

Automated Road Safety Evaluation Utilizing Convolutional Neural Networks and Edge Computing a Comprehensive Study

Mohit Nakade¹, Aniket Patil²

^{1,2}Student, CSE, YSPM's Yashoda Technical Campus, Satara, India

Abstract—The rapid degradation of municipal transport infrastructure demands automated structural evaluation frameworks. This research presents the design and implementation of an Artificial Intelligence Smart Road Monitoring System utilizing deep learning architectures for real time pavement anomaly detection. By replacing manual visual inspections, the proposed framework processes video streams captured via vehicle mounted cameras. A fine-tuned object detection network models, detects, and classifies structural surface degradation including potholes, alligator cracks, and longitudinal fractures. Deployed over a decoupled cloud native framework with a high-performance backend, the system stores geo tagged anomaly data into a relational database to generate instant alerts and prioritize maintenance. This end-to-end framework bridges the gap between raw data collection and actionable predictive maintenance strategies for smart city infrastructure management.

Index Terms—Hawkes Process, Order Book Imbalance, Flash Crash, Anomaly Detection, Limit Order Book, High-Frequency Trading, and Retrieval-Augmented Generation.

I. INTRODUCTION

Roadway networks constitute the foundational backbone of transport logistics. However, the upkeep of these expansive asset systems remains a persistent operational challenge for highway authorities. Pavement surfaces continually decay under the compounding stress of heavy axle loads, environmental thermal expansion, and moisture infiltration. Traditional pavement management strategies rely on manual ocular surveys or specialized laser profilometer vehicles.

Manual surveying is fundamentally flawed. It is logistically slow, labor intensive, exposes field

engineers to hazardous traffic environments, and suffers from extremely subjective observer bias. Specialized hardware vehicles, while precise, are prohibitively expensive to operate at scale.

To resolve these critical structural bottlenecks, this project introduces a scalable, low latency Artificial Intelligence Smart Road Monitoring System built upon deep learning and edge computing principles. By utilizing high resolution cameras mounted on vehicles, the system actively records surface data during standard transit. These visual streams are processed frame by frame via specialized Convolutional Neural Networks optimized for localization and classification tasks. Surface anomalies are assigned bounding boxes, categorized by distress severity, and appended with spatial telemetry data. This transition from manual inspections to continuous automated edge intelligence transforms pavement maintenance from an inefficient reactive model into a precise predictive infrastructure asset management system.

II. LITERATURE SURVEY

Automated pavement distress evaluation has transitioned from brittle pixel level feature engineering to highly robust deep neural network topologies.

Cord and Chambon (2012) implemented traditional mathematical morphology, two-dimensional spatial filtering, and localized luminance thresholding techniques to isolate surface crack networks from structured three-dimensional laser profiles. The algorithms demonstrated severe performance drops when exposed to non-uniform ambient illumination, shadows from roadside objects, surface moisture variations, and superficial asphalt discoloration.

Zhang et al. (2016) deployed an early stage supervised Deep Convolutional Neural Network trained on structured patch-based grid segments, shifting the domain away from manual feature extraction to learned internal representations. The patch wise classification approach lacked localization regression capabilities. It could only signify if a grid patch contained a crack, rendering it computationally inefficient and unsuitable for high-speed vehicular video processing.

Maeda et al. (2018) developed a multi category road damage dataset and trained a lightweight Single Shot MultiBox Detector MobileNet framework, proving the viability of executing real time damage inference on resource constrained smartphone devices. The lightweight architecture sacrificed accuracy, yielding high false positive rates when encountering complex visual artifacts such as road debris, skid marks, or varying asphalt patch textures.

Majidifard et al. (2020) introduced a two-tier hybrid architecture combining an object detector for bounding box isolation with a downstream semantic segmentation network to calculate the precise surface area density of the distress. The sequential pipelining of two independent deep learning networks introduced massive computational latency, making it impossible to run concurrently on edge devices without dedicated desktop grade Graphical Processing Units.

Arya et al. (2021) formulated generalized transfer learning frameworks utilizing heterogeneous multinational datasets to evaluate cross regional model performance against distinct asphalt compositions and lane markings. Despite localized accuracy gains, the model generalized poorly when exposed to out of distribution environmental factors, such as extreme weather anomalies, wet pavement specular reflections, or heavy snow accumulation.

III. METHODOLOGY

The methodology establishes a robust mathematical and computational framework for identifying and categorizing pavement distress using advanced computer vision techniques.

3.1. Dataset Acquisition and Preprocessing

The model training relies on a composite dataset aggregating diverse visual conditions from municipal roadways. The primary data source comprises high resolution images captured via vehicular mounted industrial cameras. To prevent mathematical overfitting during the training phase, the raw dataset undergoes rigorous spatial and chromatic augmentation. Techniques include random scaling, horizontal flipping, and stochastic color jittering. Histogram equalization standardizes the pixel intensity distribution, mitigating the effects of variable ambient illumination and ensuring the neural network receives standardized tensor matrices.

3.2. Deep Learning Architecture

This research employs the eighth iteration of the You Only Look Once object detection algorithm. The network topology consists of a feature extraction backbone, a fusion neck, and a decoupled prediction head.

The backbone utilizes the Cross Stage Partial bottleneck with two convolutions module, which enhances gradient flow and reduces computational bottlenecks compared to older network designs. A Spatial Pyramid Pooling Fast layer concludes the backbone architecture, extracting structural features at multiple receptive fields through sequential mathematical max pooling operations.

The neck aggregates these multi scale features using a Path Aggregation Network structure. Finally, the decoupled head processes categorical classification and bounding box regression through parallel convolutional branches, dramatically improving spatial localization precision.

3.3. Loss Function Formulation

The optimization process minimizes a composite loss function. For bounding box regression, the algorithm employs Complete Intersection over Union loss. This metric evaluates the overlap area, the distance between center points, and the aspect ratio consistency between the predicted and ground truth bounding boxes.

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha \nu$$

Where b and b^{gt} represent the central spatial coordinates of the predicted and ground truth boxes, ρ is the Euclidean distance, c is the diagonal

length of the smallest enclosing box covering both objects, ν measures aspect ratio consistency, and α is a positive trade off parameter.

For classification, the model applies Focal Loss to address severe class imbalance between prevalent minor surface cracks and rare critical potholes.

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t)$$

Where p_t is the estimated probability for the ground truth class, α_t is a weighting factor, and γ is the focusing parameter that down weights easily classified background examples.

IV. CONTENT

4.1. Edge Computing Hardware Environment

The physical deployment utilizes compact computing nodes integrated directly into municipal transit vehicles. These nodes feature an eight core Central Processing Unit paired with a dedicated discrete Graphics Processing Unit equipped with 4096 Compute Unified Device Architecture cores. This hardware matrix ensures sufficient teraflops to sustain continuous matrix multiplications without frame dropping.

4.2. Software Stack and Inference Pipeline

The inference logic operates entirely within a Python environment. Open-Source Computer Vision libraries handle asynchronous video decoding. The incoming continuous video stream is partitioned into discrete frames and tensorized. The neural model evaluates each tensor, outputting confidence scores and absolute spatial coordinates. A non-maximum suppression algorithm filters overlapping predictions, retaining only the bounding box with the highest probability threshold.

4.3. Telemetry and Backend Synchronization

To maintain minimal network latency over cellular bands, the edge nodes do not transmit heavy raw video files. Instead, they extract a localized image crop of the identified anomaly and append it to a lightweight JavaScript Object Notation payload. This payload includes the distress classification, the associated confidence percentage, and the Global Positioning System coordinates. An Asynchronous Server Gateway Interface routes this data payload to a

centralized PostgreSQL relational database enhanced with spatial extensions for geographic querying.

V. RESULT AND DISCUSSION

5.1. Evaluation Metrics

The model performance is strictly quantified using Precision, Recall, and mean Average Precision.

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

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

Where TP represents True Positives, FP represents False Positives, and FN represents False Negatives. The mean Average Precision evaluates the integral area under the Precision Recall curve across all distress categories at a strict Intersection over Union threshold of 0.50.

5.2. Quantitative Performance

Empirical testing on an independent validation subset of 4500 images yields robust results. The overall mean Average Precision reaches 84.7 percent. Pothole detection achieves the highest categorical accuracy at 91.2 percent due to their distinct geometric boundaries and structural depth shadows. Longitudinal and transverse cracks achieve a mean Average Precision of 81.5 percent. The edge inference engine maintains an average execution speed of 34 frames per second on the physical hardware, confirming its viability for high-speed vehicular deployment without data loss.

5.3. Qualitative Failure Analysis

Despite high quantitative metrics, specific failure modes persist in the wild. The model exhibits reduced precision when evaluating roads with heavy surface moisture, as water reflections distort the crack topology and confuse the convolutional filters. Additionally, complex shadows cast by overhanging foliage occasionally trigger false positive detections for alligator cracking. These edge cases highlight the inherent limitations of relying solely on two-dimensional visual spectrum data.

VI. FUTURE SCOPE

Future Scope

Future iterations of this research will focus heavily on mitigating the current limitations associated with variable ambient lighting and standing surface moisture. The integration of multi modal sensor fusion

is a primary objective. Coupling visual spectrum cameras with Light Detection and Ranging sensors will provide authentic three-dimensional depth profiles, allowing the algorithm to differentiate definitively between superficial liquid stains and genuine structural deformations. Furthermore, exploring emerging Vision Transformer architectures may improve the contextual global understanding of extensive pavement networks, potentially exceeding the localized receptive field constraints of current convolutional models.

REFERENCES

- [1] Cord and S. Chambon, “Automatic pavement surface crack detection on 3D laser images,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 741–753, Jun. 2012.
- [2] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, “Road crack detection using deep convolutional neural networks,” in Proc. IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 3708–3712.
- [3] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, “Road damage detection and classification using deep neural networks with smartphones,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7663–7670, Dec. 2021.
- [4] H. Majidifard, P. Jin, Y. Adu-Gyamfi, and W. G. Buttlar, “Pavement distress classification using deep learning models and image segmentation,” *International Journal of Pavement Engineering*, vol. 23, no. 8, pp. 2455–2469, Jul. 2022.
- [5] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, “Deep learning-based road damage detection and classification for multiple countries,” *Automation in Construction*, vol. 132, p. 103935, Dec. 2021.