection of defects at varying scales within a single forward pass. The YOLOv8-medium variant was initialized with COCO pre-trained weights and fine-tuned on the annotated road defect dataset for 100 epochs using stochastic gradient descent (SGD) with a momentum of 0.937 and a cosine learning rate schedule (initial lr=0.01, final lr=0.001). During inference, non-maximum suppression (NMS) was applied with an IoU threshold of 0.50 and a confidence threshold of 0.25 to filter redundant bounding box predictions.

D. System Deployment Pipeline

The deployment pipeline integrates three functional modules. The Data Acquisition Module streams video from vehicle-mounted cameras to an onboard NVIDIA Jetson AGX Orin processing unit. The Inference Engine runs the fine-tuned YOLOv8 model in real time at 47 FPS, annotating each frame with defect bounding boxes and confidence scores. The Reporting Module geo-tags detected defects using onboard GPS coordinates, aggregates detections by road segment, and transmits structured incident reports to a cloud-based maintenance dashboard accessible by road maintenance authorities.

IV RESULTS AND DISCUSSION

A. Experimental Setup

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), 64 GB DDR4 RAM, and Ubuntu 22.04 LTS. PyTorch 2.0 was used as the deep learning framework for both models. Model performance was assessed using accuracy, precision, recall, and F1-score for the CNN classifier, and mean average precision at IoU threshold 0.5 (mAP@0.5) alongside real-time frames-per-second (FPS) throughput for the YOLOv8 detector.

B. Quantitative Results

Table I summarizes the overall performance of both models across the complete dataset. YOLOv8 consistently outperformed the CNN across all classification metrics. The YOLOv8 model achieved an overall accuracy of 96.7%, surpassing the CNN by 4.3 percentage points. The precision and recall improvements of 4.1 pp and 3.8 pp respectively demonstrate YOLOv8's superior ability to minimize both false positive detections and missed defects. The mAP@0.5 of 94.8% confirms high-quality bounding-box predictions across all defect categories. At 47.2 FPS, YOLOv8 comfortably satisfies the real-time processing requirement for a 30 FPS video stream, providing a safety buffer for embedded deployment.

TABLE I: Overall Performance Comparison — CNN vs. YOLOv8

Metric	CNN (%)	YOLOv8 (%)	Improvement	Notes
Accuracy	92.4	96.7	+4.3 pp	Best on highways
Precision	91.8	95.9	+4.1 pp	Low FP rate
Recall	90.5	94.3	+3.8 pp	Low FN rate
F1-Score	91.1	95.1	+4.0 pp	Balanced measure
mAP@0.5	—	94.8	—	Object detection
FPS	—	47.2	—	Real-time capable

Table II presents per-scenario performance breakdowns. Both models performed best on highway on-ramp data, where surface uniformity and consistent daylight lighting conditions facilitated cleaner feature extraction. Performance was lowest in construction zones, where irregular debris, exposed aggregate, and heterogeneous pavement materials introduced classification ambiguity for both models.

The YOLOv8 model's 92.8% mAP in the construction zone scenario, while lower than other settings, still demonstrates practically useful detection capability.

TABLE II: Per-Scenario Performance Breakdown

Scenario	CNN Acc. (%)	YOLOv8 mAP (%)	Dominant Defect	Key Challenge
Urban Arterial	93.8	96.1	Potholes	Occlusion by vehicles
Rural Road	92.1	95.4	Potholes	Variable lighting
Highway On-ramp	94.2	97.3	Bumps	High vehicle speed
Construction Zone	89.2	92.8	Mixed	Heterogeneous surface

C. Qualitative Analysis

Qualitative inspection of YOLOv8 detection outputs revealed that the model reliably localizes potholes with tight bounding boxes even under partial occlusion by shadows or road debris. False positive detections were most frequent at reflective wet patches on rural roads, which share low-level textural features with shallow potholes. CNN misclassifications were concentrated in ambiguous transition zones where pothole edges merge with normal road cracking patterns, indicating a limitation of patch-level classification without spatial context.

The CNN's lower computational overhead (9.1 ms per image on CPU vs. 21 ms for YOLOv8) makes it suitable for offline batch analysis of archived road survey footage, while YOLOv8's higher accuracy and bounding-box output are better suited for real-time in-vehicle deployment. A hybrid pipeline that uses CNN for initial frame-level screening and YOLOv8 for flagged frames is a promising direction for resource-constrained deployments on low-power edge hardware.

V SYSTEM MODULE DESCRIPTION

A. Smart Road Data Collection Module

This module manages synchronized capture of visual and inertial data from vehicle-mounted sensors. High-definition cameras stream video at 30 FPS while tri-axial accelerometers record vibration signatures correlated with road surface events. Timestamped GPS coordinates are logged concurrently to enable geo-spatial mapping of detected defects. The module implements a ring-buffer management strategy to handle continuous data streams

with minimal latency and supports adaptive frame rate control based on vehicle speed.

B. Deep Learning Inference Module

The inference module hosts the fine-tuned YOLOv8 model and performs per-frame detection at 47.2 FPS on the embedded NVIDIA Jetson AGX Orin. TensorRT optimization was applied to the ONNX-exported model, reducing inference latency by 38% compared to the native PyTorch baseline. Detected defects are annotated with class labels, confidence scores, and normalized bounding box coordinates before being forwarded to the reporting module for geo-tagging and aggregation.

C. Reporting and Alert Module

Detected defects are geo-tagged with GPS coordinates and aggregated into structured incident records containing defect type, severity score, location, and a cropped evidence image. Records are transmitted over 4G/LTE to a cloud-based maintenance dashboard where road authorities can filter incidents by severity threshold, location cluster, and road segment. Automated SMS and email alerts are dispatched when high-severity potholes are detected, enabling rapid mobilization of repair teams and reducing the time between detection and maintenance action.

D. Model Monitoring and Update Module

To maintain detection accuracy over time as road conditions evolve, the system includes a model monitoring module that tracks prediction confidence distributions across data streams. When confidence degradation is detected — indicating distribution shift — the module flags the affected region for human review and queues labeled samples for incremental model re-training. This continuous learning loop ensures the deployed model remains accurate across changing seasons, new road surfaces, and regional pavement standards.

VI CONCLUSION AND FUTURE SCOPE

This paper presented an automated road defect detection framework integrating CNN classification and YOLOv8 real-time object detection for identifying potholes and bumps across diverse road environments. Validated on a 12,000-image multi-scenario dataset using five-fold cross-validation, the YOLOv8 model achieved 96.7% accuracy and 94.8% mAP@0.5 at 47.2 FPS,

demonstrating strong suitability for real-time vehicle-

mounted deployment. The end-to-end pipeline — spanning sensor data acquisition, GPU inference, geo-tagging, and cloud-based maintenance reporting — provides a practical, scalable solution for modern road asset management.

The integration of geo-tagging, severity scoring, and automated alerting enables road authorities to transition from reactive to proactive maintenance strategies, reducing the time between defect occurrence and repair. This directly translates into improved driving safety, reduced vehicle operating costs, and lower long-term infrastructure expenditure. The cross-scenario validation confirms that the framework generalizes well beyond controlled laboratory environments, making it suitable for real-world deployment across varied geographic and climatic conditions.

Future work will explore several directions. First, domain-adaptive training strategies will be investigated to improve construction-zone performance, potentially leveraging synthetic data augmentation via generative adversarial networks (GANs). Second, lightweight model distillation techniques will be developed for deployment on cost-constrained microcontroller-based edge nodes. Third, the detection scope will be extended to include additional defect categories such as surface cracks, rutting, alligator cracking, and lane-marking degradation. Finally, integration with smart city infrastructure and Vehicle-to-Everything (V2X) communication protocols will be explored to enable cooperative, fleet-wide road condition mapping that continuously updates a shared road health database.

ACKNOWLEDGMENT

The authors sincerely thank the Department of Transportation and the institution's research laboratory for providing access to road testing zones and high-performance computational resources. Gratitude is also extended to the field team members who assisted in the multi-scenario data collection campaigns and the domain experts who contributed to the manual annotation process.

REFERENCES

- [1] S. Nienaber, M. J. Booysen, and R. S. Kroon, "Detecting potholes using simple image processing techniques and real-world footage," in Proc. South African Transport Conf. (SATC), Pretoria, South Africa, Jul. 2015, pp. 1–10.
- [2] A. Eriksson, M. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: Using a mobile sensor network for road surface monitoring," in Proc. 6th ACM Int. Conf. Mobile Syst. Appl. Serv. (MobiSys), Breckenridge, CO, USA, Jun. 2008, pp. 29–39.
- [3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, Apr. 2020.
- [4] J. Zou, H. Ye, and C. Ren, "Road surface defect detection using image processing techniques," IEEE Trans. Intell. Transp. Syst., vol. 18, no. 3, pp. 754–762, Mar. 2017.
- [5] Y. Li, H. Liu, X. Hu, L. Zhao, and H. Wan, "Pavement defect detection with fully convolutional network and uncertainty estimation," ISPRS Int. J. Geo-Inf., vol. 9, no. 1, p. 48, Jan. 2020.
- [6] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiwayama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," Comput.-Aided Civ. Infrastruct. Eng., vol. 33, no. 12, pp. 1127–1141, Dec. 2018.
- [7] W. Zhang, Z. Zhang, and X. Qi, "Automatic pavement defect detection using 3D laser profiling technology," Autom. Constr., vol. 96, pp. 350–365, Dec. 2018.
- [8] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, and I. Pitas, "Pothole detection based on disparity transformation and road surface modeling," IEEE Trans. Image Process., vol. 29, pp. 897–908, 2020.
- [9] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "Deep learning-based road damage detection and classification for multiple countries," Autom. Constr., vol. 132, p. 103935, Dec. 2021.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
- [11] G. Jocher et al., "Ultralytics YOLOv8," GitHub, 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [12] P. Rathore, A. Kumar, J. C. Bandyopadhyay, P. Bose, and J. Dutta, "Road surface monitoring using smartphone sensors: A survey," Sensors, vol. 17, no. 6, p. 1270, Jun. 2017.