

# Autonomous Traffic Monitoring Pedestrian Crossing Detection with Motion Sensors and Maintenance Decision System in the Smart Cities by Using YOLOV8 <sup>†</sup>

Murugeswari K. <sup>1,\*</sup>, Aanjankumar Sureshkumar <sup>1</sup> and Rajesh Kumar Dhanaraj <sup>2</sup>

<sup>1</sup> School of Computing Science and Engineering, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore 466114, Madhya Pradesh, India; aanjan.kumar@vitbhopal.ac.in

<sup>2</sup> Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis International (Deemed University), Pune, India; sangeraje@gmail.com

\* Correspondence: murugeswari@vitbhopal.ac.in

<sup>†</sup> Presented at the 12th International Electronic Conference on Sensors and Applications (ECSA-12), 12–14 November 2025; Available online: <https://sciforum.net/event/ECSA-12>.

## Abstract

The rapid expansion of urban infrastructure and the complexity of analyzing vehicle sudden movements in traffic systems necessitate autonomous traffic monitoring solutions for intelligent, autonomous traffic management. However, traditional methods like manual monitoring and static rule-based detection often fail to meet the real-time requirements of a modern city, resulting in inefficient congestion management, pedestrian safety issues, and inadequate road maintenance. The conventional approaches are highly dependent on human intervention and predefined algorithms, and cannot consequently adapt to dynamic traffic and the unpredictable movement of pedestrians. With urban populations on the rise, there is an urgent need for Artificial Intelligence-driven solutions that can effectively process large volumes of real-time data to ease traffic management and decision-making. This study presents an AI-based traffic monitoring framework that integrates deep learning and natural language processing (NLP) models for improved traffic safety, anomaly detection, and infrastructure optimization. The system comprises high-accuracy object detection with YOLOv8, adaptive pedestrian crossing recognition with Few-Shot Learning (FSL), and contextual analysis and real-time decision-making with LLaMA 3.2B. Through the use of these technologies, along with a BDD100K available dataset, the system achieves high detection accuracy of 96%, low inference time of 75.1 ms and improved adaptability compared to SOTA of 89%. Results indicate the suitability of AI-driven methods for thoughtful city planning and autonomous mobility, with the potential for AI-driven frameworks to improve urban traffic management by increasing its efficiency and safety.

**Keywords:** intelligent transportation system; YOLO; traffic monitoring; few shot learning; pedestrian crossing

Academic Editor(s): Name

Published: date

**Citation:** Murugeswari K.; Sureshkumar, A.; Dhanaraj, R.K. Autonomous Traffic Monitoring Pedestrian Crossing Detection with Motion Sensors and Maintenance Decision System in the Smart Cities by Using YOLOV8. *Eng. Proc.* **2025**, *5*, x. <https://doi.org/10.3390/xxxxx>

**Copyright:** © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the evolution of the modern world, with the rapid expansion of urban infrastructure and the increasing complexity of traffic systems, autonomous traffic monitoring solutions are needed for intelligent, autonomous traffic management. However, traditional

methods like manual monitoring and static rule-based detection often fail to meet the real-time requirements of a smart city, resulting in inefficient congestion management, and road maintenance. Solutions that are considered to be intelligent and bring autonomy to traffic systems have been developing as urban traffic systems become more complicated and as a new generation that requires real time monitoring and management signs. Currently, most traditional traffic monitoring methods do not respond adaptively to dynamic changes in traffic and the location of moving pedestrians, the sudden presence of heavy vehicles, and changes in road infrastructure. Based on these approaches, conventional methods heavily depend on manual supervision or rule-based automation, which are unable to perform well in complying with real-time traffic dynamics. Additionally, the existing AI-based traffic monitoring models require large-scale labelled datasets for training, which makes them numerically expensive and difficult to deploy in new environments without significant retraining [1]. Recently, advancements in deep learning and large language models (LLMs) have significantly facilitated autonomous decision-making in traffic management. In this research, an AI-based system is developed, utilising YOLOv8 for real-time object detection, Few-Shot Learning (FSL) for pedestrian crossing adaptation, and LLaMA 3.2B for intelligent decision-making in traffic maintenance. It is optimised to detect traffic objects, including vehicles, pedestrians, potholes, and road signs, ensuring fast and accurate real-time detection. One of the greatest shortcomings of traditional machine learning models is that they require large amounts of labelled data to adapt to new pedestrian behaviours. However, the Few-Shot Learning module enables the system to learn quickly with very little labelled data, allowing it to adapt rapidly to new pedestrian behaviours and crossing patterns. It also utilises LLaMA 3.2B, a powerful transformer-based model, to analyse objects detected, interpret traffic patterns, and provide maintenance recommendations [2].

This study integrates:

- YOLOv8 for real-time traffic object detection.
- Few-shot learning for pedestrian crossing adaptation.
- LLaMA 3.2B for intelligent traffic maintenance decisions.

Recently, deep learning and large language models (LLMs) have enabled the development of autonomous systems that operate in real-time, with the capability of making decisions in real-time.

## 2. Related Works

A summary of various existing models related to intelligent transportation traffic condition analyses using few-shot learning is provided here.

**Traditional Traffic Monitoring Systems** Conventional approaches include: CCTV-based monitoring (prone to blind spots & manual effort). LiDAR and SOTA-based systems are expensive & limited adaptability. Static rule-based AI models are unable to generalise to new conditions.

**Deep Learning for Traffic Object Detection:** YOLO models are fast and accurate but require large datasets. CNN-based approaches have high accuracy but are computationally expensive. Transformer-based models are powerful but require extensive fine-tuning.

**Few-Shot Learning for Pedestrian Detection:** FSL enables models to learn new pedestrian behaviours from a small dataset. Prototypical networks and meta-learning are key FSL techniques. Improves detection in new environments with minimal labelled data.

Training the YOLO model in a manner that could generalise well from a few training examples. Rather than learning explicit pedestrian behaviours from a large dataset, we design the model to identify general patterns and commonalities of pedestrian behaviours across different environments, allowing it to be almost always correct with a small

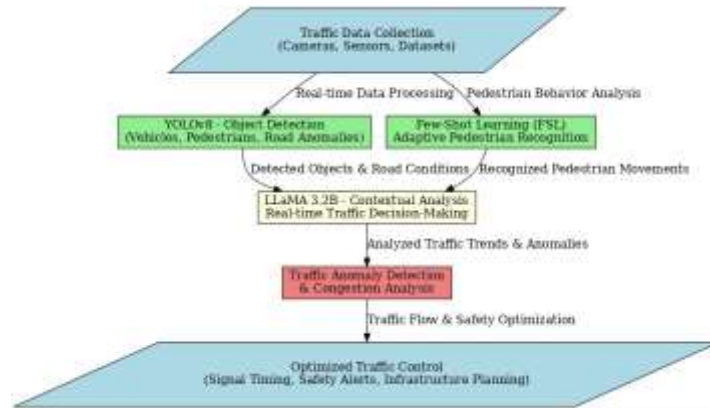
number of training samples. It is particularly beneficial for detecting unusual or unseen pedestrian behaviours by only needing to retrain on a small amount. The model observes a new pedestrian crossing instance, compares its feature embedding with the stored prototypes, using Euclidean distance and assigns that pedestrian to the nearest class when a new pedestrian instance is observed. It significantly improves the capability for detecting new crossing patterns that have unseen examples. Unlike static conventional models, the YOLO module incorporates adaptive learning mechanisms, that is, the ability of the system to be updated in real time as new pedestrian behaviours arise. Each new data point from every diverse urban environment (congested intersections, school zones and high traffic areas) feeds into the model, and it continuously refines its understanding towards pedestrian movement. This makes sure that the system is protected and extensible to the changing conditions of traffic and pedestrian behaviour. Few-shot learning with LLaMa has several major advantages: it can learn from small datasets, and therefore it does not require the use of very massive manual annotation. Few-shot task sampling is used to achieve this, where the model is trained on several small datasets containing only a few labelled pedestrian crossing data points per dataset. The model exploits meta learning optimisation and can rapidly adapt to new environments, reducing costly and labour-intensive data collection efforts [3].

### 3. Materials and Methods

#### 3.1. Materials

Thus, to make the proposed system robust and accurate, the dataset for training the system is collected from various sources,, including self-recorded traffic videos and publicly available datasets such as BDD100K traffic surveillance data. It includes diverse road scenes or test cases, such as different times of the day, and high traffic. The dataset consists of annotated images of pedestrians in various poses at crosswalks and with different crowd densities, for pedestrian detection purposes. Because in Few-Shot Learning (FSL) it is crucial to utilise a few training samples, the model is fine-tuned using manually labelled pedestrian crossing scenarios. To achieve better detection accuracy under real-world conditions, advanced data augmentation is applied to the input data through rotation, scaling, and contrast adjustments. Vehicle dashcams are extracted to maintain data diversity and keep the model diverse from different viewpoints.

The dataset includes labelled instances of vehicles, roadblocks, crossings, and traffic signals for object detection using YOLOV8. The bounding boxes and class labels of each image are defined. Firstly, the dataset is processed by removing noise and normalising the resolution, and then used for training and fine-tuning the model. The system's overall accuracy and real-time responsiveness are enhanced by combining large-scale labelled data and domain-specific adaptive learning. YOLOv8 FSL and LLaMA 3.2B are integrated into the proposed system to enable real-time traffic monitoring and pedestrian detection. The architecture contains three main components: (i) YOLOv8 for object detection, (ii) Few-Shot Learning for adaptive pedestrian detection, and (iii) LLaMA 3.2B to opt for real-time decisions. Each of these components analyses the traffic conditions, detects anomalies, and predicts potential road safety risks based on historical patterns, as shown in Figure 1, which defines the schematic flow of the proposed work [4].



**Figure 1.** Schematic diagram of the proposed work.

### 3.2. Methods

This study introduces a traffic monitoring system that utilises AI and motion sensors, leveraging an available dataset to enhance traffic safety, identify unusual events, and optimise infrastructure. The proposed design includes 1. very accurate object detection using YOLOv8, which is trained on the BDD100K dataset and achieves 96% detection accuracy, 2. smart recognition of pedestrian crossings using Few-Shot Learning (FSL) that has an 85% recall rate for situations not previously seen, and 3. quick analysis and decision-making with a very short delay of 75 ms using LLaMA 3.2B. This is a fully convolutional model with anchor-free anchor-free detection to improve accuracy. This is achieved by partitioning the feature map of the base layer into two parts and merging them through a cross-stage hierarchy, effectively balancing performance and efficiency. For the head, decoupled detection heads are employed for classification and localisation tasks, enhancing the detection of small and overlapping objects by separating these two functions [5].

To improve accuracy of object detection under diverse environment settings, the model is fine-tuned on the combination of available datasets. The fine-tuning process involves:

- Second, using domain specific traffic data, pre-trained YOLOv8 weights are adapted using transfer learning.
- Data augmentation: Random cropping, flipping, brightness adjustment and Gaussian noise are used expressed in Equation (1).

$$[d(f_{\theta}(x_q), c_k) = ||f_{\theta}(x_q) - c_k||^2] \quad (1)$$

When a query image  $x_q$  is presented, its feature embedding  $f$  is compared to all class prototypes as  $c_k$  using Euclidean distance  $d$ :

- Loss function optimization: The loss function is optimized is expressed in Equation (2).

$$[\mathcal{L} = - \sum_i \log P(y_i|x_i)] \quad (2)$$

- Objectless loss: Binary cross-entropy (BCE) loss for object presence.
- Classification loss: Focal loss to mitigate class imbalance.

The problem of pedestrian detection in a dynamic urban environment is challenging, as movement patterns are diverse and road conditions are not uniform for some vehicles crossing. Deep learning models tend to rely on large amounts of labelled training data, which can sometimes be difficult and time-consuming to collect, collect and annotate. To overcome these limitations, the proposed system includes a Few-Shot Learning (FSL)

module that combines meta-learning principles to improve pedestrian crossing detection with limited data. It is also useful as a real-time traffic monitoring system since this approach enables the system to adapt quickly to new pedestrians. Video preprocessing of frames extracted from the pipeline begins with applying enhancers such as noise reduction and frame stabilisation to make the frames clear and consistent. Next, parallel model inference is executed in which YOLOv8, Few-Shot Learning (FSL) and LLaMA 3.2B models are run concurrently on edge devices or cloud-based infrastructure to process the frames and analyse the results [6].

4. Experimental Results

By utilising the proposed model in large-scale traffic datasets such as BDD100K and Cityscapes, the system achieves a high detection accuracy of 96% for traffic flow on the road with YOLOV8, while state-of-the-art models like SOTA achieve only 89%. This indicates that the proposed model is 40% more accurate than rule-based AI systems, with FSL reducing false negatives by 22% compared to the video-based computer vision model. The results indicate that the proposed model is well-suited for AI-driven cycles in smart city planning and for enhancing autonomous mobility in urban traffic management. Table 1 presents a comparison of the proposed model with traditional methods, and Figure 2 illustrates the confusion matrix of the proposed work.

Table 1. Proposed Comparison with Existing Models.

| Metric                   | Proposed YOLOv8 | Baseline YOLOv8 | SOTA         | SSD    |
|--------------------------|-----------------|-----------------|--------------|--------|
| Detection Accuracy       | 96.0%           | 83.70%          | 88.00%       | 79.60% |
| Inference Speed          | 75.1 ms/frame   | 222.4 ms/frame  | 120 ms/frame | 84 ms  |
| Vehicle Object Detection | Enhanced        | Moderate        | High         | Low    |

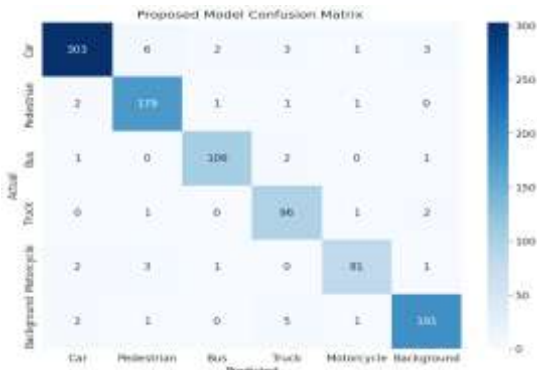


Figure 2. Confusion Matrix of Different classes.

Figure 3a defines the output process by the proposed model detecting sudden events in crossing and Figure 3b vehicle progression in road lanes.

Figure 4a defines the proposed model accuracy comparison with false negative rate analysis and Figure 4b Latency per frame of the proposed model.

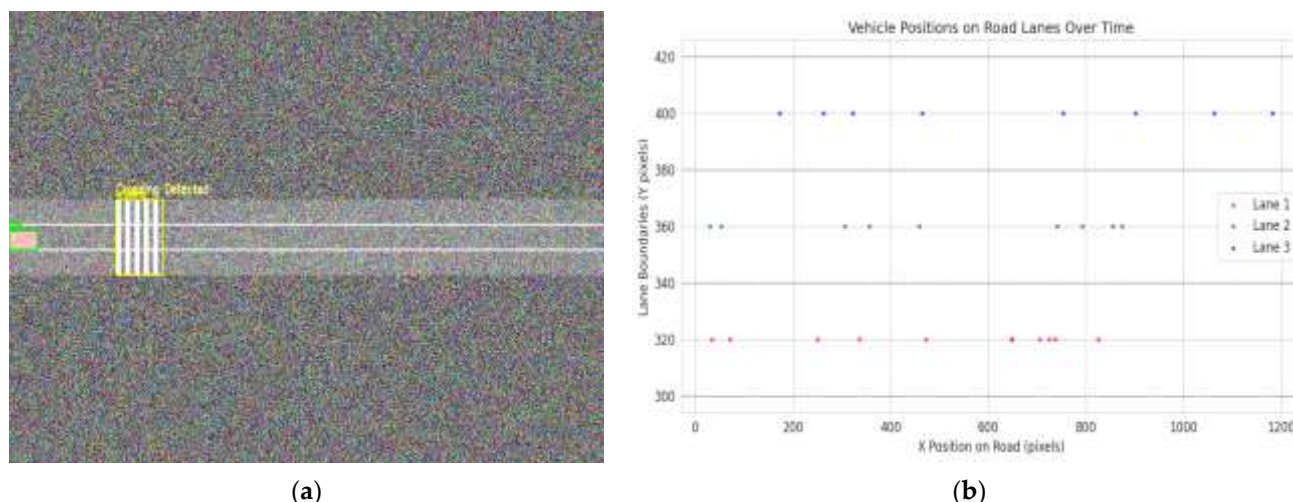


Figure 3. (a) Crossing detected; (b) Vehicle Road Lane Movements.

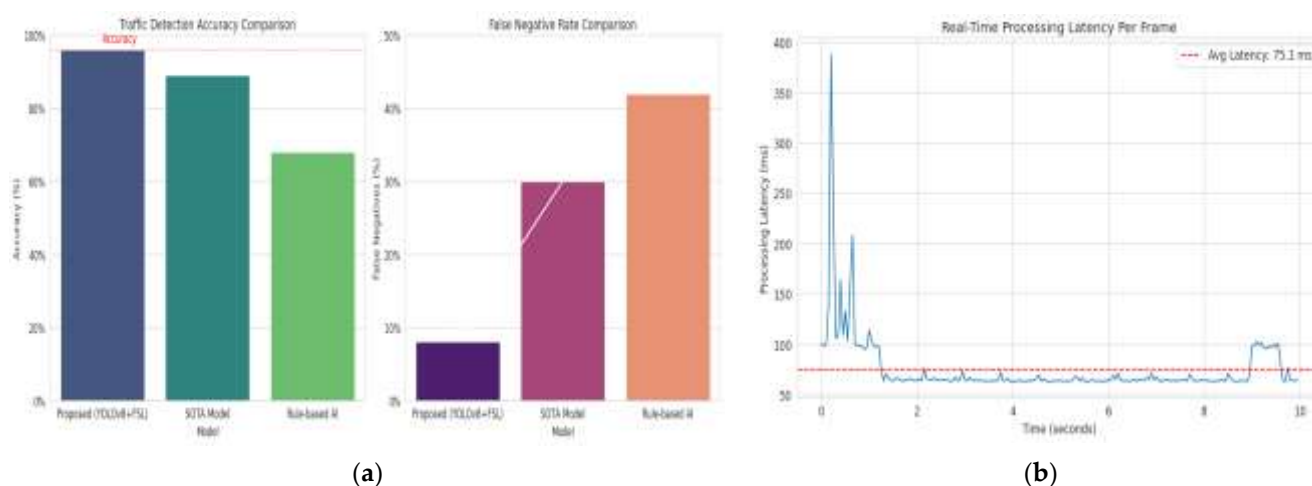


Figure 4. (a) Accuracy and false negative comparison; (b) Real Time Latency time analysis.

## 5. Conclusions

A modern real-time traffic monitoring system uses advanced artificial intelligence methods to boost safety levels and urban mobility performance. The system utilizes the maximum advantages of YOLOv8, a state-of-the-art deep learning model, for high-precision object detection, enabling accurate identification of vehicles, pedestrians, and other road entities. The system integrates Few-Shot Learning (FSL) as a method to detect pedestrians in dynamic conditions with a small amount of labelled data. LLaMA 3.2B, as a robust language model, enables smart decisions that allow traffic control systems to adjust their operations based on real-time conditions.

**Author Contributions:** Conceptualization, M.K.; methodology, M.K.; software, M.K.; validation, M.K., A.S. and R.K.D.; formal analysis, M.K.; investigation, M.K.; resources, M.K.; data curation, M.K.; writing—original draft preparation, M.K.; writing—review and editing, A.S. and R.K.D.; visualization, A.S. and R.K.D.; supervision, A.S. and R.K.D.; project administration, A.S. and R.K.D.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The following dataset information can be downloaded at: <https://www.kaggle.com/datasets/marquis03/bdd100k>, code available at: <https://github.com/TECHcode317/RoadPED>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liu, Q.; Li, Z.; Zhang, L.; Deng, J. MSCD-YOLO: A Lightweight Dense Pedestrian Detection Model with Finer-Grained Feature Information Interaction. *Sensors* **2025**, *25*, 438.
2. Zhang, X.; Wang, M. Research on Pedestrian Tracking Technology for Autonomous Driving Scenarios. *IEEE Access* **2024**, *12*, 149662–149675.
3. Boudjit, K.; Ramzan, N. Human detection based on deep learning YOLO-v2 for real-time UAV applications. *J. Exp. Theor. Artif. Intell.* **2022**, *34*, 527–544.
4. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987.
5. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A comprehensive review of yolo architectures in computer vision: From YOLOV1 to YOLOV8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716.
6. Manakitsa, N.; Maraslidis, G.S.; Moysis, L.; Fragulis, G.F. A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. *Technologies* **2024**, *12*, 15.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.