mirna-ALPR

by MeRna 190860

Submission date: 08-Jun-2023 05:36PM (UTC+0200)
Submission ID: 2090052164
File name: mirna190860-ALPR-system1.pdf (1.72M)
Word count: 13602
Character count: 73479

Automatic License Plate Recognition Using Deep-learning

Submitted To: Dr. Sherif Hussein

Submitted By: Mirna Victor 190860

Table of Contents

List of figures:

Abstract:

Artificial intelligence (AI) has several subfields, including computer vision, which can extract information from digital images, videos, and other visual inputs and then carry out activities or offer options based on that information. Computers can see and think thanks to computer vision and artificial intelligence (AI). Depending on people's needs, computer vision is used to read license plates and perform other tasks such as parking management and traffic enforcement. Because of huge growth of manufacturing vehicles, high ways and some other areas became crowded. Hence, it is not easy to check license plates manually. Automatic License Plate Recognition (ALPR) system will make those processes easier with high accuracy. Several researches have proposed different ALPR system. However, they still have some obstacles such as license plates diversity and illumination conditions. Moreover, detection and recognition accuracy need to be improved. The goal of this Project is to study different techniques that are used in ALPR system. The literature analysis provides a detailed study of the three main tasks of the ALPR system which are: vehicle license plate detection, segmentation and recognition techniques. It explains how these strategies have been tested by researchers, what techniques are being utilized or developed, what character types have been identified, what datasets have been referred to, and how much progress has been achieved and comparison between their results. and finally, Because there aren't enough datasets, especially in the Arabic plates, many ALPR approaches have been proposed that are based on classic image processing and machine learning algorithms. This project will provide a real-time ALPR system for the Egyptian license plate (LP) detection using different YOLO Algorithm versions which are: YOLOv5, YOLOv7, and YOLOv8. And using CNN model to implement the recognition phase.

Chapter 1:
Introduction:

The technology of nowadays is an important problem solver in people's life. Artificial intelligence is one of the driving technologies that is shaping people's lives. AI has a lot of applications that can provide a safe life for citizens. One of these applications is Automatic License Plate Recognition (ALPR). By increasing the number of vehicles in high ways leads to failure to control the vehicles' speed and racking them manually is practically not feasible specially Crowded places, parking areas in malls, toll gates or any other places that require checking license plates to let cars enter or leave. Criminal cases might increase without even know any information about the crime. ALPR system can solve all these problems by applying its application.

ALPR has as an efficient measure with a wide range of applications, vehicle surveillance has emerged as an invaluable asset. Its utility encompasses diverse public locations for different goals like managing traffic flow, automating tolls collection and overseeing parking areas. Fig 1. shows that ALPR system is divided into four main tasks. The first task is to capture the vehicle image, it seems to be an easy step but it is a little bit challenging task because it is incredibly challenging to take a real-time picture of a fast-moving vehicle. detection of the license plate is the second task (Plate Localization), the localization task is implemented by using traditional computer vision techniques that use the information of each images to locate the license plate or using advanced techniques using the algorithms of AI to detect the license plate automatically. Third task is character segmentation, each character will be segmented to prepare this data for the final task which is character recognition.

This project will survey previously proposed ALPR systems. Comparison between such techniques will be conducted. Moreover, various datasets used in such techniques will be examined.

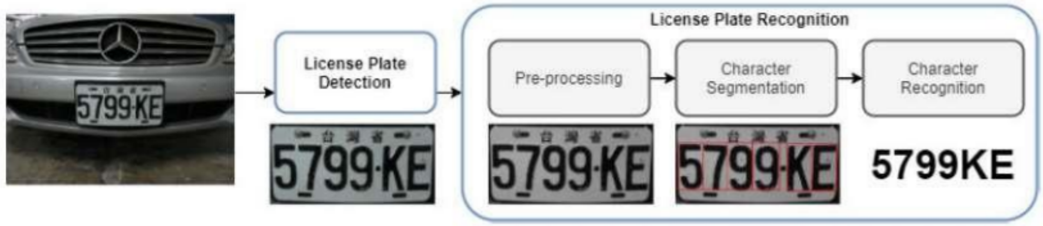


Figure 1

Figure 1: shows the four main tasks of the ALPR system.

1.2 Problem Statement:

License plate detection and recognition are made difficult by environmental factors or changesin plate types. Such observed variances are sunnitized as follows:

• **Variations in the environment:**

1- Illumination: Depending on the lighting conditions and the type of vehicle being driven, input photos may have various forms of illumination.
2- Background: The image may include patterns like vehicle license plates or other printed symbols, vertical patterns on a front, or textured flooring.

• **Differentiated Car Plates:**

1. Location: Different areas of a picture that include the plates.
2- Quantity: A plate might be one or several in a picture.
3. Size: Depending on the camera's distance and magnification level, plates may range in size.
4- Color: Character and background colors on plates might vary depending on the type of plate or the capturing technology used.
5- Font: Various languages and fonts may be used to write on plates from different countries.
6- Occlusion: Dust and other particles may cover plates.
7- Plates may be inclined.
8- Other: A plate might also have frames and screws in addition to characters.

1.3 Research objective:

The goal of this project is to present a detailed background on each task in ALPR system. Then, examine all of the different techniques used to put the ALPR system into practice and how the proposed systems solved the problems mentioned in the problem statement section. And to compare the results of these methods across various datasets while also surveying a variety of possible algorithms to address each of the difficulties that the ALPR system faced. Finally, the system will be divided into two main sections. Section number (1) is: License Plate Detection and it will be implemented using Different versions of YOLO Algorithm which are YOLO V5, YOLO V7, and YOLO V8. Then comparison between results and section number (2) is: Character Recognition using CNN model. The system will use the Egyptian vehicles dataset.

.

Chapter 2: Literature Survey:

2.1 Introduction:

Automatic License Plate Recognition (ALPR) finds multifarious usage scenarios such as traffic auditing/control methods' optimization; parking facilities efficiency enhancement; seamless toll collection procedures; secure law enforcement surveillance techniques among others through automated license plates identification practices initiated by these setups. ALPR undertakes multi-faceted objectives where accurately detecting an automobile's number plate through advanced technology followed up with efficient identification abilities without requiring much human intervention becomes crucial. These processes entail three primary steps; License Plate (LP) detection, character segmentation, and character recognition, where the initial phases must produce precise or near-perfect output for the latter stages to succeed. As such, to reduce processing time while ensuring accuracy and eliminating false positives in license plate detection procedures, ALPR systems mainly utilize the two-stage strategy of scanning vehicle images first before proceeding to their unique ID recognition.

License plates are used to identify vehicles. The owner's information is linked to these license plates. The use of license plates, Intelligent transport systems (ITS) is able to help the system in identifying vehicles for interested parties, such as law enforcement or insurance firms investigating a claim. In order for an ITS to fulfill its intended purpose effectively, it must possess the necessary capability to accurately read vehicle license plates. Despite a strong positive correlation between shooting distance and proportion of the license plate captured by an image within ANPR architecture, achieving optimal balance remains challenging at times. Additionally, owing to limitations in capturing large-area views of vehicles in motion that offer clear visibility as well as identifiable license plate details without distortion or blurriness can be extremely difficult. To tackle this obstacle head-on requires substantial efforts aimed at identifying a tiny yet distorted image accurately- especially so when it comes to recognizing the numbers on a number plate accurately. One possible solution entails deploying CCD cameras equipped with PTZ capture functions that enable them to move around while focusing on various angles simultaneously. Multiple approaches have been suggested for executing different phases as part of identifying number plates through ANPR architecture- each with distinctive advantages and disadvantages. The three primary processes involved herein include extracting number plates from images captured by sensors mounted on cameras; recognizing individual characters correctly; and processing all critical data gathered during this entire sequence of steps carefully.

Environmental conditions or changes in plate types make license plate detection and recognition more challenging. The following are some examples of reported variations: changes to the environment, such as illumination and another additional kind of challenging Differentiated Car Plates: Location, Quantity of plates, Size, and Colors: Character and background colors. ALPR system is generally consisting of three main tasks: Detection, Segmentation, And Recognition. This chapter discusses the various techniques to implement each stage of thesystem briefly and compare between them.

2.2Background:

2.2.1 Task 1: License Plate Detection Methods:

ALPR needs to extract license plate from the image of the vehicle. This task is very important and critical. However, some problems face the detection stage such as, license plate may be broken, dirty or located out of sight for ANPR camera. Moreover, some problems are related to the environment conditions such as, illumination and fog. Hence, different algorithms come up with solutions of those problems. Figure 2 shows various techniques used to detect the license plate. They are generally divided into traditional computer vision techniques and advanced techniques.

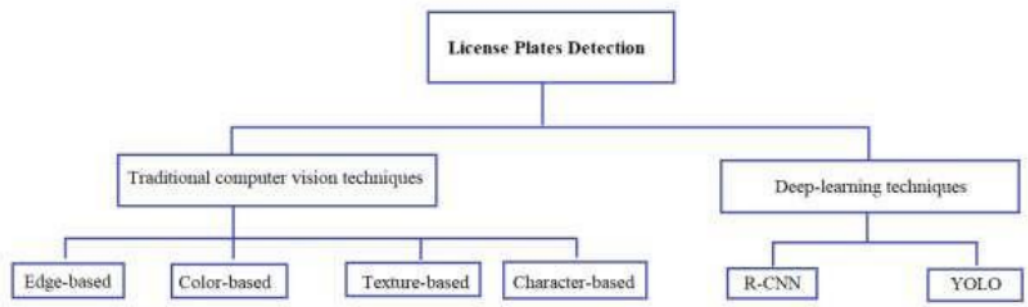


Figure 2

Figure 2: The algorithms of detection task.

2.2.1.1 Traditional techniques:

A. License Plate Detection Using Edges information:

Assuming that, the license plates of every vehicle is a rectangle with knowing their aspect ratio. The license plates are extracted using edge detection techniques, which collect all of the rectangular edges in the image. In most of vehicles, the body and the license plates have differentcolors. Hence, the boundaries of the number plate became an edge. There are two types of edges,vertical and horizontal. The lines of the rectangle are used to get the required edges. There are a lot of algorithms and filters used to detect the edges. Such as, Canny edge detection, Gabor filter,Sobel and Hough transformation. Those algorithms are explained in detail, In [1].

B. License Plate Detection Using Character features:

Character is a piece of information used in image processing and computer vision that describesthe features of a picture. It frequently relates to whether a certain area of the image. Specific components in a picture, such as edges, points, or objects, are known as features. This detection technique works by checking the whole image to find characters and identifies them. This method is done by using the technique called "Character-like region" [2]. Then using neural network classifier. For example, Matas and Zimmermann [3] proposed and approach that extracts every region in an image that resembles a character has been suggested. Then the neural network classifier is utilized to categorize the extreme locations. Second, we take it for granted that any linear spatial arrangement that we find is indicative of a possible license plate position. With an outstanding detection accuracy of 95%, this method has shown that it is resilient when dealing with a wide variety of lighting situations and angles of view.

C. License Plate Detection Using Texture-Based Methods:

Plate detection using texture-based approaches is dependent on the presence of characters that are printed on the license plate. Because of the striking contrast in color between the characters on the license plate and the plate itself, the characters on the plate regularly switch colors. Therefore, if the image is grey-scaled, the text and the platebackground stand out clearly. The plate region's surrounding pixels are thus distributed in a distinctive way. Additionally, the color transition creates a high edge density in the plate region. The Gabor filter is a common tool in texture analysis. The capacity to examine texture in limitlessdimensions and scales is a significant benefit of utilizing a Gabor filter [4].

Texture-based approaches are resistant to number plate deformation. On the other hand, this approach requires more time and has a lower success rate. As it takes comprehensive examination. In addition to this, it has a poor performance in a wide variety of lighting conditions and settings. Cho and others [5] The breadth of the characters as well as the contrast between the characters and the background of the license plate are utilized to determine which region the characters are located in. They were able to extract the specific plate region by making use of the inter-character distance, and they recorded a detection rate of 99.5%. However, using these methods requires a significant amount of effort and frequently results in errors, particularly if the input image contains extra text.

D. License Plate Detection Using Color Features:

Some nations or areas may have color-specific license plates for vehicles. Color-based license plate extraction from the vehicle image is also put to the test for ANPR systems. The concept of color coordination of the license plates is used in the general approach to license plate extraction.Additionally, outside of the plate zone, the image does not contain the color scheme of the plate and its text. The Hue, Lightness, and Saturation (HLS) color model is a useful tool for organizing the pixels that make up an input image in accordance with the various sources of illumination.. The (RGB) model does not classify pixels into 13 groups, as does the HLS model. Applying the color information-based number plate extraction approach has a benefit. In other words, you get to choose the license plates that are crooked or distorted. It also has some disadvantages, though. Using the RGB valueto describe a pixel's color might be difficult, especially in certain lighting situations. The HLS color model [6], which is employed in place of it, is very sensitive to noise. The flaw of incorrectly detecting picture components that have the same color on the license plate as the vehicle body is addressed by color projection-based methods. Grayscale was substituted with a color model.

2.2.1.2 License Plate Detection Using Advanced Techniques (Deep learning):

A. Convolution Neural Networks:

CNN are very popular when it comes to processing images and videos. CNN models recognize the contents of an images and videos. Image localization is a branch off of standard CNN vision algorithms. These methods predict discrete number classes. The object localization approach predicts a collection of 4 continuous integers, including the x coordinate, y coordinate, height, and width, to create a bounding box around an item of interest.Depending on the application, amount of data, and processing power at hand, the first convolutional neural network layers in CNN-based classifiers might range from a few to 100 layers. A significant area of inquiry in CNN is the number of layers. The output layer, which comes last, calculates the likelihood that an object will appear in a picture [7].

Figure 3, assume, that an algorithm recognizes 100 distinct items and a general CNN architecturein each image. The likelihood that an object will appear in an image is then indicated by the 100-value array in the last layer, which ranges from 0 to 1. The output layer is the only difference in an image localization algorithm. A probability value between 0 and 1 is provided by the final layer in classification algorithms. As localization is a regression problem, localization algorithms, in contrast, produce an output in four real values. As was previously said, a box is drawn around the item using those four values. The majority of ANPR system studies currently in existence process an input image that includes a vehicle. The algorithms are concerned with locating and retrieving the vehicle's license plate from the image. Faster and R-CNN (region- based CNN) In different settings, the ANPR is located using R-CNN, YOLO (very efficient object identification framework), and SSD (single shot detectors).

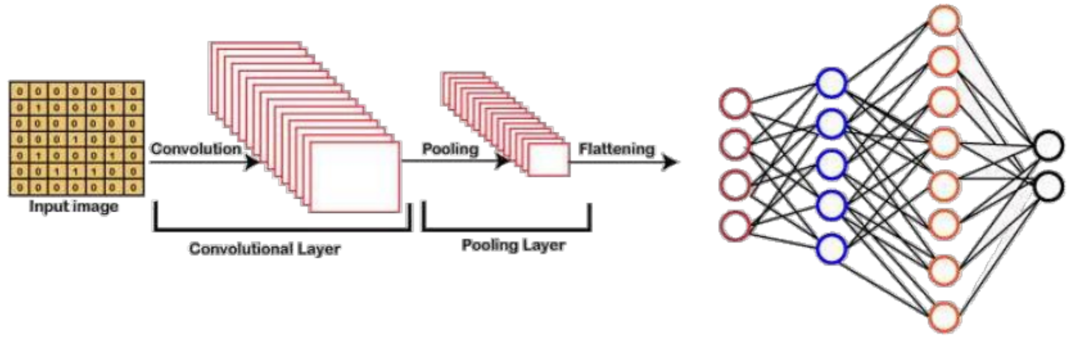


Figure 3

Figure 3: CNN architecture[7].

B. Detection Using YOLO Algorithms:

Before YOLO entered the scene of Deep Learning item recognition algorithms, many methods were published in order to identify the object from the image; nevertheless, YOLO took a completely different strategy than the other methods. When a classifier was developed once more for the purpose of acting as an object detector, it was not the same one that had been used before. The fact that YOLO was able to recognize the object after viewing it only once, as the name suggests, piqued the interest of a significant number of newly trained deep learning developers. In reality, it took a single, intelligent look. Redmonetal was the one who originally popularised the YOLO approach. (2016) Entitled "You Only Look Once: Unified, Real Time Detection." In addition to its straightforward architecture, his study demonstrates that YOLO accurately and quickly identifies objects, with an average accuracy of 88% from ImageNet 2012 validation. 2018 Rayson Laroca et al. We carried out a poll utilising the YOLO approach in order to identify and identify the licence plate. This study is titled "Robust real-time automatic license plate." This study used detection problems and videos to highlight the problem of video-based discovery. When compared to other traditional object detectors, the YOLO object detector is quicker and more accurate because it just has one stage of detection and was trained using the COCO dataset. Modern object detectors are slow when using them to the train models since, as we know, they need more GPUs for training and they train with enormous micro batches. Sometimes the training time is so long that using such models becomes impracticable.

By enabling object recognition with a single GPU and a reduced mini batched size, YOLOv4 put forth this problem. It makes training with a single 1080Ti or 2080Ti GPU extremely quick. Even though the algorithm is rather effective, we should still be careful to use better photos when working with nations like Egypt because this will aid our algorithm in some ways to increase the accuracy of the entire model[4].
Some suggestions for enhancing camera performance include paying attention to various camera settings, such as the iris size, shutter speed, angle at which the camera is pointed, and camera position. The type and size of the lens, as well as the lightning circumstances, are some other things we can concentrate on.

A deep learning approach is provided in the study [5] for identifying the category and location of solder junctions on vehicle door panels in real time.
Due to the diversity and complexity of work situations, the solder joints' locations in vehicle door panels change frequently. As a result, programming and instruction have to be repeated repeatedly by hand interference if the location of the solder connections is not exactly determined by automation. Due to this, the manufacturing line becomes less intelligent and automated, which has an impact on the productivity and quality of the welding. Due to the lower solder joint size on a car door, YOLOv1 and YOLOv2 are not appropriate. As a result, the YOLOv3 [8] method, which only changes from YOLOv2 in terms of the end result, is utilized to more accurately detect small solder joints. forecasts are made on feature maps of various sizes, and the results of these forecasts are integrated to produce the final product, using a variety of levels of prediction. There is only one object that can precisely identify solder junctions in each grid cell of 52 by 52 feature maps [6].

2.2.2 Segmentation:

Character segmentation comes before the classification step in many optical character recognition algorithms. Character segmentation can be conducted using various techniques. It considers the characters and background having contrasting colors in license plates. The background pixels and foreground (character) receive the opposite "colors" when the image is binarized, which facilitates this separation. Thresholding, region-based, edge detection-based, and morphological segmentation are examples of traditional computer vision-based image segmentation techniques. A digital image is divided into several parts by the technique of image segmentation (objects). By segmenting an image, a more meaningful and understandable representation of the image is created [8].
Neural networks can also be used in image segmentation. The following subsections describe each segmentation technique briefly

A. Segmentation using Thresholding:

Thresholding is the simplest method for segmenting photos. It involves grouping the image's pixel elements according to their intensity values. Typically, it is employed to create binary images by splitting an image into two halves. The key is to choose a threshold T, set the pixels with values below it to 0 (background pixel), and the pixels with values above it to 1 (foreground pixel). For uint8 data type photos, for instance, the most value that your image pixels may accept is 255, Since the highest value that the image's pixels may accept is 255, assigning 255 to the foreground and 0 to the background is also a viable option [8].

Thresholding selected a suitable threshold value to divide image pixels into various areas and distinguish between foreground and background. Several thresholding techniques have been created by various researchers, including OSTU thresholding, P-tile method, Histogram dependent methodology, Edge maximization technique, Mean method, and visual approach. Based on Salem Saleh Al-amri [9] edge maximization and histogram dependent methods are better than other techniques in thresholding.

B. Region-Based Segmentation:

Using region-based segmentation, a picture is cut up into different regions that all have similar characteristics. Each region is made up of a group of pixels that the algorithm discovers by utilizing a seed point as a starting point. After locating the seed points, the approach can either increase the size of regions by adding more pixels or reduce the size of regions and merge them with other points [10]. Typically, some pixels are designated as seed pixels at the beginning of the segmentation process. The algorithm then detects the seed pixels' near boundaries and categorizes them as similar or different. The subsequent processes are continueduntil the entire image has been segmented, using the nearby areas as seeds. The well-known watershed approach for segmentation is an illustration of a related technique [11].

Expanding a region, separating a region, and merging two or more regions together are the three main components that make up region-based segmentation systems [10]. A sequential method that is based on areas and is dubbed "region growth" scans the surrounding pixels and adds them to larger regions based on "seed pixels," "growing criteria," and "stop conditions" that have been specified. The process of breaking a single image into multiple distinct and unified pieces is known as area splitting. Following each split, a region merging process is put into place in order to examine the regions that are immediately adjacent to one another and, if necessary, combine them.

C. Edge-Based Segmentation:

Edge-based segmentation is a widely used method for processing photographs that recognizes the edges of various objects depicted within an image by leveraging information from each object’s respective edges. Doing so enables detection of related attributes across diverse areas within an original photograph while simultaneously decreasing overall image size and “cleaning” superfluous data-allowing easier interpretation for analysis purposes.

The complex network of edge recognition involved with this segmentation approach employs cutting-edge algorithms including ones that distinguish variance in contrast/texture/color saturation/etc., thereby yielding highly accurate boundaries between each object identified in a photo that strategically construct chains through their differentiating characteristics notable enough for attributional differentiation.

Though achieving this effect requires several varied specializations including graphical histogram/grid measures (gray) or quantitative gradients calculated mathematically with determinate operators such as Roberts Laplacian/Sobel/Prewitt/LoG + modern tech adaptations like Canny or color enhancers.

D. Image segmentation based on clustering:

Starting with the concept of Clustering before moving on to Image Segmentation with Clustering. Clustering is comparing the similarities between the provided data and creating distinct clusters as a result. The way we group the data varies depending on the clustering technique that should be chosen [14]. Clustering methods, a type of unsupervised classification technique, are used to uncover hidden information in images. They aid in the improvement of human vision by emphasizing clusters, shadings, and structures. The process divides pictures into groups of pixels with related characteristics and separates data into discrete chunks and clusters. [10]. Clustering is a potent picture segmentation technique. Numerous clustering techniques exist, such as improved fuzzy c mean algorithm, adaptive k means, and k means (IFCM). The K-Means clustering technique is an unsupervised method for separating the desired area from the background. The data is clustered or separated into K-clusters.

E. Segmentation Using Semantic Technique:

During the semantic segmentation process, the pixels in an image are broken up into different semantic classes. The segmentation model does not make reference to any other context or data, and the classification of every pixel in this model is limited to a single category. When semantic segmentation is applied to an image that contains several trees and vehicles, for instance, a mask that unites all tree types into a single class (tree) and all vehicle kinds—including buses, automobiles, and bicycles—into a single class (vehicles) will be created. This mask will unify the many types of trees and vehicles into a single class. When applying this method, the problem description is frequently vague, particularly when multiple instances are merged into a single class. This is because this method combines them all into one.

For instance, the entire crowd captured in a photograph of a busy street could be filed away in the "people" category of the photo collection. For images with such a high level of complexity, semantic segmentation is unable to offer information that is semantically specific.[10] The inputs that semantic segmentation models are given ultimately result in the generation of their outputs, which are segment maps. The number of classes that the model is supposed to segment is denoted by the letter n in these segment maps. These n-channels are all binary in their underlying structure, and the item placements are "filled" with ones where there should be empty spaces and zeros where there should be empty spaces. The ground truth map is a single channel integer array with a range of "n" segments that are "filled" with the relevant class index values (classes are indexed from 0 to n-1). These segment ranges are "filled" with the appropriate class index values. The "n-channel" binary output of the model can also be referred to by its alternative nomenclature, which is a two-dimensional one-hot encoded representation of the predictions. For segmentation, neural networks typically use an encoder, a bottleneck, and a decoder or up sampling layer that begins at the bottleneck (as in the FCN).

The model's "n-channel" binary output is also known as a two-dimensional one-hot encoded representation of the predictions. A decoder or up sampling layer that starts at the bottleneck is usually used in neural networks for segmentation (as in the FCN).

2.2.3 Character Recognition:

The last step, character recognition, which is responsible of recognizing the license plate after the stages of detection and segmentation of the plates. At this point, text is created using characters that are inserted in photos of license plates. This is a particular instance of optical character recognition that takes into account specific elements of the license plate. For example, the font and color of the license plate are subject to tight regulations in many nations, and they are often chosen to be legible. The license plates do come with a few special problems, though. For instance, because the photograph was shot outside, the system designers had to take the effects of the weather, uneven brightness, and fluctuating ambient light into account. They may still be rotated or damaged even though they have a regular license plate.

There are several categorization techniques that need learning models with fixed-size inputs. Before classifying the input segments, a rescaling is performed since the result of the segmentation stage has a broad range in size. Since the quantity of characters, their placement in relation to one another, and their potential values are frequently known, each segment is designated individually as belonging to one of the possible values. Three different circumstances can be considered when doing this:

A. Direct comparison of all raw image data pixel values with predetermined templates
B. Feature extraction by various image processing and machine learning methods before segment classification.
C. Deep learning techniques.

Those techniques will be explained in the following sections.

2.2.3.1 Recognition using Template Matching Techniques:

The easiest way for character recognition is the Template Matching Technique, which is used to distinguish segmented characters. It uses a cross-correlation method to determine how closely the extracted character resembles the template characters. These techniques are frequently utilized for binary pictures due to the direct effects that changes in lighting conditions have on the intensities of the different gray levels in the final image. For the recognition of non-rotated, shape-invariant characters, the pattern matching method is appropriate. In order to determine how well normalized characters and templates match, P. Comelli et al. employed cross correlation in [15]. Italien vehicles traveling through tollgates are subject to this algorithm's implementation. Single prototypes for character strings are created as templates for two-letter provinces. The algorithm successfully identified nearly 91% of the more than 3,000 genuine photographs it evaluated, which were taken in various weather and lighting conditions. Using an essential elements-based strategy, authors in [16] created an OCR system for Thai automobile license plate identification, which are often shaped complexly. On the basis of the fundamental components of characters, this algorithm identifies character patterns.

According to [17], Character matching depends on edge Hausdorff Distance (HD), sometimes referred to as max-min distance. It is a statistic for contrasting two groups of random points. The method almost exactly matches the recognition rate of neural network classifiers, but is a bit slower. It also has all the mathematical features of a metric. The character with the greatest correlation value is the most suitable match. However, identification using Template Matching is only effective provided the characters are intact, not skewed, have constant font sizes, and have not been altered. [18]

2.2.3.2 Recognition Deep Learning Techniques:

The benefit of using neural networks is that they may be directly fed the raw pixel data and perform the dual functions of feature extraction and classification on their own. CNN has been employed in numerous recent research [19] and has demonstrated significant promise in a variety of computer vision applications. Utilizing object detection-based approaches directly, such as YOLO, is another emerging strategy. Although deep learning-based systems need more computer horsepower than competing techniques Similar to template matching and statistical feature extractors, they frequently offer more accuracy.

A. Convolutional Neural Networks (CNNs):

The characters that appear on the segmented license plates are well-known to CNN (LPs). Figure 8 depicts the layers that make up CNN, which include convolution, pooling, and fully connected (FC) layers [20]. In [21], Mondal put out a convolutional neural network (CNN)-based system for automatically recognizing license plates that is focused on characteristics that are self-generated. CNN's self-synthesized function can determine the vehicle's condition from the number plate. They have demonstrated that their technology is reliable and efficient by showing
that more than 90% of the photos correctly identify the vehicle's license plate.

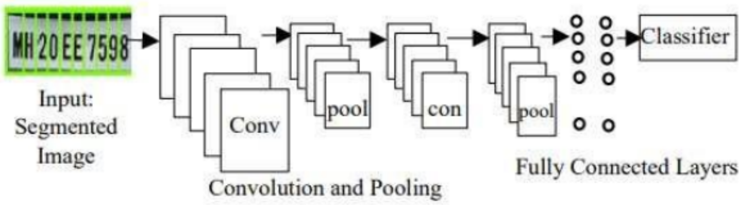


Figure 4

Fig 8. The design of CNN for character recognition [21]

Furthermore, CNN has successfully trained using the SIFT (Scale Invariant Feature Transform) feature. A feature identification method called SIFT aids in identifying the little details present in a picture. Lastly, a two-phase verification technique has been implemented. The first phase is a statistical filter in the LPD phase to effectively remove the wrong plates, and the second phase is a pipeline shortening that improves the performance of the LPDR system. The accuracy rate was roughly 84.3%, and they were able to prove that the anticipated approach does, in fact, recognize the car license plate in real time. The suggested system's recognition results are shown in Figure 9.

Figure 9: Vehicle number plate recognition results using dataset [22]

Figure 5

Zhu built a system in [22] that focuses on vehicle monitoring and data integration for the recognition of license plates on moving vehicles on city streets. Each vehicle's license plate from the video series is recognized using an object detection framework that has been trained and is centered on a plate detector. For the purpose of recognizing vehicles license plates , convolutional neural networks (CNN) have been used. Additionally, to get the outcome, the continuous frames incorporate recognition effects. Figure 10 displays the suggested LPR system configuration that combines vehicle tracking and result integration. According to their claims, in the actual conditions of urban roads, They had a recall rate of 89% and an accuracy rate of 82.5% for identifying license plates.

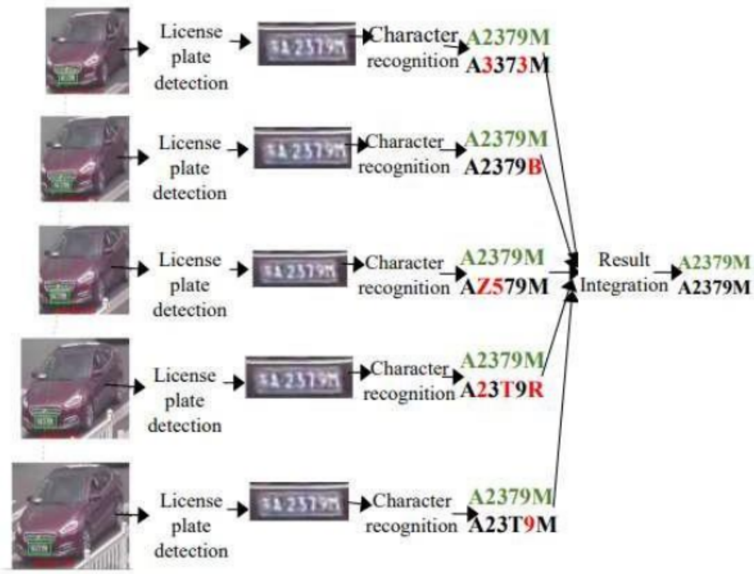


Figure 10: Structure of the LPR system centered on vehicle tracking [22]

Figure 6

B. Recognition Using YOLO Algorithms (You Only Look Once):

One of the most significant and successful object identification algorithms is YOLO, which incorporates a number of ground-breaking concepts from the scientific community in computer vision [23]. To find the item within the image, all of the prior object detection techniques used areas.

Region-based algorithms are very different from YOLO. In YOLO, a single neural network expect the bounding boxes as well as these boxes' class probabilities. YOLO can process 45 frames per second, which makes it quicker than conventional object detection methods. However, the YOLO method is constrained by the fact that it can only handle little items in the image. A reliable and efficient YOLO object detector-based ALPR system has been developed by Laroca in [24]. They applied fundamental data enhancement techniques to an inverted License Plates system for character identification and segmentation. At this time, it was determined if the Fast-YOLO and YOLOv2 models could handle both more straightforward data. Fast-YOLO should be able to accurately identify vehicles and their license plates in less time in basic scenarios. Important outcomes in two datasets have also been produced by the ensuing ALPR method. They claimed that their technology has a 93.53% recognition rate.

In [25], Laroca proposes an effective and efficient YOLO object detector-based layout-independent ALPR framework that incorporates a cogent method for detection the license plates and layout categorization (LP). This is due to the fact that YOLOv3 and FastYOLOv3 perform pretty well on little items but poorly on medium- and large-sized objects. They employed eight open datasets, and To the datasets, they applied many data enhancement methods. They confirmed that the proposed method had a 96.8% overall identification rate on the datasets. The pre-identification plate and most recent YOLO-L model were used by Min to construct a clever technique for finding car license plates in [26]. To precisely determine the location of the license plate, the suggested model alters two components. The size and quantity of the candidate boxes for a plate were initially determined using the k-means++ clustering technique. The YOLOv2 network model and depth were modified after that. They also employed a plate pre-identification technique to distinguish license plates from associated objects. They said that the suggested strategy produced precision and recall of 98.86% and 98.86%, respectively. Henry has proposed a universal system for recognizing license plates on moving vehicles in [27]

Related work:

According to Cheng-Hung Lin et al. (2019) [28], a three-stage license plate identification method based on Mask-RCNN was suggested. This system was employed for several oblique photos and different shooting angles. When detecting vehicles in the earlier stage, the author employed YOLOv2 to identify the linked conveyance. Following that, YOLOv2 was once more used to determine the license plate location. The photographs of the phase I collected cars are divided into 19 × 19 grids during this phase using YOLOv2. The author recognized characters in the last phase by employing Mask R-CNN. The study's findings show that the suggested model can categorize car license plates with bevel angles greater than 0 to 60 degrees and achieve a mAP rating of about 91%.

A technique to identify the Bangla license plate from a car image has been proposed by Nazmus Saif et al. (2019)[29]; utilizing convolutional neural networks. Due to its setup for an end-to-end pipeline, the convolution neural network was chosen as the primary emphasis in this work. For their use case, Traditional image processing methods were decisively outperformed by CNN, and generalized CNN models also outperformed in other situations. The detection experiment made use of the YOLOv3's 53 convolutional model layers. Image segmentation and character recognition make up the second stage following character identification. For segmentation and platform image identification, At this point, the device removes and transfers the number plate region to the second YOLO model. In tests utilizing 200 images, the model correctly recognized the LP number in 199 of them, or 99.5% of the time.

Digital image processing methods and neural networks are used in (2020) [30]. The suggested approach combines preprocessing and recognition phases. The preprocessing phase includes: Binarization, picture quality improvement, and image segmentation into smaller images are used to find the license plate area. The segmented sub-images will be recognized and classified as numbers and characters in the recognition step. Using the standard cross correlation approach, localization is carried out in this study. According to specified algorithms, segmentation comprises splitting the linked characters into independent characters, dividing the separated characters into individual numbers, and segmenting the license plate into three parts. Back propagation neural networks are used to recognize objects (BPNN). Two pieces of data are used by the recognizer to work. Whole pixels from the sub-images are included in the first batch of data. 16 characteristics that were taken from the sub-images make up the second batch of data. These two approaches are contrasted with one another. The environment work is done with the MATLAB software, and the system has experience with 99 photos of the provinces of Duhok and Erbil. For segmentation, localization, and recognition, the accuracy percentages are 100%, 100%, and 100%, respectively. The first method's recognition percentage is 94.5%, while the second method's is 91%. Fig 11 shows example of its dataset.

Figure 7

Fig 7. An Arabic license plate recognition.[30]

According to (2021)[31] The study implemented three-step approach that will identify the characters on the LP, segment the characters, and detect the license plate. The input picture is transformed into a bi-level image to do detection. The characters are separated from the observed license plate using region props. For the purpose of recognizing the segmented characters, a two-layer CNN model is created. The database is automatically updated by the suggested model each time an automobile enters or exits a parking space. The suggested ALPDR model has been evaluated under a variety of circumstances, including fuzzy pictures, varying camera distances, day and night settings, and stationary cars. Experimental results showed that the project improved state-of-the-art literature models in terms of license plate detection, segmentation, and recognition accuracy, achieving 91.1%, 96.7%, and 98.8% accuracy, respectively.

As stated in (2021) [32].The project proposed vehicle tracking system that uses roadside security cameras to track fast-moving vehicles. It is a rather challenging process to obtain background real-time CCTV footage. You Only Look Once, a powerful deep learning model (YOLO), is utilized for object identification to solve the problem. There are four basic phases in the entire project. The vehicles are identified from each frame of the video in the first stage, which involves turning the video material into pictures. The next phase is the detection of license plates from the identified vehicles. The license plate characters are identified from the observed license plates in the last phase.
The ImageAI library is used by the suggested deep learning model to enable effective training. The performance of the model is assessed using images of Tamil Nadu license plates. Character recognition accuracy is 90%, number plate localization accuracy is 98%, and vehicle detection accuracy is 97%.

The authors of [33] propose a real-time ALPR system that covers the entire process from beginning to end and is based on deep convolutional neural networks (CNNs). In order to get superior outcomes, the strategy involves recognizing the front and rear pictures of LPs and automobiles as they are running in a cascaded fashion. After that, characters are located and identified within the LP region that has been cut up.

An ALPR system that was focused on detecting and reading LPs in challenging settings was proposed by the authors of [34]. You Only Look Once version 3 (YOLOv3), a CNN-based model, was used for LP detection and character segmentation, and a convolutional recurrent neural network (CRNN) [35] was used for character recognition. Both of these models were trained on the same data set.

A new method for the detection of license plates (LPs) that are issued in many countries was presented in a paper that was published in [36]. LPD, Unified Character Detection (UCR), and Multinational License Plate Layout Detection were the three components that made up its architecture's three layers. The creators of this approach decided to use a new YOLOv3-SPP with a Spatial Pyramid Pooling block and a more compact YOLOv3 architecture for the LPD and CR stages of the method.

The authors of [37] proposed combining YOLO with a sliding-window method as a decision-making strategy. The sliding window, which is used in their method, is responsible for reading each character on Taiwan's automotive license plates. The YOLO framework is then able to identify each type of object, plate, digit, and letter that is seen in the window.

The authors of [38] utilized distinct CNNs for each stage of the ALPR process. The models that were used were Fast-YOLO [39], YOLOv2, and CR-NET [40], which is an architecture for character segmentation and recognition that was influenced by Fast-YOLO.

A brand-new LP detection and character identification system based on a backpropagation neural network (BPNN) and integrated feature extraction model was suggested in [41]. This system was based on a combination of the two models. This method is adaptable in situations with poor light and complicated backgrounds.

The writers of [42] have created an ALPR system that can identify Egyptian LPs using a mobile application. The approach proposes to perform character detection using segmentation, character recognition by extracting features using Speeded Up Robust Features (SURF), and finally character recognition by applying preprocessing to the acquired picture [43]

The authors of [44] have presented a solution for Chinese car LPR that makes use of a CRNN for CR stages and a YOLOv2 detector for LPD. They use a CNN for context feature extraction and a two-layered Gated Recurrent Unit (GRU) [45] for feature sequence decoding in their recognition architecture.

The LP was read using a cascade structure in [46]. The model initially identifies the character region using a CNN classifier, applied in a sliding window method across the full image, and then uses the run Length Smoothing Algorithm (RLSA) and Connected Component Analysis (CCA) to build bounding boxes separately at each scale. The resulting boxes are then refined by the edge feature of LP and filtered by geometrical restrictions. The final step was to test the remaining bounding box using a different CNN classifier.

In order to extract LP, the authors of [47] used a cascaded framework made up of an R-CNN network and a rapid region proposal network. The model's initial goal is to create LP candidates using a compact RPN network. then utilizes the sampler to extract the Region of Interest (ROIs) from the original image. Finally, the model classifies the candidate plate and regresses the four corners of the LP using the R-CNN network.

A lightweight ALPR model created by the authors in [48] can be used solely on embedded systems like the Raspberry Pi 3. They used a combination of a MobileNet [49] feature extractor with fewer parameters and a Single Shot Detection model to obtain the lowest memory consumption for the detection step. Additionally, they employed LPRNet20, a potent but computationally efficient network, for character recognition.

To implement multi-directional automobile LPD, the authors in [50] created a CNN-based technique called MD-YOLO that was motivated by the YOLO framework. To roughly represent the intersection ratio between anticipated value and tag value, they developed the angle deviation penalty factor (ADPF). In order to recognize negative rotation angle values, they chose leaky and identical functions as activation functions rather than ReLU function. Prior to the deployment of MD-YOLO, an ALMD-YOLO prepositive CNN attention model was used because the LP is frequently quite tiny.

Similar methods were used in this research by Kim et al. [51]. They replaced conventional optical character recognition (OCR) techniques with a model based on YOLOv2. The model used in this study, YOLOv3, has undergone minor updates to make it more current. A more modern foundation should result in more pleasing outcomes.

The input data must be readable before attempting to read characters from the image. A generative adversarial network (GAN) was used by Lin et al. [52] to reconstruct images of license plate numbers. Their study was based on the idea that outside factors such as closure and blurring, can impact an image's quality. To reconstruct these license plates, they specifically look at super-resolution GANS (SRGANs). similar to Lee et al. [53], who exploited super-resolution to enhance license plate recognition. These techniques were employed by both studies with varying degrees of effectiveness. It would be advantageous to investigate more recent super-resolution models, like the realESRGAN [54]. The research on license plate identification would advance as a result.

CR-NET predicts 35 classes as opposed to the 36 classes that are suggested in the dataset associated with this paper. This is due to the decision made by Laroca et al. [55] to regard the character "O" and "0" as a single class. This study attempts to develop a model that can distinguish between '0' and 'O' as well as other ambiguous characters by providing a model with training data for both characters.

Chapter 3 Methodology:

This project identifies and recognizes the license numbers and letters on the license plates of The Egyptian Vehicles. In order to extract the number from the images of the Egyptian license plates that were taken from saved images or videos, the system goes through two phases. The first phase is license plate detection. The system has to detect or extract the license plate from input car images, this phase is accomplished by using YOLO Algorithm.
The project provides the implementation of training and testing the three versions of YOLO on the dataset of Egyptian car plates and obtains different mAP values for each version of YOLO Algorithm with detailed analysis.
In addition to that, we made a comparison between results and their advantages and disadvantages of each version of the three versions of YOLO Algorithms that are used. Then the second phase is the license plate recognition, after extracting the plates from the Egyptian vehicles, then, the second phase the system should recognize what is in the license plate, which letters and numbers are included. The recognition phase has to apply OCR technique and it is implemented by using CNN model.
The following paragraphs will explain what are the three versions of YOLO Algorithms and how can YOLO extract the object from the image. Then, the detailed explanation of the CNN model architecture and how can CNN recognize the letters and numbers that are included in the plates.

1. Object Detection using YOLO Algorithm:

A powerful object detection technique in computer vision is known as YOLO. YOLO is an acronym that stands for "You Only Look Once." The YOLO algorithm can recognize objects in still images or video frames by first dividing the input image into cells and then calculating the bounding boxes and class probabilities for each cell individually. Because it has the ability to distinguish objects in real-time video with a high degree of accuracy, YOLO has earned a reputation for both its speed and its accuracy. Additionally, the system is able to recognize several pieces within a single image and can cope with objects that have varying dimensions as well as aspect ratios.

The YOLO model typically divides itself into three sections: the backbone, the neck, and the head. Backbone is a feature extraction network that is used to extract feature information from images [22,23]. Neck can fuse the features extracted from backbone, increasing the range of the features learned by the network and enhancing the performance of the detection network. Head can make precise predictions by utilizing earlier high-quality feature engineering. Backbone, neck, and head are all examples of neural networks. Nearly every iteration of the YOLO models has resulted in the same or comparable developments and improvements being made in each of these three structures. Due to the remarkable performance of the models in the YOLO series in terms of both accuracy and speed of detection, these sensors have found widespread application in a wide range of fields, including remote sensing, transportation, medical, and other industries.

A new method for the detection of license plates (LPs) that are issued in many countries was presented in a paper that was published in [6]. LPD, Unified Character Detection (UCR), and Multinational License Plate Layout Detection were the three components that made up its architecture's three layers. The creators of this approach decided to use a new YOLOv3-SPP with a Spatial Pyramid Pooling block and a more compact YOLOv3 architecture for the LPD and CR stages of the method.

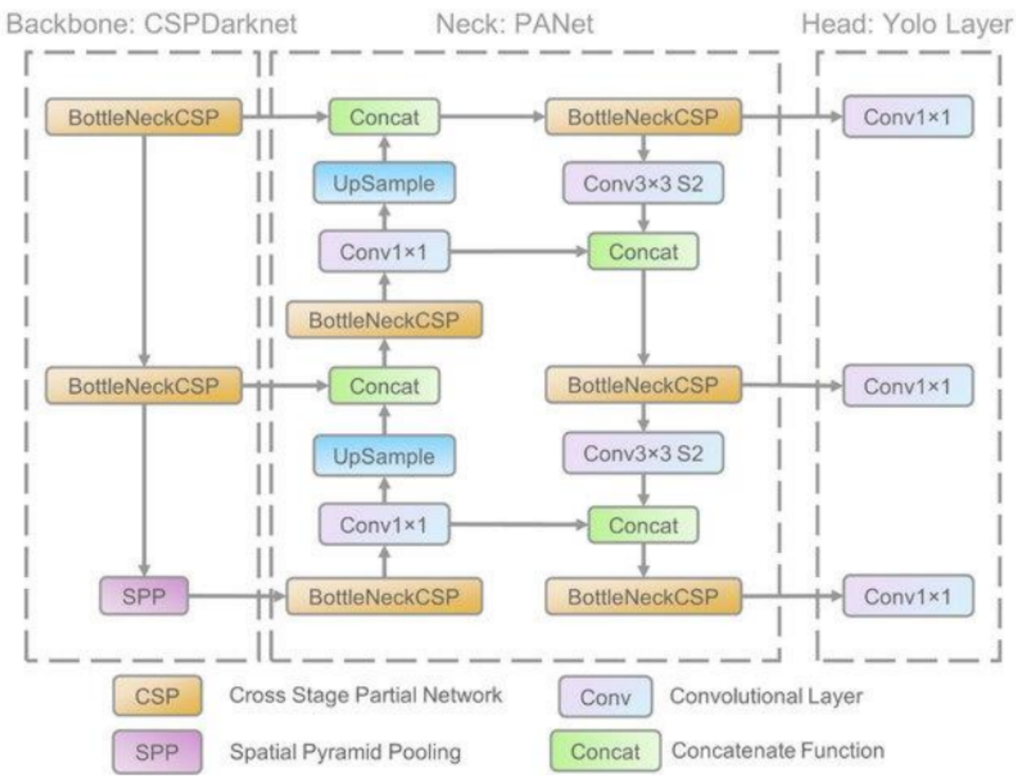
The authors of [7] proposed combining YOLO with a sliding-window method as a decision-making strategy. The sliding window, which is used in their method, is responsible for reading each character on Taiwan's automotive license plates. The YOLO framework is then able to identify each type of object, plate, digit, and letter that is seen in the window.

The authors of [8] utilized distinct CNNs for each stage of the ALPR process. The models that were used were Fast-YOLO [9], YOLOv2, and CR-NET [10], which is an architecture for character segmentation and recognition that was influenced by Fast-YOLO.

A brand-new LP detection and character identification system based on a backpropagation neural network (BPNN) and integrated feature extraction model was suggested in [11]. This system was based on a combination of the two models. This method is adaptable in situations with poor light and complicated backgrounds.

A. Object Detection Using YOLO V5:

As for YOLO (You Only Look Once) architectures, YOLOv5 was introduced in May 2020 as an upgrade to YOLOv4, It is fast and accurate as different YOLO models. YOLOv5 is a single-shot detection model which process the whole image in one path to detect objects unlike other two stages models which they detect the objects then run classification on the detected object then refining them, YOLOv5 does all of that in one try, this causes faster detection and inference.
YOLOv5 uses CSPDarknet as it's the backbone, PANet as it's neck and the head is YOLO layer, the YOLO layer is responsible for outputting the detection results, CSPDarknet is responsible for feature extraction the it is passed to PANet for feature fusion as shown in the Architecture image below:



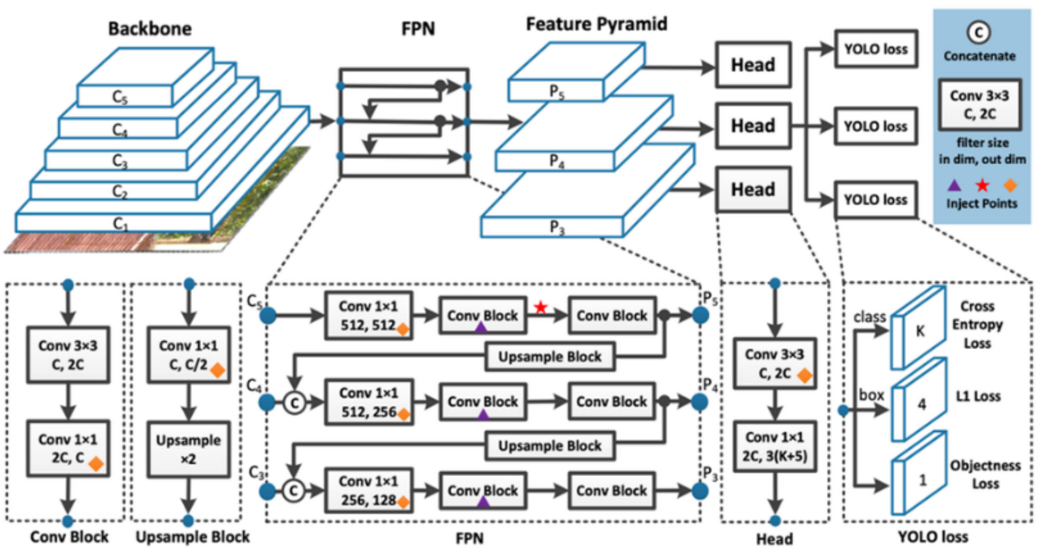
Yolo architecture [11]

Nevertheless, YOLOv5 has the upper hand when it comes to engineering. Python was chosen as the programming language for YOLOv5 as an alternative to the preceding versions' use of the C programming language. On Internet of Things devices, this makes installation and integration easier. Additionally, the Darknet community is smaller than the PyTorch community, which means that Darknet does not have as much room for future development and contributions as PyTorch does. Due to the fact that YOLOv4 and YOLOv5 were developed using two independent programming languages and frameworks, it is difficult to make an accurate comparison between their respective performance levels. However, over the course of time, YOLOv5 has shown that it performs better than YOLOv4 under some circumstances and has garnered some support from the computer vision community in addition to YOLOv4's endorsement. Because both the YOLOv5 and YOLOv4 architectures have incorporated the most recent technological advancements, there are not many substantial distinctions between the two in terms of theory. The author has merely opened a repository on Github, and updates and changes are only posted on that platform; no thorough paper has been published. Through an examination of the structural code contained in the file, the YOLOv5 model can be summed up as follows.yaml (Jocher, 2020):
1.Backbone: CSP network and focus structure.
2.Neck: PANet, SPP block.
3.Head: Using GIoU-loss, YOLOv3 head.

The author of YOLOv5 made a noteworthy comment about an engineering distinction. In YOLOv2, Joseph Redmon presented the anchor box structure and a method for choosing anchor boxes that closely mimic the ground truth bounding boxes in the training set in terms of size and shape. The authors selected the 5 best-fit anchor boxes for the COCO dataset (containing 80 classes) and used these as the default by using the k mean clustering algorithm with various k values. This shortens the training process and improves network accuracy. These 5 anchor boxes, however, are unable to swiftly adapt to the ground truth bounding boxes of a unique dataset that contains classes that do not belong to 80 classes in the COCO dataset[26].
By way of example, a giraffe dataset prefers anchor boxes that are thin and higher than square boxes. Computer vision engineers typically use the k-mean clustering approach to generate the best-fit anchor boxes for the data first in order to solve this challenge. Following that, these settings will be manually configured in the YOLO architecture.
Glenn Jocher suggested including the anchor box selection method in YOLOv5. As a result, the network need not think about which datasets to use as input because it will automatically "learn" the best anchor boxes for that dataset and apply them throughout training. (Solawetz, 2020)

B. Object Detection using YOLO V7:

YOLOv7 joined YOLO models on July 2022, at this time YOLOv7 was the most accurate and fastest object detector, Over the earlier editions, the recent YOLO version features a number of enhancements. Using anchor boxes is one of the primary advancements. An array of pre-defined boxes with various aspect ratios called "anchor boxes" are used to identify objects of various shapes. As opposed to earlier iterations, YOLO v7 employs nine anchor boxes, enabling it to recognize a greater variety of item sizes and forms, hence lowering the incidence of false positives.
YOLOv7 uses several architectures similar to YOLOv4, YOLOv7 uses E-ELAN (Extended Efficient Layer Aggregation Network) and model scaling for concatenation models, first, E-ELAN, E-ELAN is the main block of the backbone of YOLOv7, it aggregates the features from the layers taking into account that the way should be optimized. As an advantage of YOLOv7 it uses something called BoF (Bag of Freebies) which helps enhancing the performance without any additional cost in the training. We can visualize the YOLOv7 Architecture in the below figure:



In YOLO version 7, a big advancement has been made with the introduction of a new loss function that is referred to as "focal loss." In early iterations of YOLO, the standard cross-entropy loss function was used. This function is well-known for having a lower level of success when attempting to recognize small objects. This issue is addressed by focal loss, which reduces the weighting of loss for cases that have been categorized accurately while increasing the weighting of loss for cases that are difficult to identify. In addition, the resolution of YOLO v7 is far higher than that of its predecessors. It processes photographs at a resolution of 608 by 608 pixels, which is an improvement over YOLO v3, which processed photographs at a resolution of 416 by 416 pixels. Because of its improved resolution, the YOLO v7 is able to detect extremely minute objects with a higher degree of accuracy.
One of the most important advantages of YOLO v7 is how quickly it can be used. It processes images at a pace of 155 frames per second, making it significantly more efficient than other cutting-edge object detection algorithms. Even the YOLO model that was used as the foundation from the

very beginning was capable of processing at a top speed of 45 frames per second. It is suitable for such applications as surveillance and autonomous vehicles because to the fact that enhanced processing rates are necessary for sensitive real-time applications such as these.
The YOLO v7 algorithm performs quite well in terms of accuracy when compared to other object detection algorithms. On the well-known COCO dataset, it achieves an average precision of 37.2% when using an IoU (intersection over union) threshold of 0.5. This is in comparison to other cutting-edge object detection algorithms that have been developed recently. The quantitative comparison of performance is presented in the following table[59].

Model	#Param.	FLOPs	Size	AP^{val}	AP^{val}_{50}	AP^{val}_{75}	AP^{val}_{S}	AP^{val}_{M}	AP^{val}_{L}
YOLOv4 [3]	64.4M	142.8G	640	49.7%	68.2%	54.3%	32.9%	54.8%	63.7%
YOLOR-u5 (r6.1) [81]	46.5M	109.1G	640	50.2%	68.7%	54.6%	33.2%	55.5%	63.7%
YOLOv4-CSP [79]	52.9M	120.4G	640	50.3%	68.6%	54.9%	34.2%	55.6%	65.1%
YOLOR-CSP [81]	52.9M	120.4G	640	50.8%	69.5%	55.3%	33.7%	56.0%	65.4%
YOLOv7	36.9M	104.7G	640	**51.2%**	**69.7%**	**55.5%**	**35.2%**	**56.0%**	**66.7%**
improvement	-43%	-15%	-	+0.4	+0.2	+0.2	+1.5	=	+1.3
YOLOR-CSP-X [81]	96.9M	226.8G	640	52.7%	**71.3%**	57.4%	36.3%	57.5%	68.3%
YOLOv7-X	71.3M	189.9G	640	**52.9%**	71.1%	**57.5%**	**36.9%**	**57.7%**	**68.6%**
improvement	-36%	-19%	-	+0.2	-0.2	+0.1	+0.6	+0.2	+0.3
YOLOv4-tiny [79]	6.1	6.9	416	24.9%	42.1%	25.7%	8.7%	28.4%	39.2%
YOLOv7-tiny	6.2	5.8	416	**35.2%**	**52.8%**	**37.3%**	**15.7%**	**38.0%**	**53.4%**
improvement	+2%	-19%	-	+10.3	+10.7	+11.6	+7.0	+9.6	+14.2
YOLOv4-tiny-3l [79]	8.7	5.2	320	30.8%	47.3%	32.2%	**10.9%**	31.9%	51.5%
YOLOv7-tiny	6.2	3.5	320	**30.8%**	**47.3%**	**32.2%**	10.0%	**31.9%**	**52.2%**
improvement	-39%	-49%	-	=	=	=	-0.9	=	+0.7
YOLOR-E6 [81]	115.8M	683.2G	1280	55.7%	73.2%	60.7%	40.1%	**60.4%**	69.2%
YOLOv7-E6	97.2M	515.2G	1280	**55.9%**	**73.5%**	**61.1%**	**40.6%**	60.3%	**70.0%**
improvement	-19%	-33%	-	+0.2	+0.3	+0.4	+0.5	-0.1	+0.8
YOLOR-D6 [81]	151.7M	935.6G	1280	56.1%	73.9%	61.2%	**42.4%**	60.5%	69.9%
YOLOv7-D6	154.7M	806.8G	1280	56.3%	73.8%	61.4%	41.3%	60.6%	70.1%
YOLOv7-E6E	151.7M	843.2G	1280	**56.8%**	**74.4%**	**62.1%**	40.8%	**62.1%**	**70.6%**
improvement	=	-11%	-	+0.7	+0.5	+0.9	-1.6	+1.6	+0.7

The quantitative comparison of YOLO performance

C. Object Detection using YOLO V8:

YOLOv8 is the latest, highest accuracy and fastest object detection model in the YOLO family, it was released in January 2023[57]
The C3 module is replaced by the C2f module, which is based on the CSP principle, and the backbone of YOLOv8 is fundamentally the same as that of YOLOv5. To enable YOLOv8 to gather more plentiful gradient flow information while guaranteeing its portability, the C2f module built on the ELAN concept from YOLOv7 and combined C3 and ELAN to create the C2f module [22]. In order to guarantee the correctness of objects in various scales while also maintaining a lightweight, the most common SPPF module was still employed at the end of the backbone. Three Maxpools of size 5*5 were passed serially, and then, each layer was concatenated. The feature fusion approach still employed by YOLOv8 in the neck section is PAN-FPN, which improves the fusion and usage of feature layer data at various scales. The neck module was created by the authors of YOLOv8 using two up-sampling, numerous C2f modules, and the final decoupled head structure. For the final portion of the neck, YOLOv8 adopted the concept of decoupling the head from the body in YOLOx. To attain a higher level of accuracy, it integrated confidence and regression boxes. YOLOv8 is capable of supporting all YOLO versions and switching between them at anytime. Its broad hardware compatibility (CPU-GPU) further increases its adaptability. Figure 1. provides the YOLOv8 network architectural diagram. Convolution, batch normalization, and SiLu activation functions make up the CBS in Figure 11.

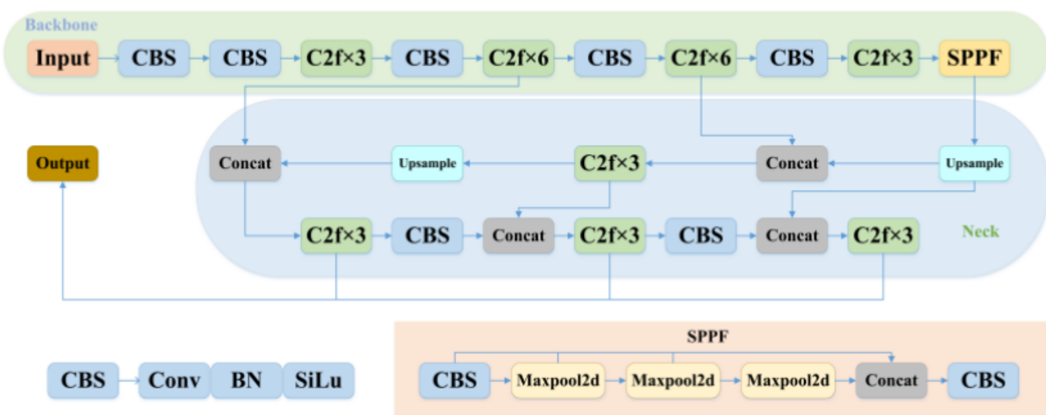
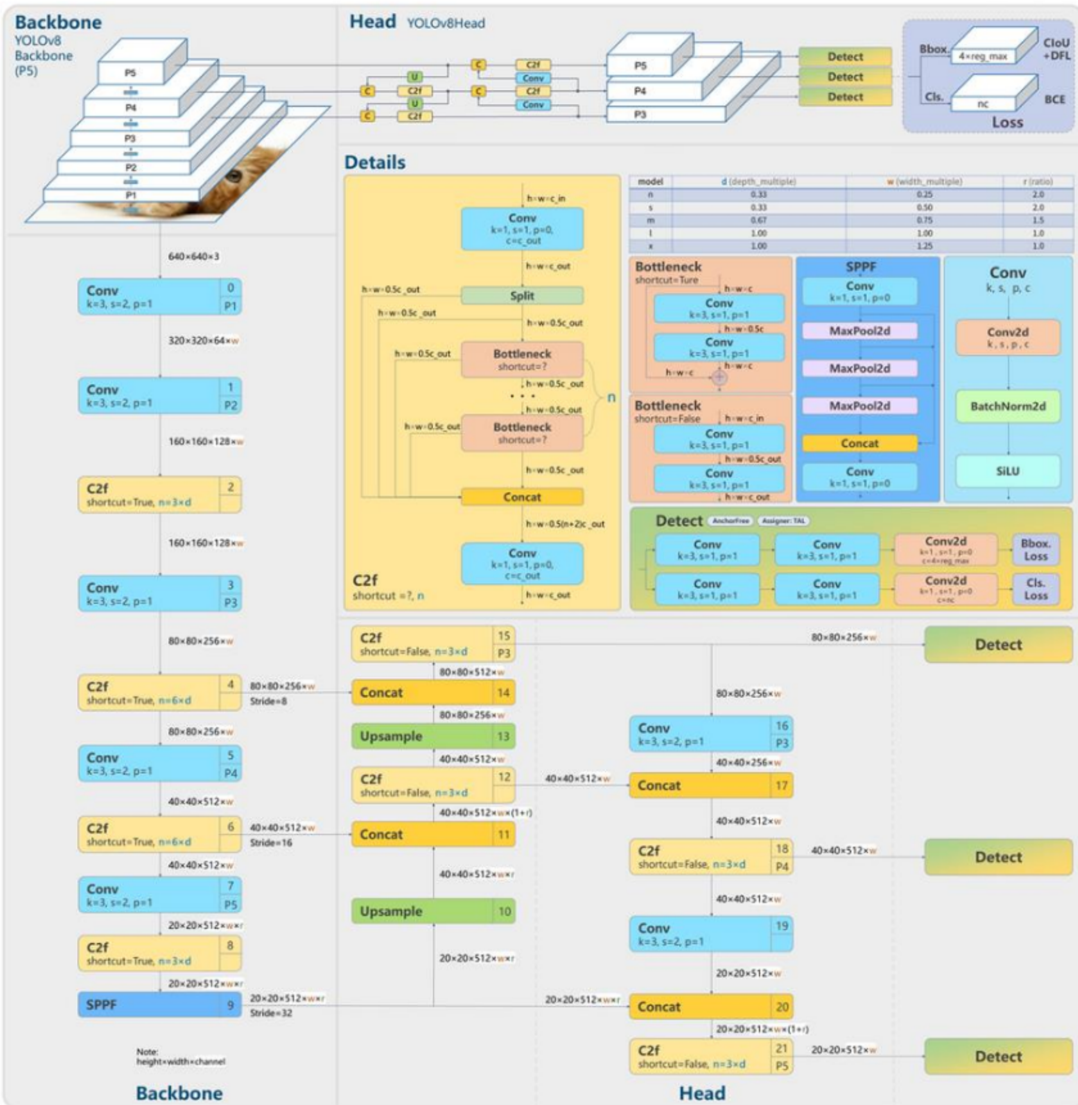
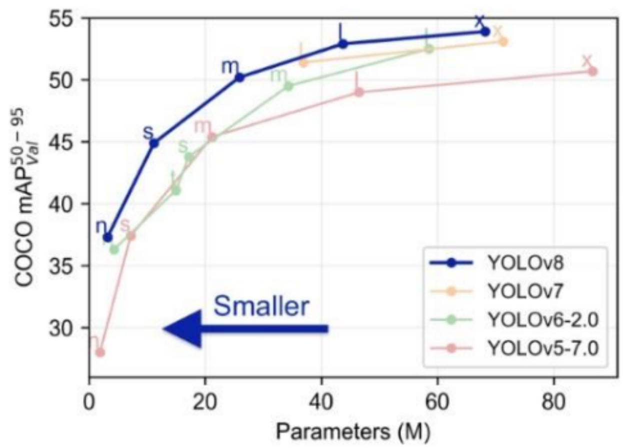


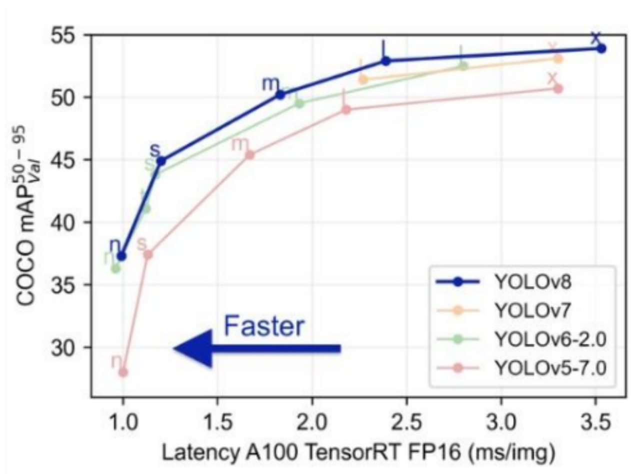
Figure 11. The network structure of yolov8 [57].



This is The YOLOv8 Architecture in the below figure [58].

As shown in the below figures, YOLOv8 uses less resources than other YOLO versions and it is faster than the others too.





Comparison between YOLO V5, YOLO V7, and YOLO V8 Algorithms:

This part introduces an comparison between the most well-liked algorithms of recent years and provides a detailed explanation of some of the paper's key ideas for YOLOv8 enhancement. Because of the following factors, YOLO is now the most well-liked real-time object detector: More precise detection outcomes, (a) a light-weight network design, (b) efficient feature fusion techniques. The two most popular algorithms right now in terms of usage are YOLOv5 and YOLOv7. In YOLOv5, deep learning technology is employed to perform object identification tasks in real-time and with high efficiency. According to model structure, training approach, and performance, YOLOv5 had been enhanced over its forerunner YOLOv4. By using the CSP (CrossStage Partial) network topology, YOLOv5 was able to efficiently lower the number of times calculations were repeated and increase processing performance. YOLOv5 has several disadvantages, though. For instance, there are still some issues with small item detection, and dense object detection has to be addressed. Furthermore, there is still room to improve YOLOv5's performance in challenging circumstances like occlusion and position change.

For enhancing the performance of real-time object detectors, YOLOv7 presented a unique training approach dubbed Trainable Bag of Freebies (TBoF). By using the TBoF approach with three distinct types of object detectors (SSD, RetinaNet, and YOLOv3), it was possible to considerably increase the accuracy and generalization ability of the object detector. The TBoF method also featured a number of trainable methods, such as data augmentation, MixUp, etc. However, YOLOv7 is also constrained by the training data, model structure, and hyperparameters, which can occasionally result in performance loss. The suggested solution also needs greater processing power and training time to operate at its best.

The 2023 release of YOLOv8 sought to bring together the greatest features of numerous real-time object detectors. It continued to incorporate the CSP concept into the SPPF module, feature fusion method (PANFPN), and YOLOv5 [28,29,30]. The following changes were the important ones: (a) It offered a brand-new SOTA model using the instance segmentation model from YOLACT and object detection networks with P5 640 and P6 1280 resolutions [31]. It also created models of various scales based on a scaling coefficient comparable to YOLOv5 to fulfil the requirements of various projects. (b) The C2f module was created using the ELAN structure from YOLOv7 with the assumption that the original concept of YOLOv5 will be preserved [22]. (c) The detection head component likewise employed the prevailing technique (separating the classification and detection heads) [32]. The majority of the other components continued to be based on YOLOv5's original concept. (d) The categorization loss for YOLOv8 utilised BCE loss. The backsliding The loss had the formula CIOU loss + DFL, and VFL suggested a mixed weighting operation [33].

DFL: A generic distribution was used to model the box's location. The probability density was as close to the site as feasible, as shown in Equation (1), and the network immediately focused on the distribution of the location close to the object position. Yi and Yi+1 are interval orders, si is the network's sigmod output, and y is a label. DFL: A generic distribution was used to model the box's location. The probability density was as close to the site as feasible, as shown in Equation (1), and the network immediately focused on the distribution of the location close to the object position. (Yi) and (Yi+1) are interval orders, (si) is the output of sigmod for the network, and (y) is a label.

YOLOv8 is a more expandable method when compared to the original YOLO algorithm. It is a framework that can handle earlier YOLO versions and swap between them, making it simple to assess how well each version performs.

$$DFL_{(s_i,s_{i+1})} = -((y_{i+1} - y)\log(s_i) + (y - y_i)\log(s_{i+1})) \quad (1)$$

Anchor-Free is used by YOLOv8 as opposed to Anchor-Base. Dynamic TaskAlignedAssigner was the matching mechanism employed by V8. Equation (2) is used to determine the alignment degree of the Anchor-level for each instance; the weight hyperparameters are and, and the classification score and IOU value are s and u, respectively. The m anchors with the highest value (t) in each instance are chosen as positive samples, while the remaining anchors are chosen as negative samples. The loss function is then trained on these examples. The most accurate detector to date is YOLOv8, which has the aforementioned enhancements and is 1% more accurate than YOLO

$$\mathbf{t} = s^{\alpha} \times \mathbf{u}^{\beta} \quad (2)$$

The ability to be extended is YOLOv8's standout feature. Researchers working on YOLO projects greatly benefit from the fact that YOLOv8 is built to function with all versions of YOLO and to transition between them, making it simple to compare their performance. As a result, the version YOLOv8 was chosen as the baseline.

2. Character Recognition using CNN:

A word on an image usually includes valuable information and immediately expresses high-level meaning, making it an important source of knowledge and one of the most contentious academic topics. The usage of CNN-based neural networks, which form the core of text recognition, has been demonstrated in numerous studies to be quite accurate and successful in classifying images. A pre-trained model that was trained using the ImageNet dataset as an initial weight can also be used to improve it further. The CNN model, one of the most effective deep learning models used in the field of image identification, can automatically learn discriminative pattern features from a huge quantity of data. This scheme's recognition accuracy is comparable to that of human vision when a sizable amount of data is used[26]. As a result, we apply it to the character recognition of license plates and adjust to the network topology, and the approach acquired a decent recognition rate.

CNN is a unique kind of feed-forward multilayer perceptron that can automatically extract features. It is learned in supervised mode using the gradient descent backpropagation learning technique. The most widely used techniques typically combine a hand-crafted feature extractor and trainable classifier. Low accuracy and less-than-ideal results could arise from this. In tasks including optical character recognition, generic object recognition, real-time face detection and position estimation, speech recognition, license plate recognition, etc., CNN has demonstrated its ability to achieve cutting-edge results. CNN incorporates three architectural concepts: shared weights, subsampling, and local receptive field.[28]

Figure 17 shows the CNN structure for the text at the character recognition stage.

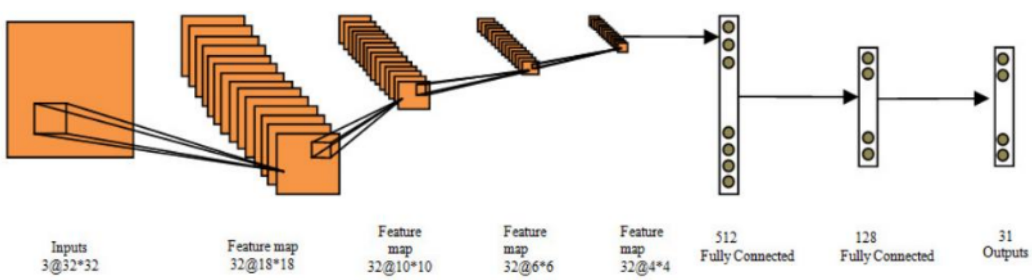


Fig.17 The structure of CNN [28]

Four convolution layers, four pooling layers, two fully linked layers, and one output layer make up the network, as illustrated in Fig. 6. An image with the dimensions 32*32*3 is used as the input. Following a Max-pooling layer and a convolution layer with a convolution kernel size of 5*5, a set of feature maps is created. Max-pooling; The convolutional output bands can be down-sampled to reduce variability using the max-pooling operator [22]. The maximum value inside a set of R activations is sent forward by the max-pooling operator. The J-related filters P(m) make up the m-th max-pooled band as shown in the following equation.

$$p_{j,m} = \max\left(h_{j,(m-1)N+r}\right) \qquad (1)$$

Max Pooling Equation

and convolution layers are applied after each of the two feature maps. The network has a Dropout (rate=0.5) The Dropout layer acts as a mask, eliminating some neurons' contributions to the subsequent layer while maintaining the functionality of all other neurons. [22] layer before the output layer and uses the activation function Leaky ReLU, The ReLU is currently the activation function that is utilised most frequently worldwide. Considering that practically all convolutional neural networks or deep learning employ it. The ReLU is, as can be seen, 50% corrected (starting at the bottom). When z is less than zero, f(z) is equal to zero; when z is more than zero, f(z) is equal to z. for the loss function and categorical_crossentropy [15] for the activation function. using Adam (lr=0.0001), Adam is a different optimization approach that produces more effective weights of neural networks by doing repeated rounds of "adaptive moment estimation." Adam uses stochastic gradient descent to expand stochastic gradient descent, which uses fewer resources and can solve non-convex problems more quickly than many existing optimisation programmes. [32], with softmax [17] serving as the output function.

3. Dataset:

The law in Egypt for the Egyptian license plate is to enforce traffic laws, traffic police or traffic officers in Egypt typically record vehicle identification numbers and characters. Errors in writing or reading the numbers and characters could occur. The widespread use of mobile phones can be utilized in the suggested study. The device converts images of the numbers and characters on a license plate taken by an officer into digital numbers and letters. Arabic characters can be difficult to read because, unlike English characters, some of them are quite similar to one another. For instance, there is little difference between letters and that could make a challenge in recognition phase those letters are: feh (ف) and Qaaf (ق), noon (ن), and ba (ب).[54]
The Egyptian General Traffic Department has prohibited the use of certain letters and numbers by not placing them on those metal plates of cars. It also determined the function of the vehicle that it performs, and this is done by following several specific methods and means. The letters that the General Egyptian Traffic Department did not write on the plates are 11 letters, and they are "ك, غ, ذ, ز, ش, ظ, ض, خ, ح, ث, ت.". Likewise, the numbers that the General Traffic Department removed from it is only one number, and it is "zero". The reason for excluding this number is due to the fact that no citizen should put a nail on the metal plate of his car, as it appears as a zero, and this has the ability to escape from traffic violations, or any other type of violation.
There is a method used to recognize the activity of the car, which has been determined by the General Egyptian Traffic Administration [56], as they have developed a certain set of colors, which refer to the activity of the car, these colors include:

1. License Plate that are in blue color, that means this car is a citizen car.
2. The red color of the metal plate, means it is a pickup car, just as the metal paintings of the tractors have a red color.
3. The plate that are orange, it indicates that this car is a taxi.
4. Plate that includes brown, it means that this car is a commercial car.
5. License plates in dark blue, it means that this car is a police car.
6. The license plates are green, which means that they are cars belonging to the main authorities in the country.
7. The plates that contain gray at the top of the plate are public sector cars and public transport buses.
8. Finally, the plates that are in yellow, it means that car is the customs car.

Egypt has been using a standard rectangular plate with three informative spaces since August 2008, as seen in Figure 1. The details of the license plate are outlined in Article (340) of the Interior Minister's Decision No. 2777, as shown in Table 1. The rectangular plate is constructed of aluminium and has a reflective color. Additionally, it was created in three parts, as seen in Figure 1. Except when the size is specified, the plate's dimensions are 35x17 cm2. Additionally, the plate numbers and the background color are displayed. (1) The color-coded section that identifies the license type for the vehicle. The term "Egypt" appears there in both English and Arabic in black lettering on a variety of colored backgrounds, depending on the type of vehicle's license plate. Similar but considerably smaller license plates with light blue (private motorcycles) and dark blue (police motorcycles) are found on motorcycles. There are only these two colors in existence. Table 2 discusses the various colors of the top rectangle with the various vehicle types. (2) The number is typeset in Hindi as seen in Figure 1. (3) In the event of non-police registration, the alphabet and number system employ either a set of Hindi type, or a set of Arabic letters, but not both. However, the police vehicles make use of both an alphabetic and numeric combination with some structure, as shown in Table 2. This table illustrates the potential numeric and alphabetic combinations for each governorate.

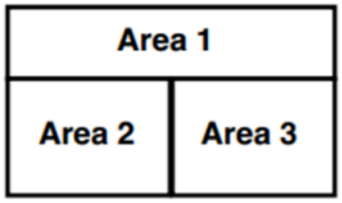


figure 19. Three regions of Egyptian license plates

Figure 1 shows all three regions of Egypt's license plates since 2008: Area 1 is the color, Area 3 has either letters or numbers, and Area 2 has either numbers or letters and numbers or a combination of both.

Challenges in Discovering and Recognizing the Egyptian license plates:

The difficulties in identifying Egyptian license plates are caused by poor picture quality in outdoor environments [54]. Besides the following obstacles:

1. It is particularly challenging to accurately identify color due to color capture information changing depending on the conditions of the capture, such as lighting, reflection, etc. Color serves as an identification to the type of registration, making it very difficult to detect color correctly.

2. Different laws apply to different vehicles that use the same road as part of the para-authority of various ministries. For instance, the Interior Minister's Decision No. 2777 did not include the unique plate made for the army and special forces, resulting in cars with different license plates, as indicated in Figure 8, none-no conformity plates.

3.Due to the vehicle or the cargo partially obscuring the license plate, it is called a license plate obstruction. Figure 20. depicts a typical license plate cover.

Figure 20. License plate occlusion due to parts of the vehicle is covering the plate[63].

4. duplicating a license plate or placing one on top of another while driving. The typical cover for a license plate is shown in Figure 4.

Figure 21. several license plates, which could confuse the detector[63].

5. using a foreign license plate when travelling. One common sight is a car going with a license plate from a different country. Figure 22. expresses an example of using a foreign license plate while driving.

Figure 22. Traveling using a license plate from a foreign country.[63]

6. travelling with a license plate that was, as we described it, "homemade." Unmachined-written license plates have been seen on several cars as they travel.

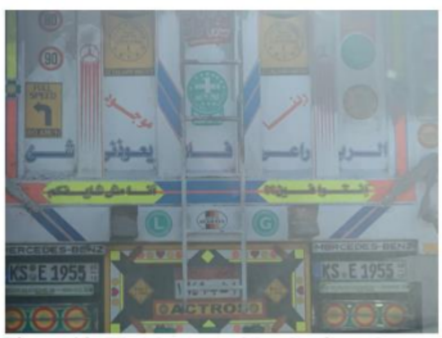
7. The common practice of altering a license plate to make it non-reflective instead of reflecting alters the picture circumstances and renders any detection inoperable. Figure 23. Depicts two instances of common tampering with a vehicle that has recently passed through a police and road authority checkpoint, at the very least

Figure 23. Painting the license plate to show tampering with it.[63]

8. In addition to a unique logo on the license plate, a specific registration outlet masks the license plate and modifies its structural elements, resulting in the creation of a new alternate category of plates that can be used to detect the license plate. Nevertheless, it might be simpler to spot the emblem than to recognize the plate. An illustration of the emblem on a license plate affixed by the registration office is shown in Figure 24.

Figure 24. Display an unfamiliar additional logo for a specific registration.[63]

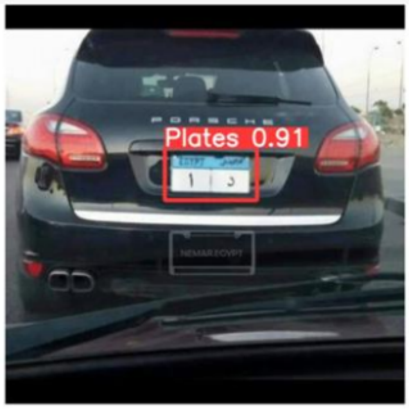
9. As shown in the Figure 25. The studded follows of trucks make it incredibly difficult to identify the proper license plate, as demonstrated in Neural Network and Artificial Intelligence approaches. These methods are defendable in a court of law under common law or natural law because a judge cannot understand the process. Any court will not accept the idea of training a black box, neural network, or deep neural network a set of photos, then having the system recognize a set of plates that are not in its learning sample. Additionally, the idea of marking suspects for additional inquiry when a computer is employed to aid a person where the machine will decide, and their verification of a human will affirm the judgment. In addition, they lack real-time, therefore performance inaccuracy is not a serious issue.



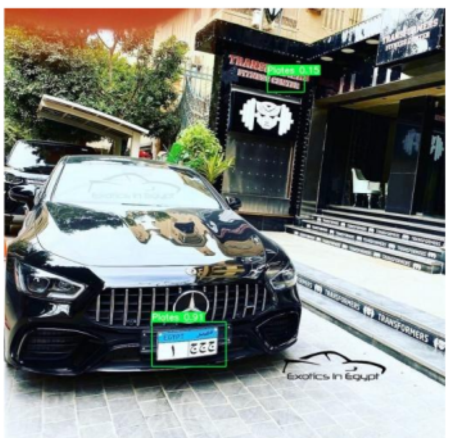
Trucks with decorative backs.[63]

Chapter 4 Results and Analysis:

The experimental analysis of our complete ALPR system is covered in this part.
The project used 2029 images of Egyptian car plates. The first step is labeling the images using the bounding box technique. Then, applying different types of augmentation techniques such as rotation, salt and pepper noise and brightness. Those methods are implemented using Roboflow program. The dataset became 5505 images after applying augmentation. The images are divided into 2 sections, 80% goes for first section, it is for training set, second section goes to testing set and it is about 20%.
The system's performance and accuracy were evaluated using the suggested techniques.
The suggested techniques are YOLO V5, YOLO V7, and YOLO V8 for detection of the license plate and, then the output from the detection phase is the input of the CNN model to make the license plate be recognized. Our system has the ability to overcome on common problems which are: Overfitting and Underfitting. when the training set of data shows low model performance. Underfitting happens when a model is overly simplistic, which can happen when there is insufficient regularization, training time, or input features.
When a machine learning model makes correct predictions for training data but not for data that is new, this is a bad machine learning behavior, known as overfitting.
The proposed system overcomes those problems by taking care of some important factors such as strong network, Increasing the dataset, using high quality images and augmentation on dataset.
The following figures show some examples of license plate detection using different of YOLO versions.



LPD using YOLO V5



LPD using YOLO V7

LPD using YOLO V8

As shown, in the figures the performance of YOLO V8 is better than the YOLO V7.

There are some factors by calculating them we can find out the model evaluation and accuracy of the YOLO Algorithm.
(1). Recall is defined as a ratio of real positives to all real (relevant) objects. Calculating recall, also known as sensitivity, involves dividing the total number of positive samples by the proportion of positive samples that were correctly identified as positive. It gauges how well a model can identify positives; the higher the recall, the more positives are identified.
(2). Precision is the ratio of true positives to all positive predictions is known as precision. To give an example, if the model found 100 license plates and 90 of them were accurate, the precision is 90%.
(3). Mean average precision (mAP) is a measurement used for evaluating YOLO object detection methods. In order to calculate a score, the mAP compares the detected box to the ground-truth bounding box. The model is more precise in its detections the higher the score.

The type of loss calculated in YOLO (You Only Look Once) models is a combination of various components, including localization loss, confidence loss, and class loss.
Loss of Localization: YOLO models try to forecast the bounding boxes for various objects in a picture. The localization loss calculates the mistake in foretelling the bounding box's dimensions (x, y, width, and height). Usually, it is estimated using smooth L1 loss or mean squared error (MSE) regression loss functions.

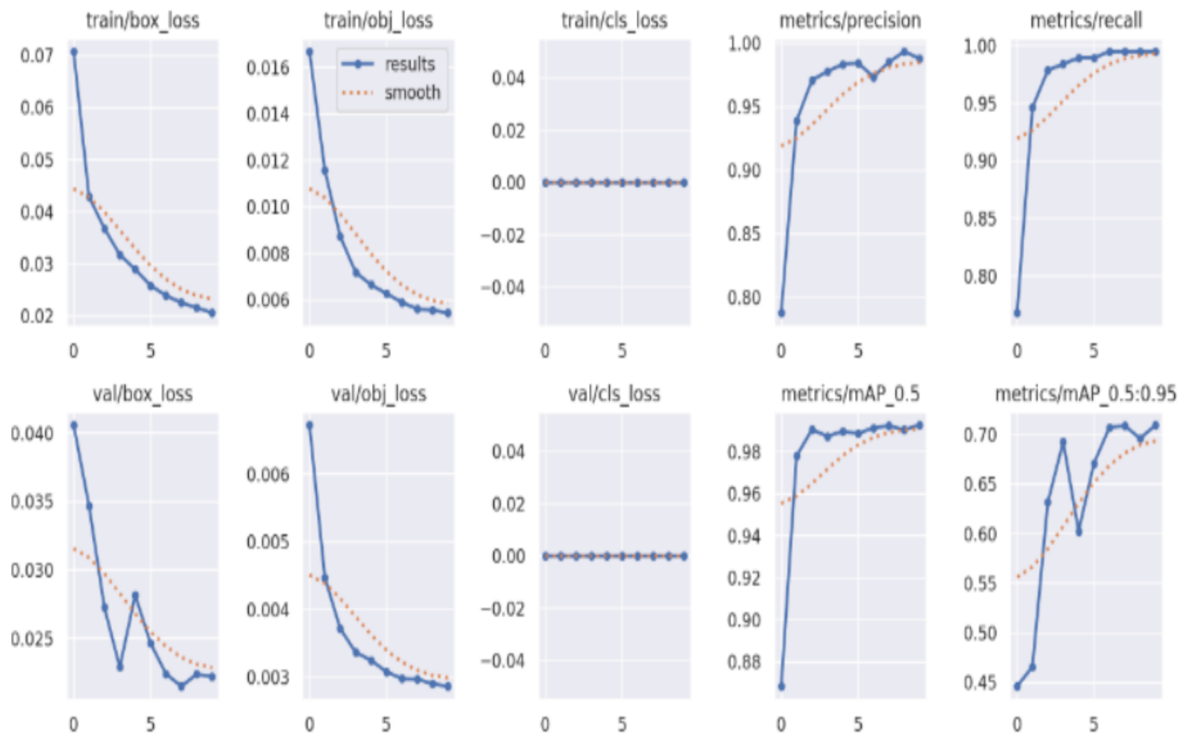
Confidence Loss: For each bounding box, the confidence loss assesses how accurately the anticipated objectness score was made. It gauges how accurately the model can determine whether or not an object is present in a specific grid cell. Binary cross-entropy loss or sigmoid cross-entropy loss are used to calculate the confidence loss.

Class Loss: YOLO models also categorize the things inside each bounding box into multiple classes. The estimated class probability's accuracy is assessed by the class loss. The difference between the ground truth labels and the anticipated class probabilities is measured using categorical cross-entropy loss.

In YOLO models, the total loss function is the weighted sum of each of these discrete losses, with varying weights often applied to each component. A balance between localization accuracy, objectness prediction, and class prediction is frequently achieved by adjusting the weights based on the significance of each loss term.
Given that the original design has experienced several modifications and enhancements since it was first introduced, it is significant to note that the precise formulation of the loss function may differ across various versions and variants of YOLO models.

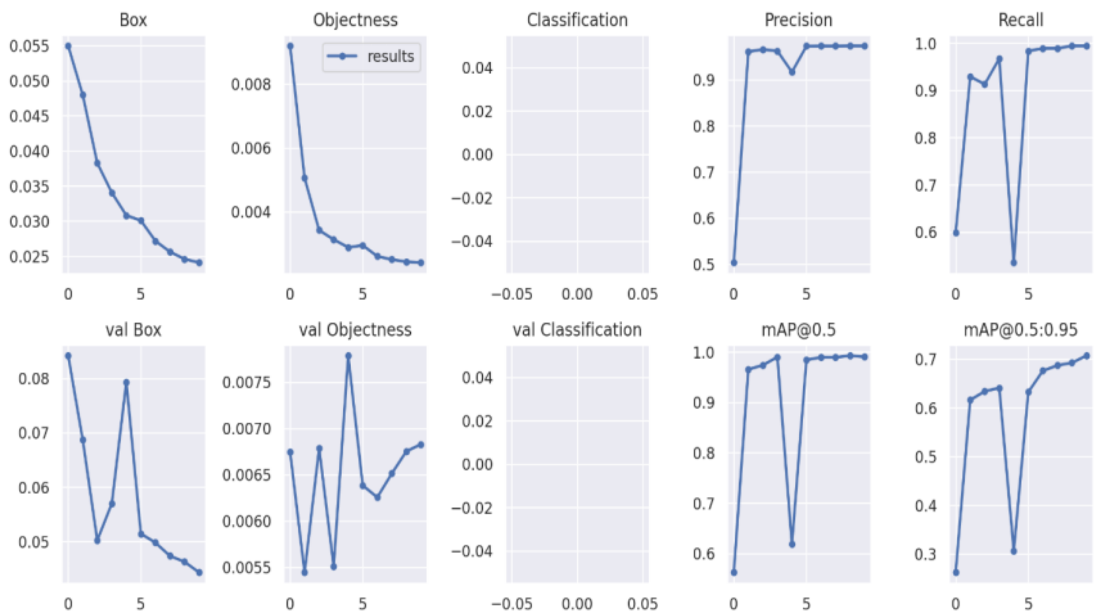
Yolo v5 results:

We used the default parameters of YOLOv5 with batch size of 16 and on just 10 Epochs, we have obtained 99% mAP, here are some detections on the test set: As it shown it made some wrong detections and others good.



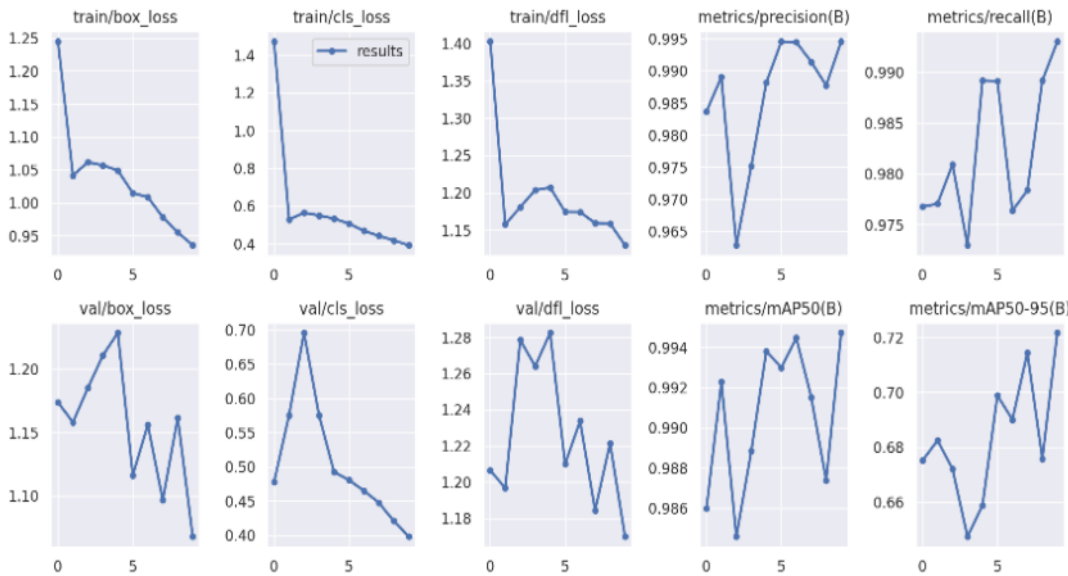
YOLOv7 Results:

We used the default parameters of YOLOv7 with batch size of 16 and on just 10 Epochs, we have obtained 98.7 mAP, here are some detections on the test set: As it shown it made some wrong detections and others good.

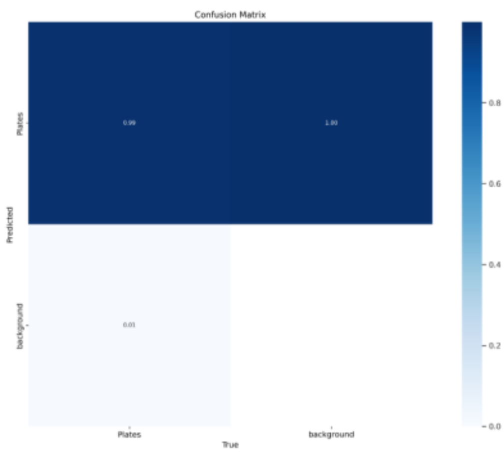


YOLOv8 Results:

We did the same things we did in the previous versions obtaining higher mAP values and in less training time.

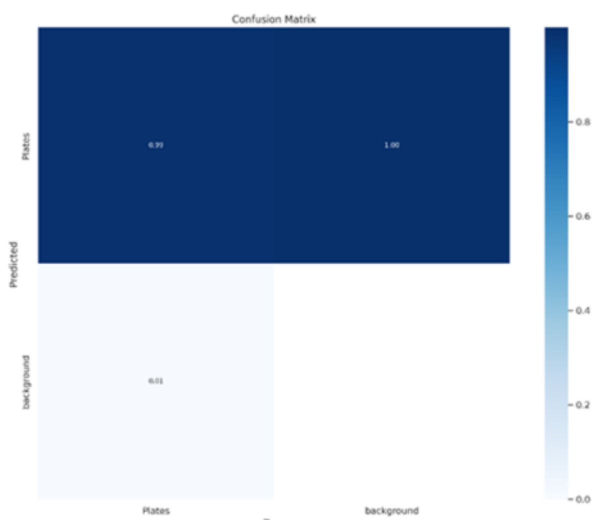


When running the model online, it augments the data so the model sees different data at each iteration, one of those augmentations is called Mosaic Augmentation, it makes the model to learn objects in new different locations. To describe the efficiency of a classification algorithm, a confusion matrix is a table, an algorithm's effectiveness in classifying data is represented and summarized by a confusion matrix. So, after comparing the results of LPD of YOLO V5, YOLO V7, and YOLO V8. We saw that YOLOv8 is the most accurate version of all YOLO versions. this is the confusion matrix of the three YOLO Versions:



Confusion matrix of YOLO V5.

Confusion matrix of YOLO V7



Confusion matrix of YOLO V8

Conclusion:

ANPR technology recognizes the numbers and characters of license plate.
This technology works even in severe weather conditions, high illumination and the speed of vehicle on high ways. Automatic Number plate recognition (ANPR) technology
Is used globally to detect, deter and disrupt criminality, in order to solve crime, protecting officers, enforcing laws and making the roads safer. ALPR technology can safe people's life. ALPR cameras are effective, robust, reliable and provide more features to use. ALPR cameras can be mounted to any place to watch any vehicle and detect its License Plate. With the ability to capture nearly 60 License Plates per second. The ANPR technology is more than just a camera, it's an efficient system that can bring all information about any passing vehicle with effective results. This easy-to-use software provides analytics tools to benefit investigations. Such as identifying known associates by vehicle. Location surveillance for knowing information about criminal activity. Historical detection for analyzing best known locations of vehicles. ALPR technology can provide people better situational awareness with more accurate results to live safe and comfortably.
The system is implemented by using different versions of YOLO Algorithms such as YOLO V5, YOLO V7, and YOLO V8 to extract the license plate from the vehicle, then CNN model recognizes the characters and letters are included in the plates, the ALPR system in this project is implemented using Egyptian dataset and it is about 5505 after applying augmentation. The best performance was using YOLO V8 for license plate extraction and then CNN for recognition with accuracy 99.5%.

Future Work:

It will be great if we implement the ALPR system in real life using SCCD Camera to get high the quality of images and all the embedded system components for hardware implementation. Increasing Egyptian license plates dataset by applying different images of different vehicles not only cars and capture images in different environmental conditions so that getting high level of variety of images so that high accuracy will be. Experiment different techniques to reduce the training time and to get high accuracy. publish a paper that contain all the needed information about the project that implemented using Egyptian dataset

mirna-ALPR

ORIGINALITY REPORT

9%	2%	8%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Haitong Lou, Xuehu Duan, Junmei Guo, Haiying Liu, Jason Gu, Lingyun Bi, Haonan Chen. "DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor", Electronics, 2023 Publication	2%
2	Zainab Ouardirhi, Sidi Ahmed Mahmoudi, Mostapha Zbakh, Mohamed El Ghmary, Mohammed Benjelloun, Hamd Ait Abdelali, Hatim Derrouz. "Chapter 25 An Efficient Real-Time Moroccan Automatic License Plate Recognition System Based ontheYOLO Object Detector", Springer Science and Business Media LLC, 2023 Publication	2%
3	wael badawy, Rahma Tarek, Alaa Yousri. "Automated License Plate Recognition in Egyptian Environment: Lessons Learned", Institute of Electrical and Electronics Engineers (IEEE), 2022 Publication	2%

4	pdfs.semanticscholar.org Internet Source	1 %
5	www.theseus.fi Internet Source	1 %
6	medium.com Internet Source	1 %
7	Submitted to University of Limerick Student Paper	1 %

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		