ABSTRACT

In this era of fast-growing technologies, there is a huge demand among the people for a secure lifestyle and travelling. In the past decade, the number of vehicles on road has been increased. Tracking of individual vehicle becomes a very challenging task with the massive growth in the vehicular sector every day. This paper suggests an automated vehicle tracking system for the fast-moving vehicles with the help of the surveillance cameras on the roadside. The process of getting CCTV footage in the real time background is very tedious process. To cater to this problem, an efficient deep learning model such as You Only Look Once (YOLO) is used for object detection.

The proposed work consists of four main steps. In the first step, video footage is converted into images and the car is detected from each of the frames. In the next step, license plate is detected from the detected cars. In the final step, the number plate characters reading are recognized from the detected number plates. The proposed deep learning model uses Image AI library to make the training process easier.

Tamil Nadu license plate images are used to analyses the performance of the model. The accuracy of 97% is achieved for car detection, accuracy of 98% is achieved for number plate localization and accuracy of 90% achieved for character recognition.

TABLES AND CONTENTS

LIST OF FIGURES

LIST OF ABBREVATIONS

ABBREVATION	EXPANSION
ANPR	Automatic Vehicle Parking System

LPR	License Plate Recognition
ANN	Artificial Neural Network
YOLO	You Only Look Once
OCR	Optical Character Recognition

CHAPTER 1

INTRODUCTION

In last few years, ANPR or license plate recognition (LPR) has been one of the useful approaches for vehicle surveillance. It is can be applied at number of public places for fulfilling some of the purposes like traffic safety enforcement, automatic toll text collection, car park system and Automatic vehicle parking system. ANPR algorithms are generally divided in four steps: (1) Vehicle image capture (2) Number plate detection (3) Character segmentation and (4) Character recognition. The first step i.e., to capture image of vehicle looks very easy but it is quite exigent task as it is very difficult to capture image of moving vehicle in real time in such a manner that none of the component of vehicle especially the vehicle number plate should be missed. Presently number plate detection and recognition processing time is less than 50 ms in many systems. The success of fourth step depends on how second and third step are able to locate vehicle number plate and separate each character. These systems follow different approaches to locate vehicle number plate from vehicle and then to extract vehicle number from that image.

Most of the ANPR systems are based on common approaches like artificial neural network (ANN) , Probabilistic neural network (PNN), Optical Character Recognition (OCR), Feature salient, MATLAB, Configurable method, Sliding concentrating window (SCW), BP neural network, support vector machine(SVM), inductive learning, region based, color segmentation, fuzzy based algorithm, scale invariant feature transform (SIFT), trichromatic imaging, Least Square Method(LSM), online license plate matching based on weighted edit distance and color-discrete characteristics. A case study of license plate reader (LPR) is well explained in. Some authors focus on improving resolution of the low-resolution image by using technique called super resolution. Sometimes it becomes necessary to assess the quality of ANPR system. In a quality assessment of visual and ANPR is well explained. A comprehensive study of License plate recognition (LPR) is well presented in. Throughout this literature, number plate and license plate are used interchangeably.

MACHINE LEARNING

7 Steps of Machine Learning

1) Step 1: Gathering Data.
2) Step 2: Preparing that Data.
3) Step 3: Choosing a Model.
4) Step 4: Training.
5) Step 5: Evaluation.
6) Step 6: Hyper parameter Tuning.
7) Step 7: Prediction.

Introduction:

In this blog, we will discuss the workflow of a Machine learning project this includes all the steps required to build the proper machine learning project from scratch.

We will also go over data pre-processing, data cleaning, feature exploration and feature engineering and show the impact that it has on Machine Learning Model Performance. We will also cover a couple of the pre-modelling steps that can help to improve the model performance.

Python Libraries that would be need to achieve the task:

1) Numpy
2) Pandas
3) Sci-kit Learn
4) Matplotlib

Understanding the machine learning workflow

We can define the machine learning workflow in 3 stages.

1) Gathering data
2) Data pre-processing
3) Researching the model that will be best for the type of data
4) Training and testing the model
5) Evaluation

The machine learning model is nothing but a piece of code; an engineer or data scientist makes it smart through training with data. So, if you give garbage to the model, you will get garbage in return, i.e., the trained model will provide false or wrong prediction

1. Gathering Data

The process of gathering data depends on the type of project we desire to make, if we want to make an ML project that uses real-time data, then we can build an IoT system that using different sensors data. The data set can be collected from various sources such as a file, database, sensor and many other such sources but the collected data cannot be used directly for performing the analysis process as there might be a lot of missing data, extremely large values, unorganized text data or noisy data. Therefore, to solve this problem Data Preparation is done.

We can also use some free data sets which are present on the internet. **Kaggle** and **UCI Machine learning Repository** are the repositories that are used the most for making Machine learning models. Kaggle is one of the most visited websites that is used for practicing machine learning algorithms, they also host competitions in which people can participate and get to test their knowledge of machine learning.

2. Data pre-processing

Data pre-processing is one of the most important steps in machine learning. It is the most important step that helps in building machine learning models more accurately. In machine learning, there is an 80/20 rule. Every data scientist should spend 80% time for data per-processing and 20% time to actually perform the analysis.

Data pre-processing is a process of cleaning the raw data i.e. the data is collected in the real world and is converted to a clean data set. In other words, whenever the data is gathered from different sources it is collected in a raw format and this data isn't feasible for the analysis. Therefore, certain steps are executed to convert the data into a small clean data set, this part of the process is called as data pre-processing.

As we know that data pre-processing is a process of cleaning the raw data into clean data, so that can be used to train the model. So, we definitely need data pre-processing to achieve good results from the applied model in machine learning and deep learning projects. Most of the real-world data is messy, some of these types of data are:

1) Missing data: Missing data can be found when it is not continuously created or due to technical issues in the application (IOT system).

2) Noisy data: This type of data is also called outliners; this can occur due to human errors (human manually gathering the data) or some technical problem of the device at the time of collection of data.

3) Inconsistent data: This type of data might be collected due to human errors (mistakes with the name or values) or duplication of data.

Three Types of Data

1) Numeric e.g., income, age

2) Categorical e.g., gender, nationality

3) Ordinal e.g., low/medium/high

These are some of the basic pre — processing techniques that can be used to convert raw data.

1) Conversion of data: As we know that Machine Learning models can only handle numeric features, hence categorical and ordinal data must be somehow converted into numeric features.

2) Ignoring the missing values: Whenever we encounter missing data in the data set then we can remove the row or column of data depending on our need. This method is known to be efficient but it shouldn't be performed if there are a lot of missing values in the dataset.

3) Filling the missing values: Whenever we encounter missing data in the data set then we can fill the missing data manually, most commonly the mean, median or highest frequency value is used.

4) Machine learning: If we have some missing data then we can predict what data shall be present at the empty position by using the existing data.

5) Outliers' detection: There are some error data that might be present in our data set that deviates drastically from other observations in a data set. [Example: human weight = 800 Kg; due to mistyping of extra 0]

Researching the model that will be best for the type of data

Our main goal is to train the best performing model possible, using the pre-processed data.

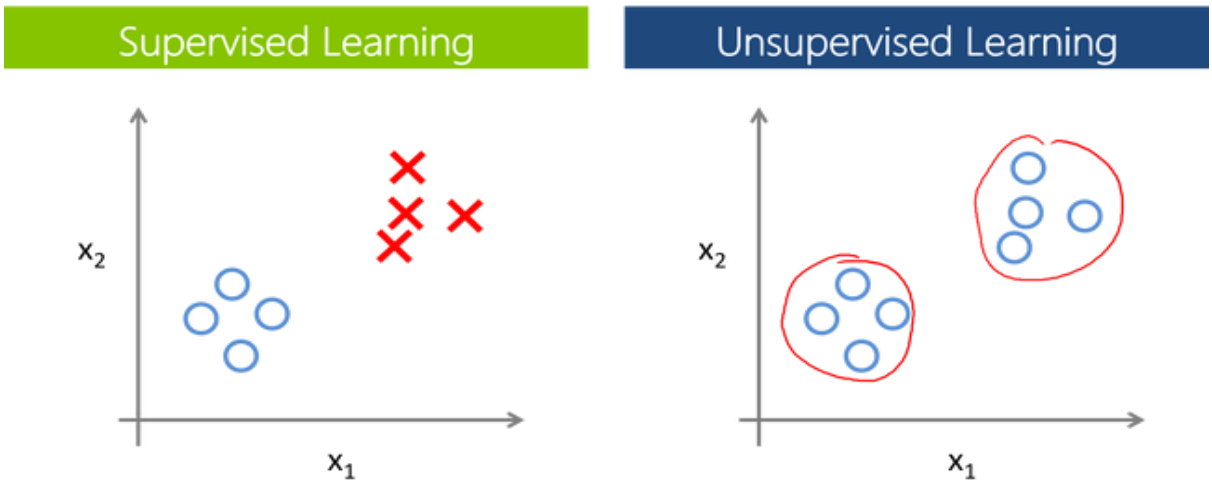

Fig 1.1 Supervised Learning and Unsupervised Learning

Supervised Learning:

In Supervised learning, an AI system is presented with data which is labelled, which means that each data tagged with the correct label.

The supervised learning is categorized into 2 other categories which are “Classification” and “Regression”.

Classification:

Classification problem is when the target variable is categorical (i.e., the output could be classified into classes — it belongs to either Class A or B or something else).

A classification problem is when the output variable is a category, such as “red” or “blue”, “disease” or “no disease” or “spam” or “not spam”.

Fig 1.2 Classification

As shown in the above representation, we have 2 classes which are plotted on the graph i.e. red and blue which can be represented as ‘setosa flower’ and ‘versicolor flower’, we can image the X-axis as there ‘Sepal Width’ and the Y-axis as the ‘Sepal Length’, so we try to create the best fit line that separates both classes of flowers.

These some most used classification algorithms.

- K-Nearest Neighbor
- Naive Bayes
- Decision Trees/Random Forest
- Support Vector Machine
- Logistic Regression

Regression:

While a Regression problem is when the target variable is continuous (i.e. the output is numeric).