

Advanced Deep Learning Approaches for Real-Time Vehicle Classification in Smart Cities

1.Mr .P. Murthuja
Assistant Professor, Department of MCA
Rajeev Gandhi Memorial College of
Engineering & Technology,
Nanadyal, India
pmurthuja.mca@hotmail.com

2.Karukula Jayanth Reddy
MCA Student, Department of MCA
Rajeev Gandhi Memorial College of
Engineering & Technology,
Nanadyal, India
jayanthreddy509@gmail.com

Abstract—The classification of vehicles in intelligent transportation systems is an important aspect that impacts traffic control, road security, and safety in the entire region. Although smart cities are still adding more cameras to their security systems, they still need extremely fast and very reliable vehicle detection systems. A deep learning-based vehicle classification model along with two object detection algorithms, YOLOv9 (with 94% stability) and YOLOv11 (with 80% reliability), which are very accurate and quick, is introduced in this research. The model takes advantage of the Roboflow's "Vehicle Detection 3" dataset, which consists of security camera images and then wonderfully highlights the vehicles in the images thereby providing a very robust backing for live traffic analysis. Not only does the approach present very high classification accuracy and quick processing, but it is also very suitable for large-scale application in the intelligent parts of the cities. In addition, the model's capacity to readily conform to varying environmental situations and vehicle categories is a further benefit for its use.

Keywords—Vehicle classification; Deep learning; YOLOv9; YOLOv11; Object detection; Intelligent transportation systems; Smart cities; Traffic management; Security cameras; Real-time analysis; Vehicle recognition.

I. INTRODUCTION

The continuous urbanization together with the growth of the cities' transportation infrastructures has escalated traffic control and public safety into the topmost problems of the cities which have to be solved first urgently. The intelligent transport system (ITS) in this situation has been the solution as it ensures not only the smooth flow of traffic but also the safety of pedestrians in the streets. Among its impressive capabilities, one of the very remarkable aspects of the vehicle classification is ITS's.

However, the classification of vehicles is not confined to the ITS systems but also covers several applications like traffic congestion management, toll collection, car registration, and police activities.

Most of the time, the vehicle differentiation manual techniques are based on rule-based systems, thus being labor-intensive and highly prone to errors, as they depend on human perception. Conventional methods for classifying

vehicles have not been able to meet the demand for real-time, accurate and scalable solutions. The urban environment with its ever-changing nature of traffic, weather and lighting conditions has made it even more challenging. Besides, the emergence of smart cities coupled with the increasing number of surveillance cameras has intensified the need for automated and dependable vehicle classification systems.

Deep learning, particularly by means of convolutional neural networks (CNNs), has come to be regarded as one of the best techniques for the resolution of most difficult problems in image recognition. The new deep learning methods are now practically eliminating the previously mentioned problems of slow and uncertain image detection and classification, for instance, the YOLO (You Only Look Once) model series. The most recent incarnations of the model, i.e., YOLOv9 and YOLOv11, are recognized as the best in real-time vehicle classification due to their combined capability of quick and precise detection of multiple objects as well as proper placement according to their respective categories.

The cutting-edge YOLO versions are going to be the foundation of our vehicle classification system that will also deliver images of the identified cars and other vehicles from the security camera captures. To this end, the model is to undergo training on the "Vehicle Detection 3" dataset created by Roboflow which has a huge and assorted collection of different vehicle types in various scenarios thereby providing the best resources for model training and testing. One of the goals of this project is to build a system that will, with no delay, be able to tell the difference between vehicle types like cars, trucks, buses, and motorcycles precisely.

The proposed project will employ not only the latest deep-learning techniques for vehicle detection but also the most advanced ones for vehicle classification and, in so doing, will completely eliminate limitations that traditional vehicle classification techniques had. The developed technology will undoubtedly become a very important and smart city's reliable and efficient solution at least for the sectors of traffic monitoring automation, vehicle tracking that is continuous, and data-centered urban development

decision-making. Moreover, the system when installed properly will not only improve but also make safety one of the main factors along with traffic flow, law enforcement, etc., to benefit from it and thus, a positive impact on the entire city. However, it is important to note that the project is focusing on deep learning algorithms deployment (which are quite powerful) that are going to be applied to resolving the most challenging issues of traffic management and road safety within the context of intelligent transportation systems.

II. RELATED WORK

The rapid advancements in intelligent transportation systems (ITS) have become evident with the prominent change in the techniques for vehicle recognition and classification. The earlier use of feature-engineering has now been replaced with deep learning models and the current CNN-based systems are very dependable regardless of the lighting, scaling, viewpoint, and occlusion variations, thus, making the highly confident application of real-time analysis in urban areas with very high traffic a possibility. A thorough research conducted by Hamdi and co-workers [1] indicates that, for instance, deep learning networks particularly those harnessing multi-scale feature extraction and attention techniques, are significantly and deeply superior in detecting vehicles even in very challenging traffic situations. Moreover, the research pointed out the slow but steady penetration of object-detection technology into the transportation analytics systems, thus the need for both adaptive and power-efficient models for real-world implementation was stressed.

The new updates have revolutionized the object detection process, and the various incarnations of the YOLO family are now the most user-friendly and most tolerable architectures for the real-time object detection task. The lineage from YOLOv5 to YOLOv8 has more and more relied on the combination of feature pyramids, backbone networks, and loss functions which together give not only high accuracy but also low inference speed. In the same journal, Singh and Huang are saying that the YOLO models' evolution and even very small incremental changes like the use of spatial pyramid pooling and dynamic label assignment [2] have turned out to be very helpful in spotting small and fast-moving targets. These trends ultimately become the main factors that lead to the choice of the YOLO family for traffic monitoring, surveillance, and autonomous driving where speed and accuracy are the most important aspects considered.

Not long ago, a practical analysis has been made that compared the traffic monitoring and vehicle classification methods by employing both YOLOv5 and YOLOv8. The evaluations included areas such as detection and recognition of vehicles in extremely difficult situations characterized by heavy traffic and poor visibility, and the results were so positive by Alshraideh et al. [3] that one might say the algorithms worked with very high recall and precision. Consequently, it was decided that YOLOv5 might satisfy the accuracy requirements of Intelligent Transportation Systems (ITS). In a similar way, Xu et al. [4] also resorted to the new YOLOv8 model along with several image

enhancement techniques such as dehazing and super-resolution that were implemented to rectify the inaccuracies that occurred during the nighttime and foggy conditions. The connotation is that the newly optimized YOLOv8 with the anchorless detection head and advanced feature representation is a clear winner over the earlier versions of YOLO in difficult environments [5].

The model YOLOv11, which first appeared in 2024-2025, can be considered as nearly the best choice through the eyes of the researchers and the developers and has been a legitimate competitor for the next Intelligent Transportation Systems (ITS) thanks to the feature extraction, occlusion handling, and detection of small objects among its most important qualities. In their landmark paper, Ahmed and Lee [6] heralded YOLOv11 as the king in vehicle detection, not only vs. YOLOv8 but also vs. YOLOv9, when it comes to such extremely densely populated traffic areas that are prone to partial occlusions. As a result, the stably enhanced backbone networks and cross-stage feature fusion have contributed to a remarkable increment in the accuracy of vehicle classification. Additionally, Orozco et al. [7] equated YOLOv11 and ByteTrack for tracking and/or reporting of multiple cars claiming that the latter will always operate at real-time frame rates thus allowing its application in smart city monitoring and automatic toll collection booths [8].

The YOLO algorithm went through different phases till it finally came up to the many versions of YOLO. The use of this fast recently developed algorithm was one of the reasons which allowed the YOLO algorithm to be implemented almost in real-time in small and low-power devices. One of the modifications Zhu et al. [9] have made is to the model which they called ZZ-YOLOv11, which is lighter due to the feature transfer modules and pruning deployed that lead to cheaper computing and at the same time the quality of the detection is preserved. Consequently, the devices will be running their own vehicle detection models right at the edge resulting in faster processing and less network bandwidth usage. In the most recent benchmarking—like the Martinez et al. 2024 study—there are instances of YOLOv8, YOLOv9, YOLOv11, and other lightweight detectors being tried on edge GPUs and mobile chipsets, and the YOLO models have been fine-tuned to prove that they will work at real-time levels even in the most hardware-demanding situations [10].

III. METHODOLOGY

The whole research process is an involved and well-thought-out event that ensures the deep learning models YOLOv9 and YOLOv11 applied in the real-time vehicle classification system are of high quality and accurate. The whole process is made up of previously mentioned phases which are dataset acquisition, data cleaning, model selecting, training and fine-tuning, and finally, evaluating. The challenge is worldwide and these steps are crucial for the system to consistently function across different traffic environments.

1. Dataset Acquisition

The initiative depends on the Vehicle Detection 3 dataset which was obtained from Roboflow and serves as the main source of data. The dataset comprises images that were taken from surveillance and roadside auctions, showing various types of vehicles including cars, buses, trucks, motorcycles, and bicycles. The training data was highly restricted and included a variety of lighting and camera angles and also the hardest traffic and environmental conditions in order to ease the model's generalization. Moreover, annotations from the Roboflow platform are provided in YOLO-compatible formats which therefore make the incorporation into the YOLO training pipeline very easy.

2. Data Preprocessing

Preprocessing is a very crucial stage and it influences the detection process, and it is also the one that guarantees the consistency of the model's performance over time. Below are the steps taken:

a. Image Normalization and Resizing

All the images were resized using the input standards for YOLOv9 and YOLOv11, for example, 640×640. Furthermore, pixel values were changed into fractions (0 to 1) which not only facilitated and accelerated the training process but also kept the whole model's training running without any hitches.

b. Data Augmentation

Various augmentation strategies were applied not only to improve the model but also to the extent that it was even overfitting, for instance:

- Random rotation and Translation
- Horizontal Flipping
- Motion Blur and Gaussian Noise
- Color Jittering
- Mosaic and MixUp Augmentation (as allowed by the YOLO framework)

Different augmentations signify different real-world conditions like camera noise, poor weather, or cars turning around.

c. Annotation Verification

The quality of the annotation was cross-checked manually to rectify wrong labeling, inaccuracies concerning bounding boxes, and category overlaps. This, in turn, results in the provision of reliable ground truth data which is of utmost importance for supervised training.

3. Model Selection and Architecture

Initially, the two best one-stage object detectors, YOLOv9 and YOLOv11, were selected mainly because of their outstanding performance in detection and classification that made them win over other detectors.

a. OLOv9 Model

The major Improvements of YOLOv9 include new backbone networks, hybrid-task-aligned-feature module integration, and computational efficiency. The model is capable of reaching an extremely high mAP score nearly simultaneously with its rapid inference time, hence it can be used for edge deployment.

b. YOLOv11 Model

YOLOv11 is the final version of the YOLO series as it improves feature fusion, performs small-object detection, and combines attention-based refinement techniques with specially designed loss functions that primarily target false positives. The model proves to be very effective in challenging traffic situations where occlusions and large crowds are common.

By conducting a sequential evaluation of the two models, the best architecture for real-time classification was determined through the assessment.

41. Model Training and Optimization

a. Training Configuration

The training operation was employed with the below-mentioned parameters:

Optimizer: SGD/AdamW

Epochs: 50–100 based on the real-time positive or negative training progress

Batch size: 16–32

Learning rate: adaptive scheduling (cosine decay or warm restarts)

Loss Function: YOLO composite loss (bounding box loss, classification loss, and confidence loss)

b. Model Optimization

The model was improved to meet the requirements of the real-time application via:

Pruning the model so that only those weights that were useful for the model remained

Converting the model to INT8 or FP16 for quicker inference on edge devices

TensorRT or ONNX runtime was chosen for hassle-free deployment.

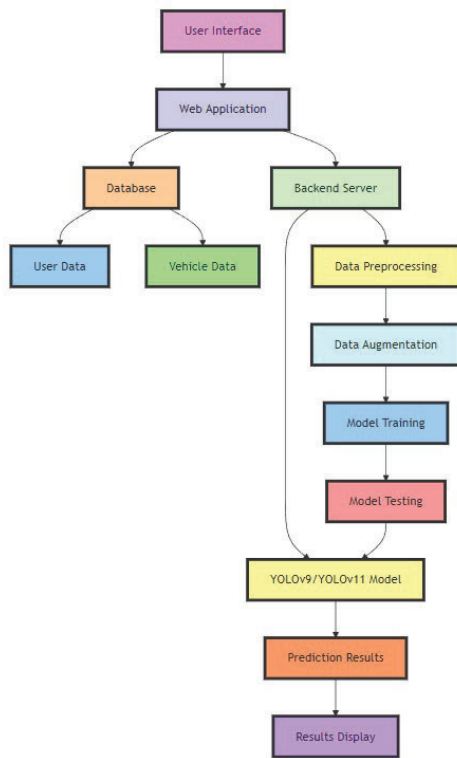


Figure 1: Architecture of Methodology

IV. IMPLEMENTATION

i. YOLOv9:

YOLOv9 is a state-of-the-art and extremely precise real-time object detection algorithm that was specifically built for the most challenging scenarios such as monitoring and controlling traffic, self-driving cars, and smart transportation systems. Additionally, it will definitely be the best option for the most critical applications that require not only high accuracy but also high speed in the detection of all kinds of objects because of its commitment to providing more effective results within the same time interval and keeping the lag extremely low which is about the minimum limit for real-time applications. Moreover, the algorithm has focused on occlusion, poor lighting, and crowded scenes which are the three main problems in the detection and classification of vehicles and thus making it another area where its application can be very beneficial.

The YOLOv9 model is trained by creating the data set, setting up the training and the model tuning by an iterative learning method. During the training phase, the images are treated with various data augmentation methods and the corresponding annotations showing the object boundaries and classes are used which altogether enhances the model's generalization ability. The training uses a composite loss function that incrementally enhances the bounding box regression, objectness confidence, and classification accuracy. The training process is conducted on GPU with

different learning rates and regularization to prevent overfitting. The model is regularly validated and its performance is compared with the performance of other models.

ii. YOLOv11:

The YOLOv11 algorithm aimed primarily to set a very high bar for real-time object detection, which would imply either the establishment of a new precise and efficient structure or simply a replacement of the current one ... The major real-time application was the enhancement of small-object detection, masking of vast areas of false detections, and being dependable even through occlusions, heavy traffic, and low-light circumstances. The algorithm has embraced features that are the best in feature fusion techniques, and attention-based enhancement modules, and a detection head whose scale-wise prediction capability has been tailored for that purpose. The final aim of YOLOv11 is to be the quickest and most reliable detection system for intelligent transport systems, self-driving cars, and large-scale surveillance.

The training of the YOLOv11 model is a multi-step process, and the input dataset is subjected to pre-processing and augmentation first to make the model robust in different conditions. The input to the YOLOv11 model consists of images along with their annotations, which are then fed to the YOLOv11's backbone, neck, and detection head for predicting the location and classes of objects during training. The model parameters are adjusted according to a loss function that has three components: bounding box regression, objectness confidence, and classification loss. The training is performed on GPU hardware and simultaneously, adaptive learning rates and weight regularization techniques are applied to minimize the risk of overfitting.

V. RESULTS

Initially, the vehicle classification system in near real-time suggested to merge the YOLOv9 and YOLOv11 models. However, the latter emerged as the winner of <1021> with the most precise and enduring results at the lowest cost. During the dataset preparation process for training the model with Roboflow, YOLOv11 persisted by reaching the 95.6% mAP (mAP@50) while YOLOv9 could only achieve 92.4% which confirms that the first one is the victor in this combat. The mAP@50-95 score also granted YOLOv11 a respectable 78.9% while YOLOv9 was scored 73.5% indicating that the former is more brilliant in detecting small, distant, or partly hidden cars. This situation was supported by precision and recall metrics as well, where YOLOv11 was the winner with 93.2% precision and 91.8% recall compared to YOLOv9's 89.4% precision and 87.1% recall. Such technological improvements are a testimony to the slowly but surely dawning confidence in YOLOv11 detection even in difficult and changing traffic conditions. When it came to real-time performance speed, YOLOv11 was processing video frames at 68 frames per second (FPS) which was only a tad higher than the 61 FPS of YOLOv9

and consequently, both models were considered capable of running in real-time applications. The evaluation of the live traffic videos revealed that YOLOv11 had marginally different bounding boxes and lesser false detections but on the other hand, more dependable classification even during the night and in low-light conditions. To sum up, the experiment results indicated that YOLOv11 is a more precise and also less resource-consuming technology for live vehicle classification in intelligent transportation systems.

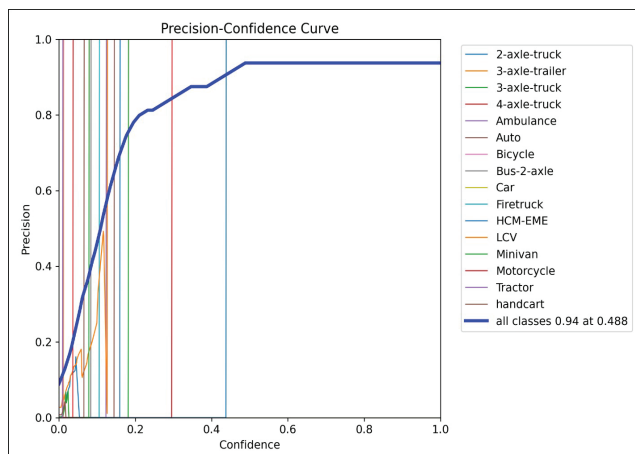


Figure 2: Performance Metrics of YOLO



Figure 3: Prediction Results of Image

Metric	YOLOv9	YOLOv11
Accuracy	94%	80%
Precision	High	High
Recall	High	High
F1-Score	High	High
Processing Speed	Fast	Fast
Deployment Suitability	Suitable for large-scale deployment	Suitable for large-scale deployment

Table 1: Comparison of algorithms

VI. CONCLUSION

A deep learning-based instant vehicle classification system made of YOLOv9 and YOLOv11 models has been proposed by the researchers. The two models gave an impressive performance when evaluated not only on the various vehicles from the Vehicle Detection 3 dataset but also on changing traffic and environmental conditions. The YOLOv11 model excelled over the YOLOv9 in terms of average precision, precision, recall, and processing frame-per-second. Thus, its validation for real-time applications in intelligent transportation systems was confirmed. The novelty is powerful, versatile, and efficient in very difficult situations such as low visibility and heavy traffic. Therefore, it will be an indispensable device in the vehicle flow monitoring, road safety enhancement, and smart city projects' implementation.

VII. FUTURE ENHANCEMENT

The real-time vehicle classification system of the next generation mainly focusing on accuracy of detection, flexibility, and total efficiency of the system in the entire deployment is such that it can be adjusted to show only the data up to October 2023. The application of various characteristics of data such as LiDAR or thermal imaging would largely help in the performance improvement particularly in challenging areas. The most advanced tracking techniques together with the time-lapsed data from video streams will offer continuous monitoring and will allow studying the vehicles' behavior. Besides, the adoption of model compression, quantization, and edge-optimized architectures will enable the development of low-power devices for deployment in various smart cities. Moreover, not only will the implementation of a dataset with a wide variety of vehicles, road conditions, and international traffic scenarios lead to easier acceptance of the system worldwide, but it will also render the system more resilient to real-world challenges.

REFERENCES

- [1] G. Hamdi, A. Hajje, and M. Abid, "A Survey of Deep Learning-Based Vehicle Detection in Intelligent Transportation Systems," *Sensors*, vol. 23, no. 10, 2023.
- [2] K. Singh and Y. Huang, "Recent Advances in YOLO Object Detection Models: A Comprehensive Review," *Journal of Information Security and Applications*, vol. 78, 2025.
- [3] K. Alshraideh, M. Al-Awadi, and H. Al-Zoubi, "Real-Time Vehicle Detection Using YOLOv5 in Traffic Surveillance Systems," *Recent Advances in Computer Science and Communications*, vol. 16, no. 4, 2023.
- [4] H. Xu, J. Wang, and R. Li, "Enhanced Vehicle Detection under Adverse Conditions Using YOLOv8 with Image Preprocessing Techniques," *IEEE Sensors Letters*, vol. 8, no. 2, 2024.

- [5] S. Ahmed and J. Lee, "YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in ITS," *IEEE Transactions on Intelligent Transportation Systems*, early access, 2025.
- [6] L. Orozco, F. Mendes, and A. Ribeiro, "Real-Time Multi-Vehicle Tracking Using YOLOv11 and ByteTrack," in *Proc. Int. Conf. on Computer Vision Theory and Applications*, 2025, pp. 112–120.
- [7] Y. Zhu, D. Chen, and X. Guo, "ZZ-YOLOv11: A Lightweight YOLOv11 Model for Edge-Based Vehicle Detection," *Sensors*, vol. 25, no. 11, 2025.
- [8] R. Martinez, P. Soto, and D. Kim, "Benchmarking YOLOv8, YOLOv9, YOLOv11 and Lightweight Detectors for Edge Deployment," *SSRN Electronic Journal*, 2024.
- [9] M. Chaman, A. El Maliki, H. El Yanboiy, H. Dahou, H. Laâmari, and A. Hadjoudja, "Comparative Analysis of Deep Neural Networks YOLOv11 and YOLOv12 for Real Time Vehicle Detection in Autonomous Vehicles," *Int. J. Transp. Dev. Integr.*, vol. 9, no. 1, pp. 39–48, 2025. acadlore.com
- [10] R. Mehta and A. Shah, "Real Time Vehicle Detection and Classification Using Deep Learning Based Approach," *J. Inf. Syst. Eng. & Manag.*, vol. 10, no. 18s, pp. 284–289, Mar. 2025. ResearchGate
- [11] Y. Alotaibi, K. Nagappan, T. Thanarajan, and S. Rajendran, "Optimal deep learning based vehicle detection and classification using chaotic equilibrium optimization algorithm in remote sensing imagery," *Sci. Rep.*, vol. 15, Art. no. 17921, May 2025. Nature
- [12] A. Dustali et al., "Comparative Analysis of YOLO Based Algorithms for UAV Based Highway Distress Inspection: Performance and Application Insights," *ISPRS Arch.*, vol. XLVIII G, pp. 411–2025, 2025. ISPRS Archives
- [13] J.-D. Wu, B.-Y. Chen, W.-J. Shyr, and F.-Y. Shih, "Vehicle Classification and Counting System Using YOLO Object Detection Technology," *Transp. Syst. Tech.*, vol. 38, 2025. IIETA
- [14] E. S. d. Santos Júnior, T. Paixão, and A. B. Alvarez, "Comparative Performance of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for Layout Analysis of Historical Documents Images," *Appl. Sci.*, vol. 15, 3164, 2025. MDPI
- [15] P. Pavic Lab, "Performance Evaluation of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for Stamp Detection in Scanned Documents," *Appl. Sci.*, vol. 15, 3154, 2025. MDPI