International Journal of Science and Engineering Invention (IJSEI)
Volume 11, Issue 01, April 2025
https://doi.org/10.23958/ijsei/vol11-i01/278

Research Article

A Deep Learning Framework for Accurate Number Plate Recognition using OWL-V2 and PaddleOCR

Ch. V.M.S.N. Pavan Kumar [1], V. Anusha *[2], V. Hemanjali [2], T. Komali [2], T. Jeevan Teja [2]

[1]Assistant Professor, Department of ECE, Bapatla Engineering College, Bapatla, Andhra Pradesh, India.
[2]UG Student, Department of ECE, Bapatla Engineering College, Bapatla, Andhra Pradesh. India.

*Corresponding author: V. Anusha; *anushavenigandla76@gmail.com*

Received 11 March 2025; Accepted 13 April 2025; Published 17 April 2025

Abstract

Recognizing number plates holds significant importance in various applications such as law enforcement, traffic management, toll collection, parking management, and security. Manual recognition is challenging one and is prone to errors. Automation of number plate recognition enables faster inference and timely response from the police personnel. Artificial intelligence techniques using deep learning models are useful in this regard. Typically, Automated Number Plate Recognition (ANPR) technology to identify and process vehicle license plates effectively Such as Image Capture, Image Processing, Character Recognition, Alerts or Actions. This paper presents a vision-language model-based approach for automatic detection. Number Plate Recognition (NPR) is a technique that identifies alphanumeric characters from license plates and converts them into text format. In this work, different deep learning models were utilized for the detection and recognition of number plates. The model was tested on a car image, processed and number plate datasets. For this recognition task, zero-shot models such as OWLVIT and Grounding DINO, were employed. Additionally, techniques like PaddleOCR were integrated. The proposed tests demonstrated that the system can accurately detect number plates with an impressive accuracy of 92.91%, even under challenging conditions.

Keywords: ***Convolutional Neural Network, Deep Learning, Zero-Shot learning and OCR technologies.***

1. Introduction

Number Plate Recognition (ANPR) systems play a crucial role in modernizing vehicle identification and management processes. They enhance traffic control by monitoring violations and ensuring road safety, while also strengthening security by identifying stolen vehicles and helping law enforcement. ANPR streamlines toll collection and parking management, making these services more efficient and user-friendly. As a key component of smart city initiatives, these systems contribute to optimizing urban mobility and reducing congestion. Additionally, ANPR saves time and costs by automating tasks, minimizing errors, and providing reliable data for investigations and legal proceedings. Their integration into various sectors makes ANPR indispensable for improving operational efficiency and safety. Automated Number Plate Recognition (ANPR) systems provide invaluable support to law enforcement agencies by enhancing their efficiency and effectiveness. These systems enable real-time identification of stolen or suspect vehicles, helping in crime prevention and investigation. By automating surveillance and monitoring traffic violations, ANPR ensures better enforcement of laws while reducing the workload on officers. Integration with databases allows seamless access to critical information, and customizable alerts facilitate targeted operations. Furthermore, ANPR systems collect reliable evidence with timestamps and location data, which can be crucial in legal proceedings. Overall, ANPR empowers law enforcement to maintain public safety and uphold the law more effectively. Automated Number Plate Recognition (ANPR) systems have revolutionized traffic management by providing efficient and accurate vehicle monitoring capabilities. These systems help in real-time tracking of vehicles, enabling authorities to manage traffic flow effectively and reduce congestion. ANPR plays a key role in detecting traffic violations, such as speeding or unauthorized lane usage, and helps ensure compliance with regulations. Additionally, ANPR aids in toll collection by automating the process and minimizing delays at toll booths. It also supports urban planning initiatives by providing valuable data on traffic patterns, helping cities design better infrastructure. In cases of emergencies or accidents, ANPR assists in identifying vehicles involved and facilitates quicker responses. By integrating ANPR with smart city technologies, governments and municipalities can create intelligent traffic systems that enhance safety, reduce travel times, and optimize resource utilization. Overall, ANPR has become an indispensable tool for improving traffic management and ensuring smoother transportation systems [1]. Automated Number Plate Recognition (ANPR) systems are becoming essential tools for vehicle identification and management worldwide, with applications growing across diverse regions. In developed regions like North America and Europe, ANPR is widely used for traffic law enforcement, toll collection, and surveillance, supported by advanced infrastructure and stringent traffic regulations. Asia-Pacific, particularly countries like China and India, is witnessing rapid adoption of ANPR due to urbanization, smart city projects, and increasing vehicle volumes. In emerging markets such as Latin America and Africa, the uptake of ANPR is slower but steadily growing, often used for crime prevention and improving traffic flow. In the Middle East, countries like the UAE are integrating ANPR into large-scale smart city and security initiatives. Globally, ANPR systems are transforming traffic management, tolling systems, and security operations, while also raising debates around data privacy and ethical concerns. Automated Number Plate Recognition (ANPR) systems are increasingly being adopted in India to address challenges like traffic congestion, law enforcement, and toll collection. These systems are used for monitoring traffic activities, such as detecting red-light violations, enforcing speed limits, and identifying stolen vehicles. ANPR also plays a significant role in electronic toll collection, reducing delays and improving efficiency on highways. India's unique challenges, such as diverse license plate formats and high traffic density, have led to the development of advanced ANPR technologies tailored to the country's needs. Cities like Hyderabad and Bengaluru are integrating ANPR into smart city projects to enhance urban mobility and traffic management. With the growing focus on digital infrastructure and smart cities, ANPR systems are becoming a vital tool for improving road safety, reducing congestion, and supporting law enforcement efforts across India [2]. A zero-shot approach in automated number plate recognition (ANPR) involves using pre-trained models that can generalize to new tasks or domains without requiring additional training on specific datasets. This method leverages advanced machine learning techniques, such as transfer learning and natural language processing, to recognize and interpret number plates without prior exposure to the specific data. For instance, models like OWL-ViT (Object-Wise Vision Transformer) can be used to detect and localize number plates in images, while tools like PaddleOCR handle the text extraction process. These models are pre-trained on diverse datasets, enabling them to perform well even in unfamiliar scenarios. This approach is particularly useful in applications like traffic monitoring, toll collection, and law enforcement, where adaptability and efficiency are crucial [3]. Recent studies in Automated Number Plate Recognition (ANPR) highlight the integration of advanced machine learning techniques and technologies like IoT to address challenges such as varying plate formats and environmental conditions. Modern approaches, including Faster R-CNN and other deep learning models, focus on enhancing recognition accuracy through sophisticated pre-processing techniques like Gaussian Blur and edge detection. Comprehensive reviews also compare traditional template-based matching with cutting-edge methods, showcasing the evolution of ANPR systems in applications like traffic management and law enforcement. These advancements underline the growing importance of robust datasets and adaptable algorithms in the field [4]. This paper is organized as follows: A brief explanation of related works is given in section2. The dataset description is mentioned in section3. Proposed algorithm is explained in section4. Proposed work is explained section5. Results and simulations are described in section6. Conclusion is presented in section 7.

2 Related Works

Automated Number Plate Recognition (ANPR) is essential for smart transportation, aiding in traffic management, toll collection, and law enforcement. Recent studies leverage deep learning, including CNNs, YOLO models, and Transformer-based approaches, to enhance accuracy and efficiency under challenging conditions. The following references highlight key advancements in plate detection, character recognition, and system optimization for real-world applications. Lubna [5] explains the importance of Automated Number Plate Recognition in smart travel systems, focusing on how computers analyze images to identify vehicles by their number plates, despite challenges like damaged plates or poor weather. The study explores methods like Convolutional Neural Networks for plate extraction and highlights integration with GPS, RFID, and the internet to improve traffic management, toll collection, and law enforcement. Chris Henry [6] developed a system that efficiently handles diverse number plates and challenging conditions, using Tiny YOLOv3 for plate detection and YOLOv3-SPP for quick letter recognition by treating them as separate objects. A unique rule ensures accurate sequencing for singleor double-line plates, and tests on global datasets, like the Korean KarPlate set, showed excellent speed and performance without extra data. Anamika Rakshe [7] created a system to automate toll collection and improve traffic flow, using image processing steps like noise removal, edge detection, and letter recognition with Convolutional Neural Networks. It captures images from high-quality cameras, storing them in a centralized database for easy monitoring, and performed well in tests with real and synthetic data across varying light, speeds, and angles. Jithmi Shashiranga [8] examines challenges like lighting, weather, and varied plate designs, comparing step-by-step methods for detecting, segmenting, and reading plates with all-in-one approaches. It highlights how deep learning tools, such as Convolutional Neural Networks and YOLO, boost accuracy, emphasizing the importance of robust datasets and practical solutions. Han Luo [9] introduced DIOU-NMS to improve plate detection accuracy and a streamlined system for speed, using LPRNet to read plates end-to-end without segmentation. Mohammed Sameel Shaikh [10] described a deep learning system for vehicle plate recognition, using YOLOv3 and YOLOv4 with Darknet 53 for accurate detection, and Tesseract for letter reading from video frames. Hengliang Shi [11] developed an advanced deep learning system for detecting and reading plates in challenging natural environments, enhancing YOLOv5 with a new L-SE focus technique for small objects and combining GRU with Connectionist Temporal Classification for seamless letter recognition. Faraz Imtiyaz Mir [12] improved Automated Number Plate Recognition using advanced imaging and smart learning like Faster R-CNN to monitor vehicles and enhance safety, employing high-quality cameras, Gaussian Blur, edge detection, and Optical Character Recognition for accurate readings. Ao Xiong and Yuqi Yang [13] introduced YOLO-M4ST, an enhanced YOLOv8 system with VTR, tailored for detecting and reading railway mast plates, using MobileNetV4 for speed, C2f-Star for better detail extraction, and a three-step VTR for seamless letter recognition. It's ideal for complex railway environments. Zilu Wang [14] developed an improved plate recognition system for adverse weather, utilizing Histogram-YOLO for plate detection and an Enhanced Transformer Block to extract details and correct images in real time.

3. Dataset Details

Dataset has an important contribution towards the learning of any deep learning model as the model learns and derives feature from it. The dataset1 is labelled as processed and it contains 1001 images. The dataset 2 is labelled as cars-imgs and it contains 866 images. The dataset 3 is labelled as number plate and it contains 1700 images. The dataset 3 is extracted from [15]. This collection of pictures is made to help find numbers on number plates or license plates. People gathered these pictures from different vehicles around Bapatla Engineering College and Bapatla town in Andhra Pradesh, India. There are 1700 pictures taken with cameras like Realme 8i (50MP), IQOO Z9 5G (50MP), Vivo T2X 5G (50MP), and Redmi

Note 13 Pro 5G (200MP). The pictures are all different because they were taken from many angles, far and close distances, and in different light. So, this set can be used to check if systems that read number plates, turn pictures into words, or recognize letters work well. Inside the zipped file, there's a folder with all the pictures and an excel file too. The excel file has the name of each picture and the real number plate number that goes with it.

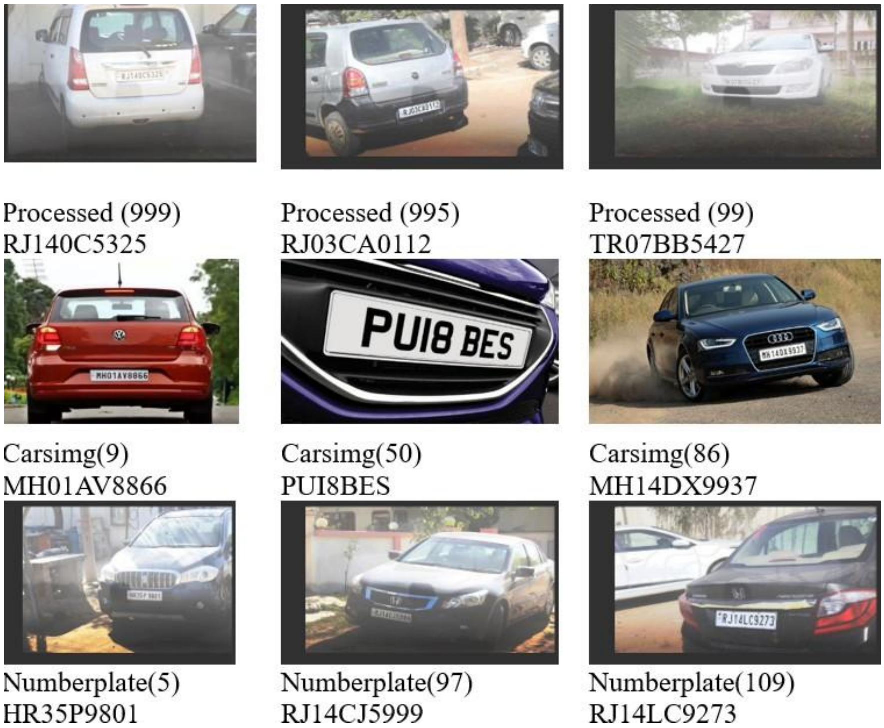

Figure 1: Sample Images of Datasets

4. Proposed Algorithm

The flowchart outlines a systematic and well-defined process that incorporates the principles of contrastive learning, the creation of dataset classifier the application of a zero-shot methodologies. OWL-ViT is an advanced AI model designed for zero-shot text-conditioned object detection, meaning it can identify objects in images based on text descriptions without being trained on specific categories. The process begins by receiving an image and a text prompt describing the object to be detected. The model then extracts visual features from the image using a Vision Transformer (ViT) and processes the text query using a masked self-attention Transformer. To align image and text features, OWL-ViT leverages CLIP contrastive learning, which enhances the model's ability to relate visual elements to words. Instead of relying on predefined classification labels, the model dynamically generates object labels based on text inputs, making it highly adaptable for detecting unfamiliar objects. The object detection mechanism is refined through bipartite matching loss, ensuring accurate classification and bounding box predictions. The final output consists of bounding boxes and labels that identify the objects found in the image, adjusted for image size and confidence thresholds. OWL-ViT's ability to detect objects without prior category-specific training makes it a powerful tool for tasks requiring open-world detection, such as identifying rare or newly introduced objects based on descriptive text prompts. OWL-ViT is an advanced AI model designed for zero-shot text-conditioned object detection, meaning it can identify objects in images based on text descriptions without being trained on specific categories. The process begins by receiving an image and a text prompt describing the object to be detected. The model then extracts visual features from the image using a Vision Transformer (ViT) and processes the text query using a masked self-attention Transformer. To align image and text features, OWL-ViT leverages CLIP contrastive learning, which enhances the model's ability to relate visual elements to words. Instead of relying on predefined classification labels, the model dynamically generates object labels based on text inputs, making it highly adaptable for detecting unfamiliar objects. The object detection mechanism is refined through bipartite matching loss, ensuring accurate classification and bounding box predictions. The final output consists of bounding boxes and labels that identify the objects found in the image, adjusted for image size and confidence thresholds. OWL-ViT's ability to detect objects without prior category-specific training makes it a powerful tool for tasks requiring open-world detection, such as identifying rare or newly introduced objects based on descriptive text prompts.

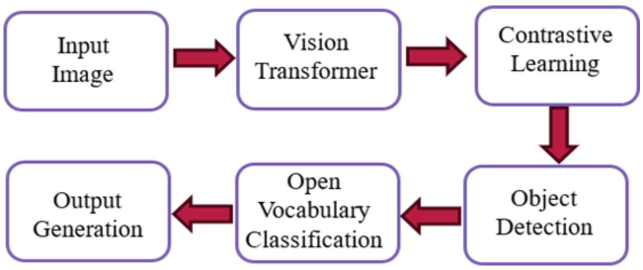

Figure 2: Block diagram of OWL-ViT

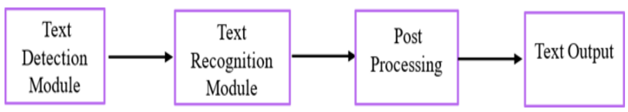

Figure 3: Flow chart of PaddleOCR

Paddle OCR is an advanced Optical Character Recognition (OCR) framework to efficiently extract text from images. It supports multiple languages and is optimized for real-world applications such as document scanning and license plate recognition. The algorithm consists of three main components. First, the text detection module identifies areas in the image containing text using models like EAST or DBNet, ensuring accurate localization. Next, the text recognition module extracts the actual text from the detected regions using sequence-to-sequence models like CRNN (Convolutional Recurrent Neural Network). Finally, the post-processing step refines the output by correcting errors and formatting the extracted text for better accuracy. The workflow begins with image preprocessing, which enhances text visibility through resizing and normalization. The text detection module then scans the image to locate regions containing text, producing bounding boxes around them. Once detected, the text recognition module extracts the text from those regions, converting images into readable information. The final output includes extracted text along with its position in the image.

5. Proposed Deep Learning Framework for Number Plate Detection

The flowchart demonstrates an automated system designed for vehicle number plate recognition, beginning with an input image containing the plate. The image undergoes preprocessing to enhance quality through tasks such as noise reduction and edge detection. A zero-shot model is then employed to detect and localize the number plate, functioning effectively without the need for extensive labelled training data. Once identified, the number plate area is cropped and processed using an Optical Character Recognition (OCR) model, which extracts the text characters.

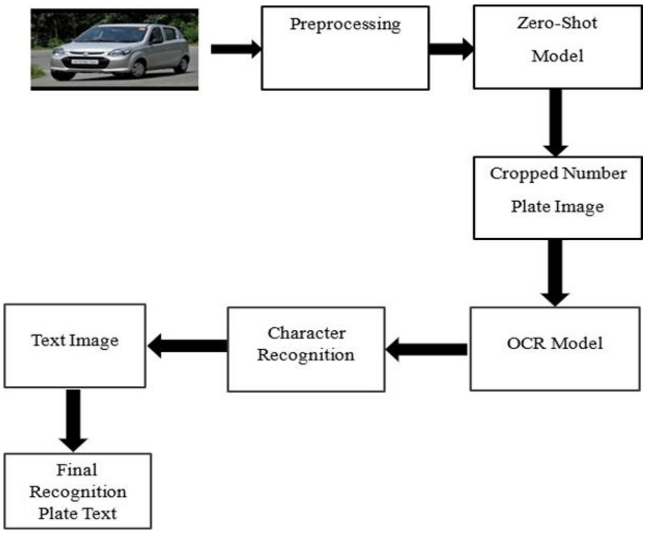

Figure 4: Flowchart of Number Plate Recognition

These characters are further refined through a character recognition step, ensuring high accuracy. The recognized text is then converted into an image format for validation. The final output is the accurately recognized plate text, which can be displayed or stored for applications like traffic monitoring, security enforcement, and automated toll collection. This efficient system highlights the practicality and benefits of automating number plate recognition in real world scenarios.

6. Results and Discussions

All Results and simulation are done by using the Google Colab Notebook and the PyTorch framework. In this paper OCR model, Zero Shot Models are used for number plate recognition. Accuracy: This refers to the proportion of correctly detected images divided by total number of images in the dataset.

$$\text{Accuracy} = \frac{True\ Positives + True\ Negatives}{Total\ Observations} \quad \text{....(1)}$$

Precision: Precision measures how accurate the detected or recognized plates are. It calculates the ratio of true positive results (correctly detected and recognized plates) to all detected plates.

$$\text{Precision} = \frac{True\ Positives}{Positives + False\ Positives}\ \text{True} \quad \text{.....(2)}$$

True Positives (TP): Correctly detected and recognized plates. False Positives (FP): Plates that were incorrectly detected or falsely recognized.

False Negatives (FN): Plates that the system missed.

F1-Score: This mixes precision and recall together to make one number that shows how well they balance. We get it by doing a special average of precision and recall so both matter the same. The way to count it is:

$$\text{F1-Score} = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad \text{......(3)}$$

Recall: This, sometimes called sensitivity, shows how good the model is at finding all the real yes things. We work it out by taking the number of right guesses (true positives) and dividing by all the real yes things (true positives plus false negatives). The way to do it is:

$$\text{Recall} = \frac{True\ Positives}{Positives + False\ Negatives}\ \text{True} \quad (4)$$

Article geometry caption [Table]xcolor margin=1in

The table 1 refers the comparison and performance of two zero shot models, owlv2-base patch16 and grounding-dino-base as well as PaddleOCR, across three datasets processed, cars-imgs, and numberplate. On the processed dataset, owlv2 achieves higher Accuracy, F1 Score, and Recall compared to grounding-dino, while both models maintain perfect Precision. For the cars images dataset, the models show similar results, with owlv2 slightly exceeding groundingdino in Accuracy and F1 Score, but grounding-dino performs better in Recall. Precision remains perfect for both. On the numberplate dataset, grounding-dino excels in Accuracy and F1 Score, while owlv2 shows marginally better Recall both maintain a perfect Precision score. Overall, each model demonstrates unique strengths depending on the dataset, making them

Table 1: Comparison of Zero-Shot Models with PaddleOCR on 3 Datasets

Zero Shot Models	**owlv2-base patch16**	**Grounding dino base**
OCR Model	**Paddle OCR**	
PROCESSED		
Accuracy	92.91%	83.65%
Precision	1	1
F1 score	0.946391	0.903628
Recall	0.952875	0.862953
CARSIMGS		
Accuracy	75.33%	74.83%
Precision	1	1
F1 score	0.916749	0.916382
Recall	0.856429	0.874293
NUMBERPLATE		
Accuracy	80.15%	86.45%
Precision	1	1
F1 score	0.917212	0.924765
Recall	0.92345	0.85832

Suitable for different application contexts.

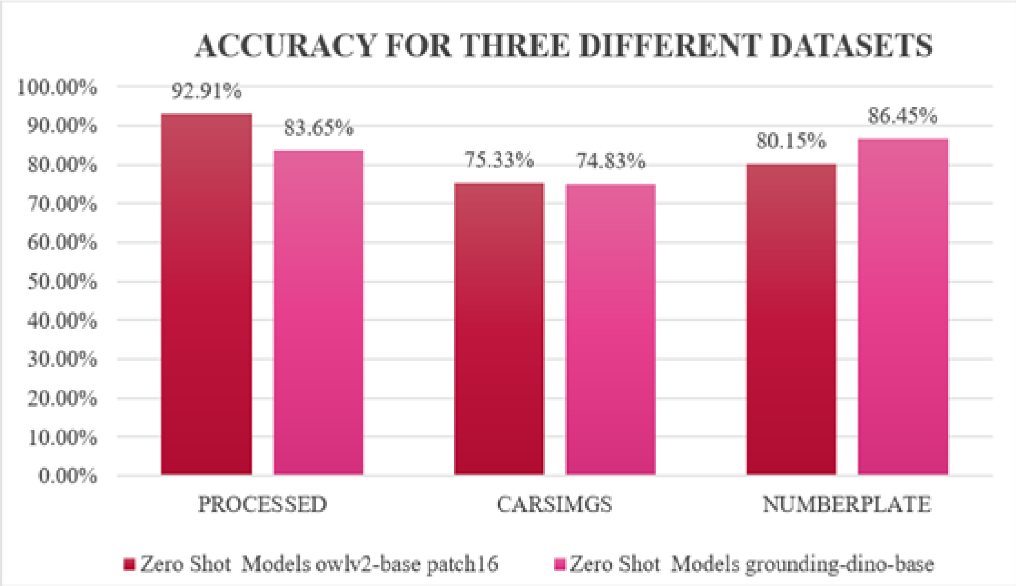

Figure 5: Results of accuracy

The Accuracy results for all the models are represented in graph in figure 5 and tabular representation in table 1. In the processed dataset the owlv2-base patch16 gets the highest accuracy value of 92.91% and grounding dino gets accuracy of 83.65%. In the cars-imgs dataset the owlv2-base patch16 gets 75.33 of accuracy and grounding dino gets accuracy of 74.83%. In the numberplate dataset the owlv2-base patch16 gets accuracy of 80.15%and the grounding dino gets the accuracy of 86.45%. In the above graph the highest accuracy for owlv2-base patch16 gets 92.91% and lowest accuracy for that model is 75.33% and for grounding dino model gets highest accuracy of 86.45% and the lowest accuracy for grounding dino model gets 74.83%. The figure 6 shows the graphical representation of Precision, F1 score and Recall results of two zero shot models as well as OCR model.

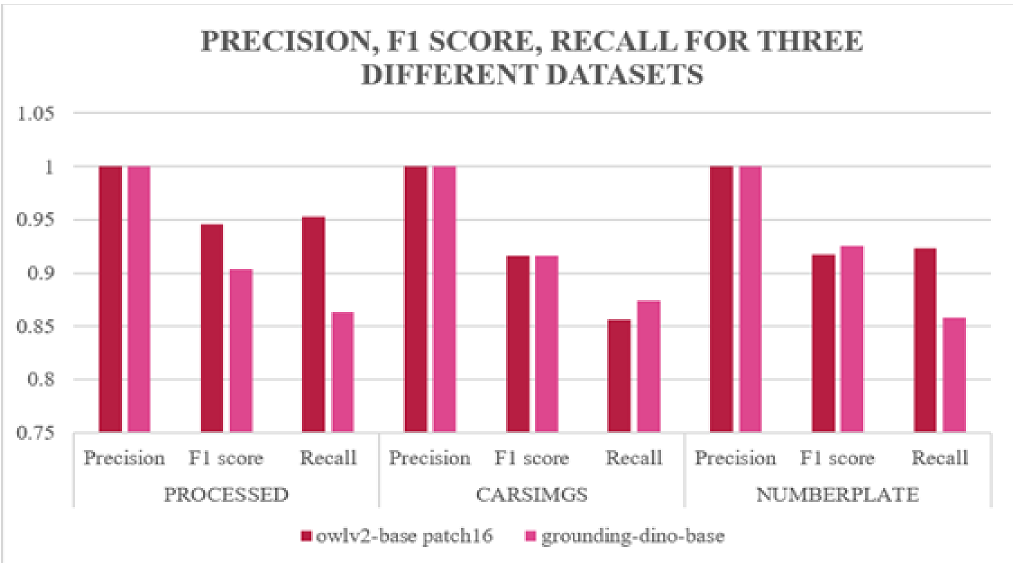

Figure 6: Precision, F1 score and Recall for various zero-shot algorithms on 3 data sets

The Precision, Recall and F1 score results of two zero shot models and paddleOCR model is represented in tabular form in table 1. The highest Recall and F1 score for owlv2-base patch 16 for dataset processed. For all models and all datasets the precision value gets 1. The lowest Recall and F1 score for grounding dino.

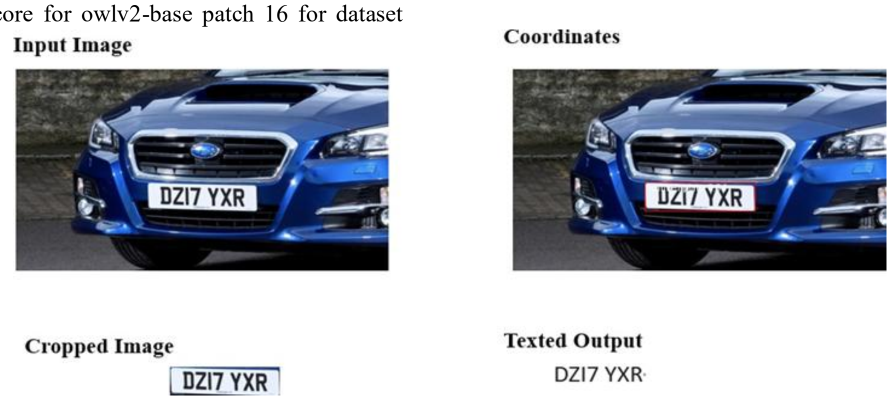

Figure 7: sample images of the output

This research demonstrates the effective integration of advanced models such as OWLV2, Grounding DINO, and PaddleOCR for automatic number plate recognition. The process begins with providing an input image containing a vehicle and its number plate. The zero-shot detection models, OWL-V2 and Grounding DINO, analyze the image and accurately detect the location of the number plate by generating a bounding box around it. The identified region is then cropped, isolating the plate for further processing. This cropped image is passed through PaddleOCR, an optical character recognition model that extracts and converts the text from the number plate into digital format.

7. Conclusion

This research shows that using zero-shot object detection along with optical character recognition (OCR) can greatly improve automatic number plate recognition. By training on a custom dataset, the system reaches a high accuracy of 92.91%, proving that combining these advanced methods makes plate detection more effective. The study highlights the need for constant improvements in ANPR systems to keep up with the changing demands of modern transportation. Overall, it demonstrates how zero-shot learning helps enhance plate recognition, making it more adaptable and efficient in real-world use.

References

[1] N. Nagul, “Automated Number Plate Recognition (ANPR) Systems: Enhancing Traffic Management and Security,” Unpublished, March 2025.

[2] J. Y. Tu, “Research on license plate recognition algorithm based on deep learning in complex scenes,” M.S. thesis, School Comput., Zhejiang Univ. Technol., Hangzhou, China, 2020.

[3] Laxmi Narayan Soni, “Integrating Template Matching for License Plate Detection and CNN-based Face Recognition for Driver Identification in Vehicle Monitoring Systems”, SSRN Electronic Journal, 2024.

[4] Hsin-Yi Lin, Yi-Quan Li, Daw-Tung Lin, “System Implementation of Multiple License Plate Detection and Correction on Wide-Angle Images Using an Instance Segmentation Network Model”, IEEE Transactions on Consumer Electronics, vol.70, no.1, pp.4425-4434, 2024.

[5] Lubna, Mufti, N., Shah, S. A. A., “Automatic Number Plate Recognition: A Detailed Survey of Relevant Algorithms”, Sensors, vol. 21, no. 9, pp. 3028, 2021.

[6] C. Henry, S. Y. Ahn, and S.-W. Lee, “Multinational License Plate Recognition Using Generalized Character Sequence Detection,” IEEE Access, vol. 8, pp. 35185-35192, Feb. 2020. DOI: 10.1109/ACCESS.2020.2974973.

[7] A. Rakshe and N. Dongre, “An Efficient Methodology of Automatic Vehicle Number Plate Detection Using Deep Learning,” International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 12s, pp. 640-651, 2024. [Online].

[8] J. Shashirangana, H. Padmasiri, D. Meedeniya, and C. Perera, “Automated License Plate Recognition: A Survey on Methods and Techniques,” IEEE Access, vol. 9, pp. 11203-11220, Jan. 2021.

[9] S. Luo and J. Liu, “Research on Car License Plate Recognition Based on Improved YOLOv5m and LPRNet,” IEEE Access, vol. 10, pp. 93692-93700, Sept. 2022. DOI: 10.1109/ACCESS.2022.3203388.

[10] M. S. Shaikh, Y. Nilekani, H. Shinde, and S. L. Varma, “Automatic Number Plate Detection and Recognition Using Deep Learning,” International Journal of Research and Analytical Reviews (IJRAR), vol. 10, no. 2, pp. 666-673, Apr. 2023.

[11] H. Shi and D. Zhao, “License Plate Recognition System Based on Improved YOLOv5 and GRU,” IEEE Access, vol. 11, pp. 10429-10439, Feb. 2023. DOI: 10.1109/ACCESS.2023.3240439.

[12] F. I. Mir, N. N. Sarkar, Y. Arora, and A. K. Sharma, “Recent Developments in Automatic Number Plate Detection and Recognition,” International Journal of Innovative Research in Computer Science and Technology (IJIRCST), vol. 12, no. 5, pp. 28-35, Sept. 2024. DOI: 10.55524/ijircst.2024.12.5.5.

[13] X. Huang, S. Yang, A. Xiong, and Y. Yang, “Enhanced YOLOv8 With VTR Integration: A Robust Solution for Automated Recognition of OCS Mast Number Plates,” IEEE Access, vol. 12, pp. 179648-179655, Dec. 2024. DOI: 10.1109/ACCESS.2024.3505208.

[14] Z. Wang, L. Zheng, and G. Li, “Design of Enhanced License Plate Information Recognition Algorithm Based on Environment Perception,” IEEE Access, vol. 13, pp. 38609-38627, Mar. 2025. DOI: 10.1109/ACCESS.2025.3544634.

[15] https://zenodo.org/records/13954136.