

# Automatic Number Plate Recognition System for Indian Number Plates using Machine Learning Techniques

Gayatri Hajare<sup>1</sup>, Utkarsh Kharche<sup>2</sup>, Pritam Mahajan<sup>3</sup>, and Apurva Shinde<sup>4</sup>

Department of Computer Engineering, Ramrao Adik Institute of Technology D.Y Patil Deemed to be University, Nerul

<sup>1</sup>[gayatriumesh34@gmail.com](mailto:gayatriumesh34@gmail.com), <sup>2</sup>[utkarshkharche29101998@gmail.com](mailto:utkarshkharche29101998@gmail.com), <sup>3</sup>[ironpritam@gmail.com](mailto:ironpritam@gmail.com), <sup>4</sup>[apurva.karkhanis@rait.ac.in](mailto:apurva.karkhanis@rait.ac.in)

**Abstract.** India being a country where the population is above 1.3 billion where each person has at least one car of his/her use. Considering this, the number of cars driven on the roads of India must be greater than the population of the people in the country. India being a diverse country, diversity is not only seen in the language of the number plates but also in size, spacing between the letters on the number plate and the font of the number plate. Diversity differs from state to state. Even though most of the people are using English Number plates, there is no certain law as to how a number plate should be, so some people tend to have number plates according to their preferences. To withstand these problems, we have created a system using You Only Look Once version 5 (YOLOv5) for number plate detection and Google Tesseract for Character Recognition.

**Keywords:** YOLOv5, Indian Number Plates, Tesseract Engine, Optical Character Recognition.

## 1 Introduction

Automatic Number Plate Recognition (ANPR) is a tool that can be used not only at toll booths on highways, express ways, etc. to speed up the process of toll collection but also used by car parking managements at malls, movie theatres, etc. This system can also help government authorities to catch and fine individuals who are trying to break traffic rules and regulations as well as help us locate criminals who use vehicles to flee. By detecting vehicle's number plate from CCTV feeds can help to locate stolen vehicles.

India, with a population of 1.3 billion, has the second-largest population [13] and the second largest road network [1] in the world having a road network of 63.72 lakh km has its own needs and requirements for an ANPR system, even though there are certain rules regarding the size of number plate of certain vehicles [2]. These rules aren't strictly followed. Different kinds of number plate patterns are found like number plates with different fonts having fancy lettering while number plates are sometimes hand painted on the vehicle especially seen on trucks, etc. The ANPR system along with a combination of sensors is used to track vehicle speed and specifically detect cars that are speeding. Here, a new ticket is generated automatically by estimating the distance between the two cameras. This assists authorities to maintain legislation and rules, which

also in return help them to decrease vehicle collisions.

ANPR System provides us an appropriate solution to provide safe parking management in residences. Registered Resident Vehicles are directly allowed to enter and park in their fixed parking spot whereas the non-registered vehicles will be first added into guest list then it will assign them an empty parking spot. In India, around 2 lakh cars are there that are being stolen each year [13]. This number can be decreased if proper precautions are taken. ANPR system can be employed for locating and tracking cars such that if any vehicle is stolen, authorities would be able to detect it from the CCTV feed and locate the path taken by the stolen vehicle and help them to find the vehicle in less time.

Most ANPR systems are made specific to certain type of Number plates [8,11,12] while some are developed based on certain specific criteria [10,14]. Like [11], Silva and Jung proposed a system specifically for Brazilian number plates by taking a combination of European number plates for training number plate detection model. While in [10], They have given a review of all different ANPR techniques that can be used based on certain criteria.

## 2 Literature Survey:

Silva, et al. [11] proposed a 4 stage ANPR system in. Unlike other ANPR models where license detection is the first stage, they decided to do vehicle detection first. So, no vehicle with a visible license plate is missed. For vehicle detection, instead of

training a model from the start they decided to use an existing model (YOLOv2) based on certain criteria. License detection is done by using a CNN named WPOD-NET which was developed for taking features from YOLO, SSD and STN. For Character Recognition (OCR), They created their own LP characteristics-based YOLO network which can detect LP without any issues.

Shidore and Narote [6] proposed an ANPR Model divided into three sections (1) License Plate Extraction (2) Character Segmentation and (3) Character Recognition. An image is acquired using a good ANPR camera and then various image pre-processing techniques like Gray scaling, Sobel Edge Detection and Thresholding are applied on the image to detect various candidate LPs in the image. Then using Bounding Box Analysis True LP is extracted from the image. For Character Segmentation, Integration of Character Region Enhancement, Connected Component Analysis and Vertical Projection Analysis is used. Character Recognition is done in two parts 1) Character Normalization and Feature Extraction 2) Character Recognition using SVM as classifier. The accuracy level stated in [6] for segmentation is 80% and 79.84% recognition.

Kulkarni, et al. [8] proposed an ANPR system for toll booths for faster toll collection by implementing a 4 stage ANPR model [10] which is a combination of different algorithms specially for Indian LPs. As [10], There is vehicle detection module but by using inductive sensors which then triggers the capture of the rear of the car and then various pre-processing techniques as applied to acquired image and then license localization is done using a 'Feature based number plate localization' which specially developed just for Indian LPs. For Segmentation, they have implemented a method known as Image Scissoring where the LP is scanned and scissored in such a way that there is no white pixel present there and is copied into a matrix. For Character Recognition, statistical feature extraction is implemented.

Kashyup, et al. [4] proposed an ANPR system consisting of 3 modules. Before extracting the number plate, a combination of image processing techniques was applied. Segmentation of Characters is done using Regionprops function. Character recognition is done in two parts, first Feature Extraction and then the Actual recognition is done by Template Matching. Template matching is the process of finding a small image region called a template and comparing it with the same template in the database. The accuracy of the system mentioned here is 82.6%.

Devpriya, et al. [14] proposed a system similar to [4]. The image is pre-processed by converting the RGB image into gray scale image then morphological operation is carried out then the number plate detection is done by applying Canny Edge Detection and Closing and opening morphological analysis. Here, Character Segmentation is done by Connected Component Analysis (CCA) [6]. Then similar to [4] each segmented characters are recognized using Template Matching.

YOLOv3 model which is commonly used for object detection is used for all the stages of ANPR [5]. Tensorflow is used for everything. The input images are converted into grayscale. Here for LP detection annotated dataset is required. For this task, they have used LabelImg for labeling data such that in the input image, a bounded box is around the number plate which helps in detection. The character segmentation and recognition are done using YOLO models. The accuracy of character recognition by the YOLO model is 99% while that of OpenCV is 93%.

Tiwari and Choudhary [12] proposed an ANPR system using Artificial Neural Networks (ANNs) based on Indian Number Plates for the system. As other ANPR systems pre-processing methods like Gaussian Filter, Wavelet Transform and Image Binarization on the acquired image and then the Number Plate is extracted and then Character Segmentation is done and then using Vector generation, Character Recognition is done.

Naren Babu et al. [7] proposed an ANPR system which is divided into two stages License Plate Detection and Character Recognition. They have trained a 37 class CNN single YOLO model for both detecting the license plate as well as recognizing the characters the license plate. They also have achieved the 100% accuracy in detecting the number plate and 91% accuracy in recognizing the characters of the number plate.

### 3 Proposed System

The proposed system is divided into 3 different stages to understand the working of the system more effectively. A custom YOLOv3 model was created for all the three stages to achieve better accuracy in number plate recognition [7]. In the proposed system, YOLOv5 model used for number plate detection only unlike [7] and further this detected number plate is segmented. We are using YOLOv5 instead of earlier YOLO versions as it can detect smaller objects better while training unlike its predecessor. The system uses Google OCR

Tesseract to recognise and transform the characters on the licence plate into text after segmentation.

### 3.1 The proposed system is divided into the following steps

- 1) Acquisition of the image
- 2) License Plate Detection using the YOLOv5 model.
- 3) Plate Segmentation
- 4) Character Recognition of the License plate using Optical Character Recognition (OCR).

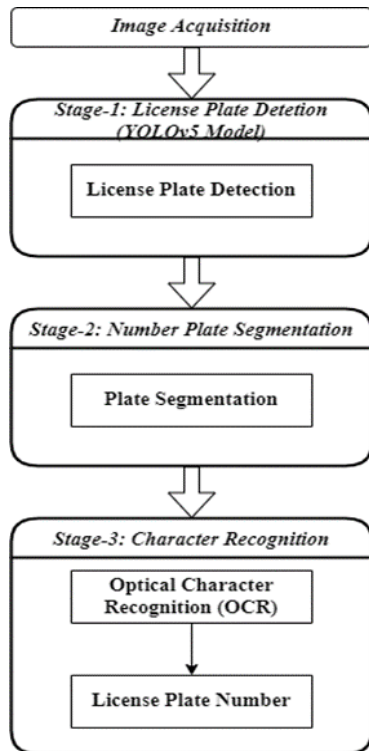


Fig. 1. Architecture of the proposed system

### 3.2 Stepwise explanation of each module

#### 3.2.1 Image Acquisition

Image acquisition of the vehicles is done using CCTV cameras at the toll booths and parking lots. For our system, we used a dataset consisting of different images of vehicles on the internet as well as we captured images from mobile devices that are shown in fig.1.

Our dataset consists of nearly 500 images of different types of vehicles such as bikes, cars, etc. with both front and rear-view. We also have images of license plates without a vehicle for defining the number plate to our YOLO network for license plate detection. Though our dataset doesn't have any specific dimensions i.e., not all images have the same size, while training the dataset we have converted all images to the same size like 640x640.

Also, we annotated the images manually using LabelImg, an open-source API to create a training set in YOLO format.



Fig. 2. Image Acquisition of the vehicle

#### 3.2.1 Stage 1: License Plate Detection

An Automatic Number Plate detection and Recognition system is used to locate the number plate o the vehicle and create the popularity of a plate that will extract text from the image and all others due to calculation modules using local algorithms, plate separation, and character recognition.

We used the YOLOv5 model for license plate detection. It is a Convolutional Neural Network (CNN) used for real-time acquisition where we will use this algorithm to obtain a license plate for a vehicle. CNN can process input images very quickly and detect patterns in them while maintaining its accuracy level.

These types of algorithms are used in AI programs to find specific types of objects based on a selected subject. The YOLOv5 Algorithm is used to locate different types of objects and to distinguish them. YOLOv5 is implemented using the Pytorch library mainly present in Python.

We decided to go with YOLOv5s as its smallest and lightest model with fastest output to cope up with the video frames. Here we can see the difference in different sized models of v5 in accuracy and responsiveness.

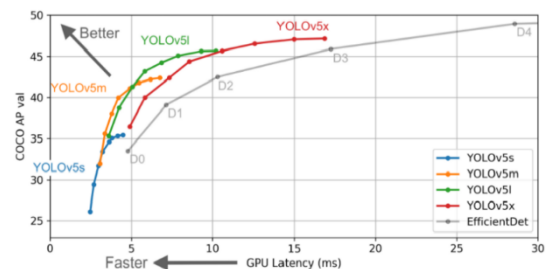


Fig. 3. Accuracy of YOLOv5 compared to other YOLO versions

For the detection of the license plate, The YOLO algorithm uses the following steps:

- 1)Residual Blocks: In this technique, The whole image is converted in form of square grids and each

grid has an dimension  $N*N$ . In this it will highlight the grids which have objects detected in them.

2) Bounding Box Regression: In this technique, A bounding box is formed around the objects found from the whole image. This bounding box has different attributes namely; bounding box width ( $bb_w$ ), bounding box height ( $bb_h$ ), Class  $c$  (this describes which object is what in. Here License plate) and bounding box centre coordinates ( $bb_c$ ,  $bb_r$ ). Here we want to detect the license plate. In Figure 1, The green outlined bounding box along with its class Plate in green/pink represents the detected license plate from the image using YOLOv5 model.

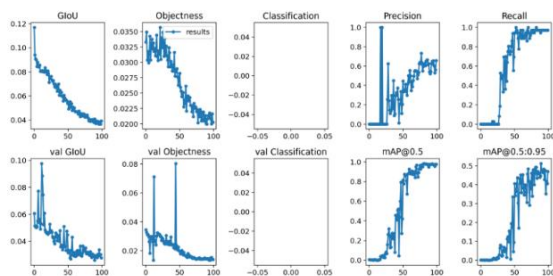


**Fig. 4.** Image after the detection of license plate using YOLOv5 Model

Proposed System has achieved the following results after training over 100 epochs.

Precision	0.629
Recall	0.943
map@0.5	0.0958
map@0.5	94% confidence as 0.417

**Table 1.** Training results of proposed system



**Fig. 5.** Accuracy of detection model while training

### 3.2.2 Stage 2: Plate Segmentation

Plate Segmentation: Segmentation is the next important step in the identification of License Plate that separates the License Plate from the original image and then we apply OCR Tesseract on it for Character Recognition.



**Fig. 6.** Image of the cropped License Plate

### 3.2.3 Stage 3: Character Recognition

Character Recognition here is done using OCR Tesseract is the last step of LP recognition that identifies each character separated by a text format [1]. Optical Character Recognition (OCR) is a method that allows a program while otherwise human intervention to identify texts or words written in users' oral communication. Optical Character identification has grown to become the most successful application of the information in the field of pattern acquisition and AI.

As we can see in every stage, we have also tried for one of the regional language number plates in India i.e., for Marathi number plate which should be roughly of following showed format. We used Google Translate API to translate Marathi words to English after passed through the same YOLOv5 model. Though the accuracy we got it is not that great, but it can further improve as more dataset will be available in future.

Segmented License plate	Character Recognition of License plate
	MH02LA7975
	MH 20 EE 7598

**Table 2.** Output after performing Optical Character Recognition (OCR)

## 4 Conclusion

In our proposed program, we have a habit of providing the entire ANPR comprehensive learning program with free results. Our results show that the strategic approach exceeds existing and external strategies in complex data sets, consisting of LPs taken with a strong point where rational retention ends up in many controlled repositories.

Most cars have the same single license plates while bikes, buses, rickshaws, etc., there are usually double license plates. So, this makes LP detection almost impossible. So, to overcome this difficulty, we can create a database with various types of

images of all License Plate patterns that can be found in India.

We further wish to work towards creating a dataset containing different Indian number plate types like in different regional languages, different font size and license plate sizes and also there are four types of number plates having different colour indicating the purpose of the vehicle driven e.g., yellow background number plates are generally used by taxis while White background Number plates indicate that the vehicle is for personal use etc. Hence, we hope to create an ANPR model which can detect all these types of number plates with maximum accuracy and speed.

## References

- [1] Andrew S. Agbemenu, Jephthah Yankey, Ernest O. Addo, IJCA, *An Automatic Number Plate Recognition System using OpenCV and Tesseract OCR Engine*, **43**, (2018)
- [2] Annual Report 21-22, Government of India, Ministry of Road Transport & Highways.
- [3] Du S., Ibrahim M., Shehata M., Badawy W., *Automatic License Plate Recognition (ALPR): A State-of-the-Art Review. IEEE Transactions on Circuits and Systems for Video Technology*, **23**(2), 311–325 (Feb 2013)
- [4] Kashyap A., Suresh B., Patil A., Sharma S., & Jaiswal A., ICACCCN, *Automatic Number Plate Recognition*, 6 (2008)
- [5] K. Prema, P. Arivubrakan, V. Suganya, TJCME, *Automatic Number Plate Recognition Using Deep Learning*, **12**, 3 (2021)
- [6] M. M. Shidore, S. P. Narote, IJCSNS, *Number Plate Recognition for Indian Vehicles*, **11**, 4 (2011)
- [7] Naren Babu, R., Sowmya, V., & Soman, K. P., ICICICT, *Indian Car Number Plate Recognition using Deep Learning*, (2019)
- [8] P. Kulkarni, A. Khatri, P. Banga, K. Shah, ICR, *Automatic Number Plate Recognition (ANPR) System for Indian conditions*, 4 (2009)
- [9] Rule 50 & 51 of Central Motor Vehicles Rule 1989
- [10] S. Ghadage, S. Khedkar, IJERT, *A Review Paper on Automatic Number Plate Recognition System using Machine Learning Algorithms*, **8**, 4 (2019)
- [11] S. Silva, C. Jung, ECCV 2018, *License Plate Detection and Recognition in Unconstrained Scenarios*, 17 (2018)
- [12] S. Tiwari, N. Choudhary, IJEAT, *An Efficient Method for Indian Number Plate Recognition*, **7**, 5 (2017)
- [13] S. Tripathi, S. Shetty, S. Jain, V. Sharma, IJITEE, *Automatic Number Plate Recognition System (ANPR): The Implementation*, 10, 8 (2021)
- [14] W. Devapriya, C. Nelson Kennedy Babu, T. Srihari, IJCIE, *Indian License Plate Detection and Recognition Using Morphological Operation and Template Matching*, 9, 7 (2015)