License Plate Recognition with Vehicle Validation for Fraud Detection

Kishore L[1], Karthikeyan K[2], Mr. Gnanapriyan F[3]
[1,2] *UG Scholar, Rajalakshmi Institute of Technology, Chennai*
[3]*Assistant Professor, Rajalakshmi Institute of Technology, Chennai*

Abstract: **Over the past few years, the overall prevalence of counterfeit and illegal vehicle license plates has been a major challenge to transport authorities and law enforcement agencies. To combat this issue, the current paper proposes an end-to-end computer vision and deep learning-based system to detect counterfeit vehicle license plates effectively. The system begins by identifying the make and model of a vehicle using a trained deep neural network, thereby enabling verification of the physical features of the vehicle. It then identifies and crops the license plate area from the vehicle image using advanced object detection techniques. Then, Optical Character Recognition (OCR) is used to scan the alphanumeric characters on the license plate. Instead of relying on a centralized API, the system uses web scraping techniques to fetch corresponding registration information from publicly available government databases or transport portals. Any mismatch between the predicted make and model and the fetched registration information, or evidence of tampering on the license plate, triggers an automatic alert for further investigation. The end-to-end solution offers a scalable, cost-effective, and efficient solution for vehicle authentication and prevention of the spread of counterfeit license plates.**

Keywords: **Computer Vision, Deep Learning, Fake Plate Identification, Optical Character Recognition (OCR), Vehicle License Plate Detection, Vehicle Make and Model Recognition (VMMR), Vehicle Verification, Web Scraping.**

1. INTRODUCTION

The identification of fake vehicle number plates is becoming increasingly significant due to its link to criminal behavior, toll cheating, and fraud. Computer vision-based Automatic License Plate Recognition (ALPR) systems and deep learning algorithms have become essential tools for traffic monitoring and law enforcement agencies [1][2]. Detecting tampered or fake plates remains an uphill task due to changing lighting, a heterogeneous variety of cars, and random plate arrangements. The latest deep learning models, including YOLOv8, and EfficientNet, have improved recognition significantly [3][4][5]. Experiments conducted by Mustafa and Karabatak [1] and Kothai et al. [6] confirm the efficacy of real-time detection and stable feature extraction in surpassing environmental constraints.

While most ALPR systems capture only plate numbers, they do not authenticate the data from official vehicle databases. To address the limitation, vehicle make and model recognition combined with ALPR has been more effective in fraud detection [5][7][8]. This paper proposes a more sophisticated system that not only identifies and detects license plates but also uses web scraping to retrieve official registration data for authentication. Web scraping is differentiated from API-based approaches as a cost-effective and flexible approach, especially in regions with no central access [9][10]. The combination of deep learning-based detection, CNN-based classification, and data verification provides a solid end-to-end solution for identifying fake plates and improving vehicle authentication accuracy [11].

2. LITERATURE SURVEY

Vehicle License Plate Recognition (LPR) systems are essential for traffic surveillance and police enforcement. Following the adoption of deep learning, their accuracy and effectiveness have notably increased. Mustafa and Karabatak [1] proposed a real-time system for license plate and vehicle detection, which confirmed the potential of deep learning in challenging situations such as occlusion and varying illumination. Models like YOLOv8 have proven to be robust in plate detection, with Al-Hasan et al. [2] and Naidu et al. [5] providing improvements to cope with regional differences and environmental adversities.
Besides plate identification, make and model classification of vehicles is needed to detect fraud.

Hassan et al. [7] applied CNNs for detecting inconsistencies between vehicle features and official records, while Pustokhina et al. [8] utilized CNNs and K-means clustering for enhancing identification rates. Kothai et al. [6] also presented a new feature extraction method that enhances detection when images are low quality and have low lighting.To complement data validation, web scraping has proven to be an economical substitute for APIs in obtaining vehicle registration information. Shrivastava et al. [9] and Ağgül and Erdemir [10] highlighted its use in data consistency verification and fraud detection. Generally, the combination of deep learning for detection and classification, and web scraping for verification, enhances LPR systems and the security of transport networks.

3. RESEARCH SUMMARY

It is important to develop a safe and smart vehicle LPR system that can identify suspicious activities like toll evasion and car-related offenses. The system in question combines deep learning methods for detecting license plates and classifying the make and model of the vehicle, and web scraping in order to cross-match data from the government and third-party sources.

In contrast to other API-based verification processes, web scraping presents a more flexible and cheaper alternative in areas without a centralized database of vehicles. YOLOv8 are utilized for real-time detection under difficult conditions, and pre-trained CNNs for recognizing vehicle model and make in order to identify inconsistencies that might reveal counterfeit plates.By integrating LPR, classification, and data validation via web scraping, the system promotes high accuracy, scalability, and real-time applicability towards improving traffic monitoring and law enforcement operations.

4. PROPOSED METHODOLOGY

System Overview

The proposed architecture combines two powerful deep learning models: YOLOv8n for vehicle license plate detection and EfficientNet-B0 for vehicle make and model classification. The license plate detection model detects and localizes the license plates from the images of the vehicle, and the classification model identifies the make and model of the vehicle. In order to authenticate the vehicle, the detected license plate is compared against government databases via web scraping, thus confirming that the information of the vehicle matches the officially registered information. This holistic approach significantly enhances the accuracy and reliability of the vehicle identification process and offers an efficient method of fraud detection in vehicular systems.

Dataset Used:

Vehicle Classification Dataset:

Dataset Name: Stanford Cars Dataset

Content: The database contains over 16,000 annotated and tagged images of automobiles by their model and make.

Purpose: To create a model through which the vehicle can be classified according to appearances, to enable system identification and predict the make and model of vehicles.

License Plate Detection Dataset:

Dataset Name: Roboflow 'vehicle-registration-plates-truck' dataset

Content: It contains over 8,000 images of car license plates.

Objective: To train a model that can accurately detect and localize license plates in car images.

1. License Plate Detection

For license plate detection, YOLOv8n is used, a lightweight and real-time object detection model with a trade-off between speed and accuracy. The model is trained for 50 epochs with a batch size of 8 and an input size of 512×512 pixels. It uses YOLO's composite loss function that takes bounding box regression and classification loss into account, optimized using the Adam optimizer with a dynamic learning rate scheduler.

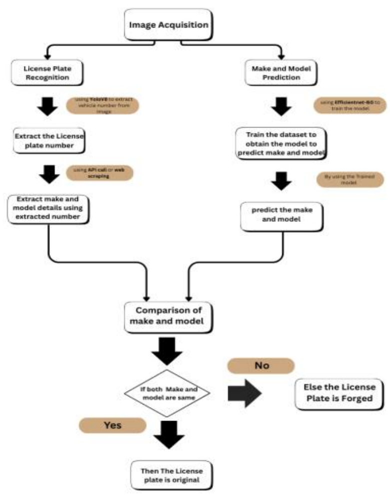

Fig.1 System Architecture

2. License Plate Extraction

Once the license plate region is localized properly, it is cropped and passed to PaddleOCR, a strong optical character recognition engine that is extremely robust and precise for distorted and multilingual text. PaddleOCR performs character segmentation and recognition and reads out the license plate number from the image.

3. Vehicle Make and Model Recognition

For vehicle make and model identification, EfficientNet-B0 is employed, utilizing its effective scaling and high performance in image fine-grained classification. The pre-trained ImageNet model is fine-tuned by replacing the last layer with a variant that adjusts the number of vehicle classes. The model is trained for 20 epochs with a batch size of 32 and CrossEntropyLoss and Adam optimizer with a learning rate of 0.001. All the input images are resized to 224×224 pixels to accommodate the model's input.

4. Data Verification

After extracting the license plate number and making subsequent predictions on the vehicle make and model, the system performs data verification via web scraping of official vehicle registration data from government and credible third-party online databases.The data gathered is then compared with the classification results; any inconsistency between the vehicle characteristics detected and the official records could be an indication of a possible case of license plate forgery.

5. IMPLEMENTATION

The system is designed as a modular pipeline that combines several stages, with each stage revolving around a significant vehicle authentication feature. It is implemented in Python, leveraging popular deep learning libraries such as PyTorch and OpenCV, as well as OCR libraries such as PaddleOCR. The key modules are described below:

1. Image Acquisition:

Vehicle images are captured in real time from surveillance camera feeds or downloaded from available benchmark data sets such as the Stanford Cars data set and the Roboflow license plate data sets. These images form the basis of input for the system.

2. Vehicle and License Plate Detection:

Detected vehicles and their license plates are identified using a YOLOv8n model, which is selected based on its fast processing speed and accuracy. Vehicle and license plate bounding box annotated datasets are employed to train the model. Training is configured with an image size of 512×512, batch size of 8, and 50 epochs. The output of this process includes cropped license plate regions and vehicle bounding boxes.

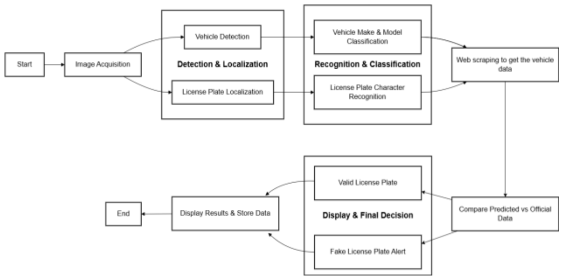

Fig.2 Proposed System Flow Chart

3. License Plate Recognition using PaddleOCR:

The cropped license plate images are processed by PaddleOCR, which recognizes and detects characters. PaddleOCR is chosen because it supports multilingual languages, has fast inference, and high accuracy, especially on distorted or partially occluded plates. The output of this module is a string of the recognized license plate number.

4. Vehicle Make and Model Classification:

To confirm the authenticity of the vehicle, make and model are predicted by a CNN model. EfficientNet-B0, pre-trained on ImageNet and fine-tuned on the Stanford Cars dataset, is employed in this system. Input images of 224×224 pixels are employed, and the model is trained with a batch size of 32 for 20 epochs. Cross-entropy loss and the Adam optimizer (learning rate 0.001) are employed to optimize the classification performance.

5. Obtaining Official Data using Web Scraping:

After license plate number extraction, a web scraping module is employed to collect official vehicle registration information from open government or third-party vehicle data websites. This functionality is obtained through the utilization of libraries such as BeautifulSoup and Selenium, which offer dynamic content parsing support. The information extracted typically includes information about the make, model, fuel type, year of registration, and ownership status of the registered vehicle.

6. Attribute Matching for Fraud Detection:

The predicted vehicle attributes (make and model) are programmatically compared with the official statistics collected through web scraping. A mismatch between the two data sets is regarded as a potential fraud signal, i.e., a counterfeit or manipulated license plate.

7. Result Display:

The final results include the license plate number, the expected make and model of the vehicle, official registration details, and whether there's a fraud match or mismatch. These results are presented clearly, either through an easy-to-use graphical interface or a command line, so that users or law enforcement can quickly understand and act on the information.

6. RESULT

The proposed vehicle license plate verification system was tested on different datasets and compared in real-world scenarios for determining its validity in identifying spurious license plates by image processing, classifying, and matching with official records.

License Plate Recognition

YOLOv8 was employed for vehicle detection as well as their respective license plates. The model performed well on conventional object detection metrics:

mAP@50: 0.98

mAP@50-95: 0.71

Precision: 0.99

Recall: 0.95

The results show the model's robust ability to distinctly recognize license plates under varying real-world conditions, including occlusion and lighting.

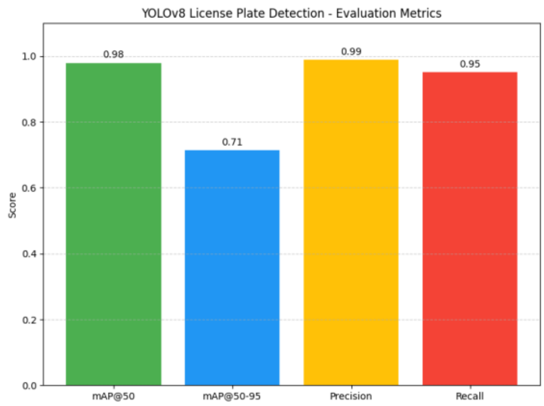

Fig.3 Evaluation Metric for License Plate Detection

Vehicle Make and Model Recognition

The model reached a test accuracy of 98.67%, reflecting a strong capacity to generalize to unseen images of vehicles. The classification report showed high precision, recall, and F1-scores across most of the classes. From Figure 4, the confusion matrix reflects a distinct diagonal pattern, validating a low level of misclassifications. Mistakes in predictions occasionally happened among visually similar models, as one can expect in fine-grained classification tasks.

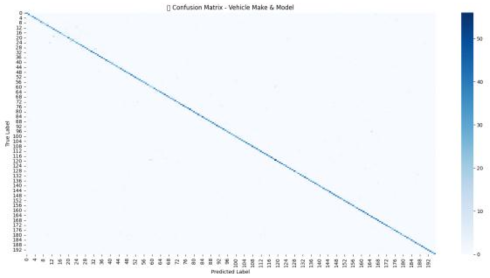

Fig.4 Confusion matrix for VMMR

Figures 5 and 6 display the result of the license plate localization and extraction procedures, respectively. The YOLOv8n model was used in these procedures to successfully identify the license plate area in the car image. The identified license plate was subsequently extracted and fed to the PaddleOCR engine, which successfully extracted and read the alphanumeric characters in the license plate, thereby enabling the subsequent verification stages in the pipeline.

Fig.5 License Plate Localization

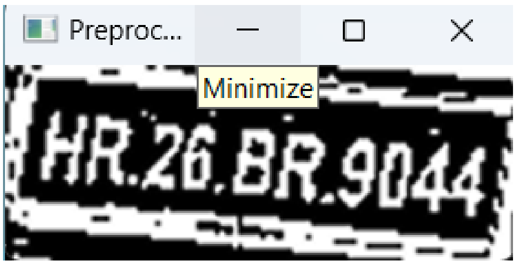

Fig.6 License Plate Extraction

Figure 7 presents the predicted result for vehicle make and model from a test dataset image, while Figure 8 showcases the output generated from a real-world car image, demonstrating the model's effectiveness in practical scenarios.

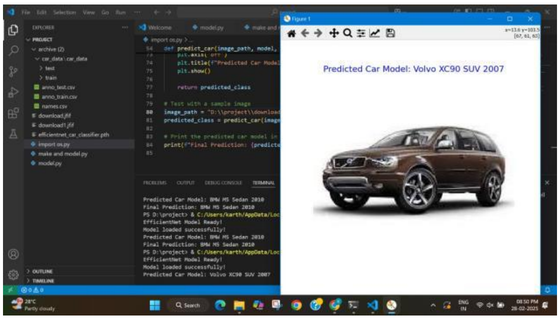

Fig.7 Predicted Output fromTest Dataset

Fig.8 Real-world Prediction Output

Figures 9 and 10 show the system's final output, where the extracted license plate number and predicted vehicle make and model are compared against official registration data obtained via web scraping. The official information retrieved from the corresponding database was then cross-checked against the predictions.A discrepancy was observed between the documented and anticipated vehicle information , which prompted the system to flag it as a possible case of fraud. The outcome is therefore clearly marked, offering helpful guidance to law enforcement agencies or automated monitoring systems.This outcome confirms the effectiveness of the proposed system in real-time detection of fraudulent license plates, presenting a robust system for fraud detection and intelligent traffic surveillance

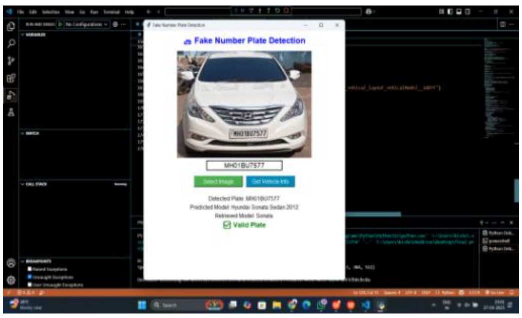

Fig.9 Valid License Plate

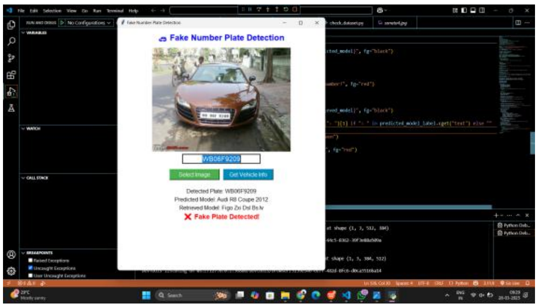

Fig.10 Invalid License Plate

These results confirm the overall effectiveness, accuracy, and real-world applicability of the proposed system in detecting counterfeit license plates and assisting law enforcement in real-time.

7. CONCLUSION

This work proposed an end-to-end and low-cost vehicle license plate verification system that can detect false vehicle registrations using deep learning and web-based data verification techniques. By integrating strong models like YOLOv8 for license plate detection, PaddleOCR for text identification, and pre-trained CNNs like EfficientNet for classifying vehicle make and model, the system demonstrated strong performance.The use of web scraping as an alternative to traditional APIs in the retrieval of official information has proved to be highly flexible and scalable, especially where there are no centralized vehicle registration databases. The end-to-end pipeline was able to effectively identify inconsistencies between the data collected on vehicles and the respective official records, effectively identifying potentially fraudulent or fake license plates.The robustness of the system to changing environmental conditions and its generalizability to real-world surveillance videos also validate its efficacy.Briefly, the solution here presented provides a practical solution for intelligent traffic monitoring and fraud detection to promote public security and support law enforcement operations with automation and AI-based decision-making.

REFERENCES

[1] T. Mustafa and M. Karabatak, "Real Time Car Model and Plate Detection System by Using Deep Learning Architectures," *IEEE Access*, vol. 12, pp. 107616–107630, 2024, doi: 10.1109/ACCESS.2024.3430857.

[2] T. M. Al-Hasan, V. Bonnefille, and F. Bensaali, “Enhanced YOLOv8-Based System for Automatic Number Plate Recognition,” *Technologies*, vol. 12, no. 9, p. 164, 2024, doi: 10.3390/technologies12090164.

[3] U. G. Naidu, R. Thiruvengatanadhan, S. Narayana, and P. Dhanalakshmi, “A Novel Method for Indian Number Plate Detection and Recognition Using EfficientNet,” *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 6, pp. 2408–2418, 2024.

[4] H. Shi and D. Zhao, "License Plate Recognition System Based on Improved YOLOv5 and GRU," *IEEE Access*, vol. 11, pp. 10429–10439, 2023, doi: 10.1109/ACCESS.2023.3240439.

[5] A. Hassan, M. Ali, N. M. Durrani, and M. A. Tahir, "An Empirical Analysis of Deep Learning Architectures for Vehicle Make and Model Recognition," *IEEE Access*, vol. 9, pp. 91487–91499, 2021, doi: 10.1109/ACCESS.2021.3090766.

[6] I. O. De Oliveira, R. Laroca, D. Menotti, K. V. O. Fonseca, and R. Minetto, "Vehicle-Rear: A New Dataset to Explore Feature Fusion for Vehicle Identification Using Convolutional Neural Networks," *IEEE Access*, vol. 9, pp. 101065–101077, 2021, doi: 10.1109/ACCESS.2021.3097964.

[7] I. V. Pustokhina *et al.*, "Automatic Vehicle License Plate Recognition Using Optimal K-Means With Convolutional Neural Network for Intelligent Transportation Systems," *IEEE Access*, vol. 8, pp. 92907–92917, 2020, doi: 10.1109/ACCESS.2020.2993008.

[8] G. Kothai, E. Povammal, S. Amutha, and V. Deepa, “An Efficient Deep Learning Approach for Automatic License Plate Detection with Novel Feature Extraction,” *Procedia Computer Science*, vol. 235, pp. 2822–2832, 2024, doi: 10.1016/j.procs.2024.04.267.

[9] N.-A. Alam, M. Ahsan, M. A. Based, and M. Haider, “Intelligent System for Vehicles Number Plate Detection and Recognition Using Convolutional Neural Networks,” *Technologies*, vol. 9, no. 1, p. 9, 2021, doi: 10.3390/technologies9010009.

[10] S. Shrivastava, S. K. Singh, K. Shrivastava, and V. Sharma, "CNN Based Automated Vehicle Registration Number Plate Recognition System," in *Proc. 2020 2nd Int. Conf. Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida, India, 2020, pp. 795–802, doi: 10.1109/ICACCCN51052.2020.9362737.

[11] B. Ağgül and G. Erdemir, “Development of a Counterfeit Vehicle License Plate Detection System by Using Deep Learning,” *Balkan Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 252–257, 2022, doi: 10.17694/bajece.1093158.

[12] C. Henry, S. Y. Ahn, and S.-W. Lee, "Multinational License Plate Recognition Using Generalized Character Sequence Detection," *IEEE Access*, vol. 8, pp. 35185–35199, 2020, doi: 10.1109/ACCESS.2020.2974973.

[13] S. Luo and J. Liu, "Research on Car License Plate Recognition Based on Improved YOLOv5m and LPRNet," *IEEE Access*, vol. 10, pp. 93692–93700, 2022, doi: 10.1109/ACCESS.2022.3203388.

[14] M.-A. Chung, Y.-J. Lin, and C.-W. Lin, "YOLO-SLD: An Attention Mechanism-Improved YOLO for License Plate Detection," *IEEE Access*, vol. 12, pp. 89035–89045, 2024, doi: 10.1109/ACCESS.2024.3419587.

[15] S. H. Tan, J. H. Chuah, C.-O. Chow, and J. Kanesan, "Coarse-to-Fine Context Aggregation Network for Vehicle Make and Model Recognition," *IEEE Access*, vol. 11, pp. 126733–126747, 2023, doi: 10.1109/ACCESS.2023.3330114.

[16] Q. Huang, Z. Cai, and T. Lan, "A Single Neural Network for Mixed Style License Plate Detection and Recognition," *IEEE Access*, vol. 9, pp. 21777–21785, 2021, doi: 10.1109/ACCESS.2021.3055243.

[17] H. Moussaoui, N. E. Akkad, M. Benslimane, *et al.*, “Enhancing Automated Vehicle Identification by Integrating YOLO v8 and OCR Techniques for High-Precision License Plate Detection and Recognition,” *Scientific Reports*, vol. 14, p. 14389, 2024, doi: 10.1038/s41598-024-65272-1.

[18] X. Liu, Y. Wang, D. Yu, and Z. Yuan, "YOLOv8-FDD: A Real-Time Vehicle Detection Method Based on Improved YOLOv8," *IEEE Access*, vol. 12, pp. 136280–136296, 2024, doi: 10.1109/ACCESS.2024.3453298.

[19] X. Shao, C. Wei, Y. Shen, and Z. Wang, "Feature Enhancement Based on CycleGAN

for Nighttime Vehicle Detection," *IEEE Access*, vol. 9, pp. 849–859, 2021, doi: 10.1109/ACCESS.2020.3046498.

[20] A. Tourani, A. Shahbahrami, S. Soroori, S. Khazaee, and C. Y. Suen, "A Robust Deep Learning Approach for Automatic Iranian Vehicle License Plate Detection and Recognition for Surveillance Systems," *IEEE Access*, vol. 8, pp. 201317–201330, 2020, doi: 10.1109/ACCESS.2020.3035992.