



International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com
editor@ijerst.com

Smart Vehicle Classification Framework for Traffic Analysis and Sustainable Mobility

Mr. Ch. Sagar¹|Muppalla Sarvani²|Gaveriboina Siva Mani Santosh³|Panguluri Syam
Pranav⁴|Tirupathi Muni Himanth⁵.

^{1,2,3,4,5}Department of CSE-AI, Chalapathi Institute of Engineering and Technology,
LAM, Guntur, Andhra Pradesh, India

ABSTRACT:The rapid growth of vehicle imagery on digital platforms has increased the demand for accurate and efficient vehicle classification systems, particularly for applications such as traffic surveillance, autonomous driving, and intelligent transportation management. In this study, we propose a Convolutional Neural Network (CNN)based multi-class classification model capable of recognizing seven vehicle categories: Auto Rickshaws, Bikes, Cars, Motorcycles, Planes, Ships, and Trains. The dataset consists of 5,007 JPG images, of which 3,504 images are used for training and 1,503 for testing. The CNN architecture is carefully designed to extract discriminative and hierarchical features unique to each vehicle type. Extensive experimentation and hyperparameter tuning are conducted to optimize the model's performance. The results demonstrate the effectiveness of CNNs in vehicle recognition tasks. Accurate vehicle classification contributes to improved traffic flow, reduced congestion, and enhanced transportation efficiency, thereby supporting sustainable urban development. Furthermore, the proposed system aligns with goals related to responsible consumption and production, industrial innovation, infrastructure development, and climate action by enabling smarter and more environmentally efficient transportation solutions.

KeyWords: Vehicle Classification, Convolutional Neural Network, Deep Learning, Image Recognition, Intelligent Transportation Systems, Multi-Class Classification

1. INTRODUCTION

The rise of smart cities and the integration of advanced technologies into urban infrastructure have dramatically transformed how cities manage resources, public safety, and transportation systems.

As urban populations increase, one of the most pressing concerns for city planners and authorities is traffic management. With the development of Intelligent Transportation Systems (ITS), real-time vehicle classification plays a vital role in

enhancing traffic flow, enforcing regulations, and improving overall road safety. Automated vehicle classification systems, especially those using real-time surveillance data, offer promising solutions for these challenges. This research addresses the need for accurate, real-time vehicle classification using deep learning, specifically Convolutional Neural Networks (CNNs), applied to CCTV-captured images. Vehicle classification is essential for a wide range of applications, such as traffic monitoring, road pricing, toll collection, law enforcement, and autonomous driving. Accurate vehicle identification is a fundamental element of ITS, which aims to provide efficient traffic flow, minimize accidents, and contribute to the optimization of urban mobility. Traditional vehicle identification methods, such as manual counting or sensor-based systems, are limited in scalability, accuracy, and flexibility. These methods often require significant human involvement or fail to perform well in dynamic environments with varying conditions, such as changing lighting, occlusions, and diverse vehicle types. However, with the advent of deep learning, particularly CNNs, there is a substantial potential to automate and improve the accuracy of vehicle classification systems. Convolutional Neural Networks have become the cornerstone of modern

computer vision applications due to their ability to automatically learn hierarchical features from images. These networks excel at extracting patterns from complex visual data and are particularly effective for tasks such as object recognition, image segmentation, and classification. Despite the progress in CNN-based vehicle classification models, there remain significant challenges when applying these models to CCTV images in real-world environments. CCTV cameras are often subject to various constraints, such as varying image resolution, lighting conditions, occlusions, and camera angles. These factors complicate the task of vehicle classification, as they can distort or obscure critical features needed for accurate identification.

2. LITERATURE SURVEY

Vehicle classification from CCTV images has gained significant attention in recent years due to the rise of intelligent transportation systems (ITS) and the growing need for smart city initiatives. Traditional vehicle identification techniques, such as manual classification and sensor-based systems, have significant limitations, including low scalability, high human intervention, and limited performance in dynamic environments [1]. With the advancements in computer vision and deep learning, the use of Convolutional Neural Networks (CNNs)

has become a preferred approach for vehicle classification. However, deploying deep learning models in real-world scenarios, particularly from CCTV images, remains a challenge due to variations in lighting, resolution, occlusions, and viewpoint distortions. This literature review aims to highlight the various methods employed for vehicle classification and discuss the limitations faced by these approaches in real-world applications.

A pivotal study by [2] proposed a deep learning model utilizing CNNs for vehicle classification from CCTV images. The authors employed a single CNN architecture trained on a large dataset of vehicle images to classify vehicles into multiple categories. This model was shown to achieve high classification accuracy in controlled conditions, but it struggled when applied to real-time surveillance footage, where variations in resolution, lighting, and occlusions reduced performance. The study highlighted the importance of robust feature extraction, as minor distortions could significantly affect classification results. To address this, the authors proposed a data augmentation strategy to simulate different environmental conditions, such as motion blur and occlusion. Despite these improvements, the method's reliance on a single CNN

model limited its adaptability to diverse vehicle types and environmental conditions.

3. SYSTEM ARCHITECTURE

The system architecture for the smart vehicle classification framework is designed as an end-to-end pipeline that integrates data acquisition, preprocessing, feature extraction, classification, and real-time deployment components to ensure accurate and efficient vehicle recognition. Initially, vehicle images are collected from sources such as CCTV cameras and datasets, forming the input layer of the system. These images undergo preprocessing, where they are resized, normalized, and augmented to improve quality and robustness against variations like lighting, occlusion, and motion blur. The processed images are then passed into the core classification module, which in the existing system uses a standard Convolutional Neural Network (CNN) consisting of convolution, pooling, and fully connected layers to extract hierarchical features and perform classification. In contrast, the proposed system employs a Multi Stream Transfer Fusion with Contextual Attention (MSTF CA) architecture, which integrates multiple deep learning models such as DenseNet, InceptionNet, and MobileNet to capture complementary features including

fine textures, multi-scale patterns, and global representations.

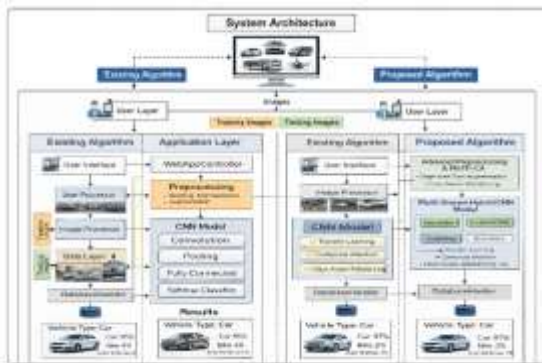


Fig1: System architecture

3.1 Methodology

The proposed system architecture for smart vehicle classification is designed as an end-to-end deep learning pipeline integrating data acquisition, preprocessing, hybrid feature extraction, and classification modules. Initially, vehicle images are collected from diverse sources such as CCTV footage and publicly available datasets and organized into predefined categories. In the preprocessing stage, images are resized to a uniform resolution, normalized and enhanced using advanced data augmentation techniques including rotation, flipping, brightness adjustment, motion blur simulation, and occlusion handling to improve robustness. The processed images are then fed into the proposed Multi Stream Transfer Fusion with Contextual Attention (MSTF CA) model, which employs a hybrid architecture combining DenseNet, Inception Net, and MobileNet encoders. Each encoder extracts complementary

features, where DenseNet focuses on fine-grained textures, InceptionNet captures multi-scale spatial patterns, and MobileNet ensures computational efficiency with global feature representation.

3.2 Dataset Description: The dataset contains 5,007 vehicle images across seven classes: Auto Rickshaw, Bike, Car, Motorcycle, Plane, Ship, and Train. It is split into training and testing sets and includes variations in lighting, background, and angles to improve model generalization for real-world conditions.

3.3 Data Preprocessing: All images are resized to a uniform resolution and normalized to improve training stability and efficiency. Data augmentation techniques such as flipping, rotation, zooming, and brightness adjustment are applied to enhance robustness and handle class imbalance before splitting into training and testing sets.

3.4 Model Development: A baseline CNN model is used for feature extraction and classification, while the proposed MSTF-CA model integrates DenseNet, InceptionNet, and MobileNet in a multi-stream fusion framework. A contextual attention mechanism enhances important features, improving classification accuracy and robustness.

3.5 Performance Evaluation: Model performance is evaluated using accuracy, precision, recall, F1-score, and confusion

matrix analysis. The proposed MSTF-CA model outperforms the standard CNN by achieving higher accuracy and better class-wise prediction reliability under real-world variations.

4. DESIGN AND CONSTRUCTION

The design and construction of the Smart Vehicle Classification Framework for Traffic Analysis and Sustainable Mobility are centered on the implementation of the proposed Multi Stream Transfer Fusion with Contextual Attention (MSTF CA) model to achieve high accuracy and robustness in real-world environments. The system is structured as a modular pipeline consisting of data acquisition, preprocessing, hybrid model integration, and deployment components. During the design phase, a multi-stream architecture is developed by integrating pre-trained DenseNet, InceptionNet, and MobileNet models, enabling the extraction of complementary features such as fine textures, multi-scale spatial patterns, and efficient global representations. A contextual attention mechanism is incorporated to enhance feature relevance by focusing on critical regions within vehicle images, especially under challenging conditions like occlusion and motion blur. In the construction phase, the system is implemented using deep learning frameworks, where pre-processed images are fed into the hybrid model, and feature

fusion is performed using transfer learning techniques. Fully connected layers and a softmax classifier are used for final prediction across multiple vehicle categories. The system also integrates class-aware rebalancing to address dataset imbalance and includes a real-time feedback mechanism for continuous learning and performance improvement after deployment. This well-structured design ensures scalability, adaptability, and high efficiency, making the system suitable for intelligent traffic monitoring and sustainable urban mobility applications

5. RESULTS AND DISCUSSION

The proposed CNN model was tested on 1,503 vehicle images across seven classes and showed strong overall classification performance. It achieved stable training convergence with decreasing loss and effective feature learning from both low- and high-level patterns. While highly distinct classes like planes, ships, and trains were classified well, some confusion occurred between similar categories such as bikes and motorcycles.



Fig 2: Python Environment

Fig 2 shows that the project is developed using Python, TensorFlow, and Django. A virtual environment is used to manage dependencies.



Fig 3: Home Page

Fig 3 shows the home page with a simple and user-friendly interface. It includes navigation options like Home, About Us, Contact Us, Login, Admin, and Register. Users can easily access different features of the system through these options.



Fig 4: Loading Dataset

Figure 4 shows the image upload page that allows users to select and upload vehicle images for classification. The uploaded images are processed by the trained deep learning model to predict the vehicle type.



Fig 5: Prediction Inceptionv3

Figure 5 illustrates the vehicle classification interface with the InceptionV3 model selected. The user uploads an image and submits it for processing. The system analyses the input and predicts the vehicle category.



Fig 6: InceptionV3 Result

Fig 6 shows the classification result with the predicted vehicle type, model accuracy, and successful execution status. It also includes a performance graph for visualization efficiently to generate accurate results.



Fig 7: Mobile net Result

Fig 7 presents the final prediction displaying the identified vehicle type along with the MobileNet model performance. It highlights the accuracy score and confirms successful model execution with a visual performance chart.

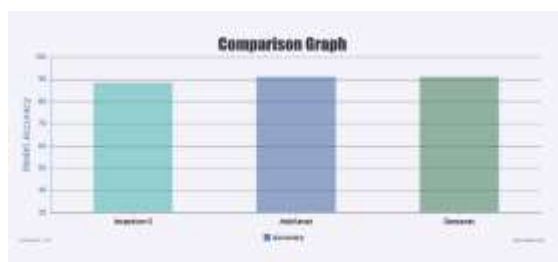


Fig 8: Model Accuracy Comparison Graph

Fig 8 shows a comparative analysis of model accuracies for InceptionV3, MobileNet, and DenseNet in the vehicle classification system. It highlights the performance differences among the models in identifying vehicle types. The results indicate that MobileNet and DenseNet achieve slightly higher accuracy, making them more suitable for real-time and reliable classification.

6. CONCLUSION

The vehicle classification system effectively leverages deep learning models, such as DenseNet, InceptionNet, and MobileNet, to achieve high accuracy in classifying vehicles from CCTV images. The integration of techniques like multi-stream fusion, contextual attention, and data augmentation ensures that the system remains robust under varying conditions such as motion blur, lighting changes, and occlusion. Furthermore, the use of a feedback loop allows the system to continuously learn from user input, further enhancing its accuracy and adaptability in dynamic environments. The real-time performance and high

classification accuracy make this system a valuable tool for intelligent transportation systems, smart city applications, and automated vehicle monitoring.

REFERENCES

1. MA Berwo, A Khan, Y Fang, H Fahim, S Javaid. (2023). Deep learning techniques for vehicle detection and classification from images/videos: A survey. *Sensors*.
2. D Shokri, C Larouche, S Homayouni. (2023). A comparative analysis of multi-label deep learning classifiers for real-time vehicle detection to support intelligent transportation systems. *Smart Cities*.
3. MI Basheer Ahmed, R Zaghdoud, MS Ahmed. (2023). A real-time computer vision based approach to detection and classification of traffic incidents. *Big Data and Cognitive Computing*.
4. AJ Moshayedi, AS Roy, A Taravet, L Liao, J Wu. (2023). A secure traffic police remote sensing approach via a deep learning-based low-altitude vehicle speed detector through UAVs in smart cities. *Future*.
5. HF Yang, J Cai, C Liu, R Ke, Y Wang. (2023). Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning. *Transportation Research Part C*.

6. M Hasanujjaman, MZ Chowdhury, YM Jang. (2023). Sensor fusion in autonomous vehicles with traffic surveillance camera system: detection, localization, and AI networking. *Sensors*.
7. M Qasim Gandapur, E Verdú. (2023). ConvGRU-CNN: Spatiotemporal deep learning for real-world anomaly detection in video surveillance systems. *Sensors*.