CNN-Based Vehicle Rules Violation Detection System

A PROJECT REPORT

Submitted By

Khasrul Alam

ID No. 22209023

Reg No. 22209023

In partial fulfillment for the award of the degree

of

MASTER OF SCIENCE (ENGG.)

IN

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

FEBRUARY 2025

COMILLA UNIVERSITY, CUMILLA-3506

ACKNOWLEDGEMENT

First and foremost, I express my deepest gratitude to the Almighty Allah, whose boundless blessings and guidance have been my greatest strength throughout this journey.

I extend my sincerest appreciation to my respected supervisor, whose unwavering support, patience, and insightful guidance have been the cornerstone of this project. From the inception of the proposal to the final execution, her encouragement and expertise have been invaluable, making this endeavor not only possible but also a rewarding learning experience.

I am profoundly thankful to all the faculty members of my department, whose mentorship, constructive feedback, and continuous encouragement have greatly enriched my academic journey. Their wisdom and knowledge have nurtured an environment that fosters growth, innovation, and excellence.

Lastly, my heartfelt appreciation goes to my family, friends, and well-wishers, whose unwavering support, encouragement, and belief in me have fueled my determination. Their motivation has been a constant source of inspiration in overcoming challenges and achieving this milestone.

This project is the result of collective efforts, and I am truly grateful to everyone who has played a role directly or indirectly in its successful completion.

ABSTRACT

One of the most well-known components of the cutting-edge ai and machine learning sector that tracks numerous functional activities and their influence on the growth of modern smart cities worldwide is the video based intelligent transportation system v-its these days the high accident rate in a crowded metropolis is a major concern it is now crucial to address all of the causes of the significant increase in traffic accidents and the percentage of fatalities that are frequently attributed to breaking traffic laws. An enhanced V-ITS system for identifying and classifying cars, drivers, and their license plates while travelling on highways has been developed in order to overcome those difficulties. The technique is reliable for reducing the rate of violations because the upgraded V-ITS system can perform the aforementioned tasks with precision from real-time traffic footage.

This real-time detection approach quickly recognizes any breaches of yellow or red traffic signals a pretrained Convolutional Neural Network (CNN) featuring a three-layer to improve the feasibility of the vita system the structure was developed subsequently a model founded on a Deep Belief Network (DBN) was designed and developed.

The open cv yolov3 and other python tools as well as free internet video data sets are mostly used to complete the project the outcome of the real-time experiment of the v-its systems can applied to reduce vehicle violations ensuring of safety of both pedestrians and passengers.

LIST OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview:

Traffic offences are becoming increasingly serious in Bangladesh and around the world these days due to the growing number of cars in cities and metropolitan areas which can lead to heavy traffic congestion this results in significant property damage lost time and other mishaps that could put peoples lives in jeopardy. Systems for detecting traffic violations are desperately needed to address the terrifying issue and avoid such unimaginable repercussions traffic offences are becoming increasingly serious in Bangladesh and around the world. These days due to the growing number of cars in cities and metropolitan areas which can lead to heavy traffic congestion this results in significant property damage lost time and other mishaps that could put peoples lives in jeopardy systems for detecting traffic violations are desperately needed to address the terrifying issue and avoid such unimaginable repercussions due to the increasing number of cars in cities and metropolitan areas which can cause severe traffic congestion traffic crimes are getting more serious both in Bangladesh and globally significant property damage lost time and other accidents that could endanger life are the results of this to solve the horrifying problem and prevent such unthinkable consequences systems for detecting traffic infractions are critically needed.

1.2 Motivation and Purpose:

It has been discovered through the literature that it is extremely difficult to both recognize and track objects in video sequences because of the volume of information in the video object tracking can be a laborious procedure numerous background removal methods are available that are effective in both interior as well as outdoor surveillance systems according to the literature review in order to identify darkened regions julio et al [3] suggested a backdrop modelling technique and implemented a different algorithm however the object tracking computer has to pay for the shadow removal procedure by creating an effective algorithm that can accurately distinguish between the foreground element and background and eliminate incorrect foreground pixels from detection it will be preferable if the shadow can be eliminated during the foreground object detection process then no extra calculation are going to be needed for the identification and elimination of shadows the technique has an object tracking algorithm overhead.

The most actively researched topic in vision technology for both human beings and vehicles is surveillance footage. Here, the goal is to replace the obsolete, manual monitoring method with an intelligent visual surveillance system. The goal is to create a detector for motion and object tracking video surveillance system for vehicles such as cars, trucks, buses, bicycles, and more. The key principle supporting this topic is that it will include all of the physical duties -

1. All across globally, robotics and machine learning are in circulation.
2. Following and recognizing items saves worker time and increases usefulness.
3. It's important because it might make it easier to for individuals to remember small details about certain goods and cut down on worker labour.

The smart systems of today are further developed by machine recognition and retrieval.

1.3 Objectives:

The goal of this project is to automate the traffic signal violation detection framework, which will facilitate the road safety agency's ability to monitor traffic and swiftly and effectively take action against the owner of the violated vehicle. The technology's top goal is to precisely detect and track the vehicle and its activities. This work proposes a novel vehicle detection model based on YOLOv3 to address the issues with current vehicle detection, including poor speed, low detection accuracy, and lack of vehicle-type authentication. More in-depth research is done on the Vision-Based Intelligent Transportation Systems (V-ITS), planning for transit, and movement engineering applications for traffic image analysis and count, vehicle trajectory, tracking, flow, classification, density of congestion, auto velocity, traffic lane changes, number identification, etc. [1-4]

1.4 Expected Outcome:

In determining success, it would be impossible to gauge the worth of any undertaking. It must be focused on precise, quantifiable, and significant results. The two parts of the investigation are object recognition of achievement and locating and sorting.

1. Recognition of Achievement:

This is the method for determining the category of an object based on its dimensions along with other weighted attributes.

The exactitude and recall of the graph with the Collision Over Union (IOU) values subsequent to it has been generated after we tested our object detector on just a handful of visuals to see how accurately it was training.

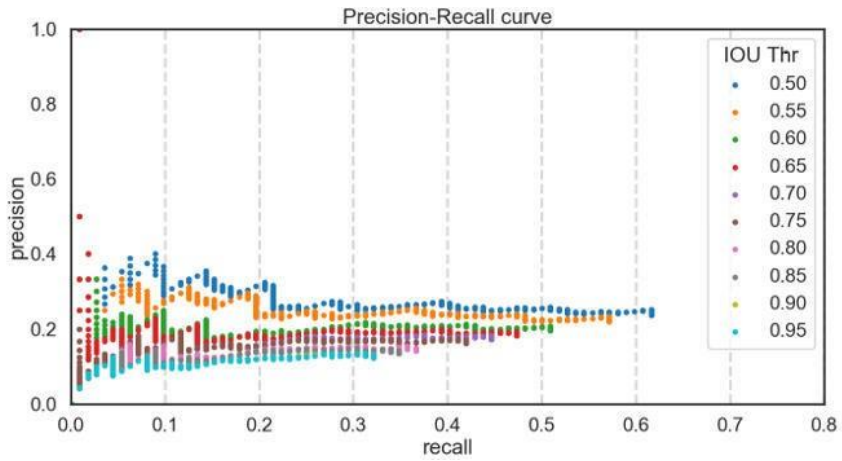

Figure 1.1: Precision recall graph with IoU threshold value

2. Locating and Sorting:

It is the last step in the object's detection and sorting process. After utilising YOLOv3 for object identification and arranging, respectively, the recorded footage will be separated into frames and saved with the recognising and sorting parameters gathered for every single input video.

Figure 1.2 : IoU threshold bounding box

IoU threshold: In the identification of objects, convergence over union is a the metric system that indicates the amount that an object's current and projected bounding boxes interact.

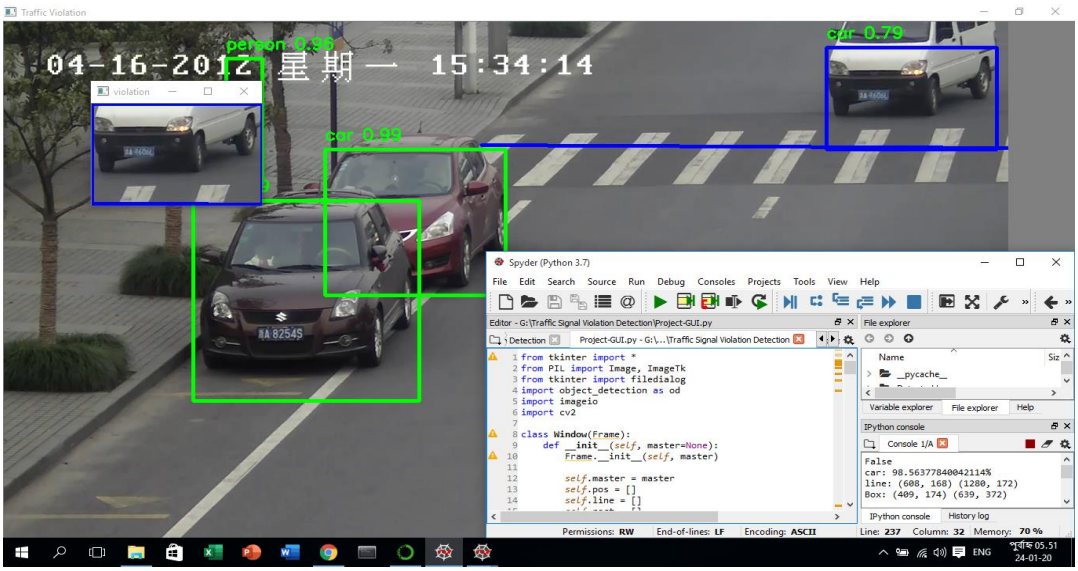

Figure 1.2: Expected output with detection and sorting

1.5 Report Layout:

The goal of the recognition of objects is to create a sense of a variety of linked tasks associated with computer vision, such as distinguishing things in digital snapshots and motion pictures.

The following subsection outlines each project chapter that leads to the final product's impeccable outcome. They can be described outlined below:

Chapter 1: This part of the text summarizes the project by discussing its motivation and objectives, which together support the realization of the expected outcomes

Chapter 2: This analysis showcases the background procedures alongside various comparisons, employing distinct graphical views for clarity.

Chapter 3: The approaches used to fulfill the project requirements are elaborated upon in this section.

Chapter 4: The following text outlines the different types of datasets that are integral to the successful completion of the project.

Chapter 5: In this part, the project's experimental evaluation is articulated through the use of various perspective video recordings, along with the results obtained.

Chapter 6: This section presents the outcomes of the a project's quantitative assessment which carried out using multiple types of aspect recordings of footage.

CHAPTER 2

BACKGROUND

2.1 Introduction:

The underlying process that is used to finish a project is covered in this part of the article. It is done using a variety of software setup multi-vehicle video datasets. Vehicles are detected, sorted, and another focused output is provided on this project at each stage. The entire process, which might aid in completing the task, consists of a number of linked works and a contrast of the results utilising multiple scopes and view films. Utilising the previously mentioned steps, the final and intelligent output has been produced, and this will further aid the Automated Violation Detection System profession.

2.2 Related Works:

The implementation of immediate traffic complaint identification in a monitoring stream is shown in [1], demonstrating the use of synchronous processing algorithms to take advantage of simultaneously video streams from multiple cameras. Another technique of putting real-time traffic violations detecting to work appeared in [2], where they employed a feature-based techniques tracking method to monitor moving cars after using video-based traffic surveillance with an enhanced background-updating methodology.

Although the aforementioned project served as inspiration for this one, a developed by itself methodology was used to carry it out. Vehicle detection is typically referred to as a substance recognition problem. The YOLOv3 model, which makes use of Darknet-53, is used to identify moving vehicular objects from the path of traffic. Violation situations are reviewed when vehicles have been detected.

2.3 Comparative Studies:

This section presents an analysis between previous and present-day versions of the the YOLO algorithm. There are not many variations between the two methods in terms of the fundamental process of identifying and classifying things such as devices, cars, wildlife, etc. The primary areas of variation were in the degree of precision level and the detection of small objects. Even though it is no longer the most precise method for the technique of detection, it is still an excellent option when continuous monitoring is required no significantly sacrificing efficiency.

1. YOLO version 2 algorithm analysis: YOLO commits a lot of mistakes when it comes to localisation. Additionally, YOLO has a comparatively low retention. Therefore, in order to maintain classification accuracy, the subsequent release of the algorithm is primarily focused on increasing recall and minimising localisation. To obtain greater functionality various significant strategies are applied.

2. High Resolution Classifier:

This is how the first YOLO stands was developed: The classifier network is initially trained at 224 x 224. After that, the sensor resolution was increased to 448. This implies that the neural network must concurrently transition to learning object detection and synthesise to the new input resolution when transitioning to diagnosis. In contrast, for YOLOv2, they first trained the model on 224x224 images before making a little tweak to achieve the desired performance of the classification network at a total resolution of 448x448 for 10 epochs on ImageNet. This is prior to testing for spotting. This gives the server a chance to set up its filters so they can function better with input from greater accuracy. The gain provided by this better resolution classification circuit is only roughly 4.5 percent mAP.

3. Batch Normalization:

Researchers achieve a mAP increase of more than 2% furthermore to performing batch normalisation on each convolutional layer within the YOLO stands method.

4. Convolution with anchor boxes:

This suggests that multiple objects are predicted for every class column. The YOLO stands method attempts to place the object in the grid cell that is located in the component's centre.

According to this theory, the red cell in the preceding images needs to identify several things simultaneously, but since each grid cell can only identify one object, an issue will arise. It has been attempted to use a k perimeter to allow the matrix cell to detect many objects in order to address this issue.

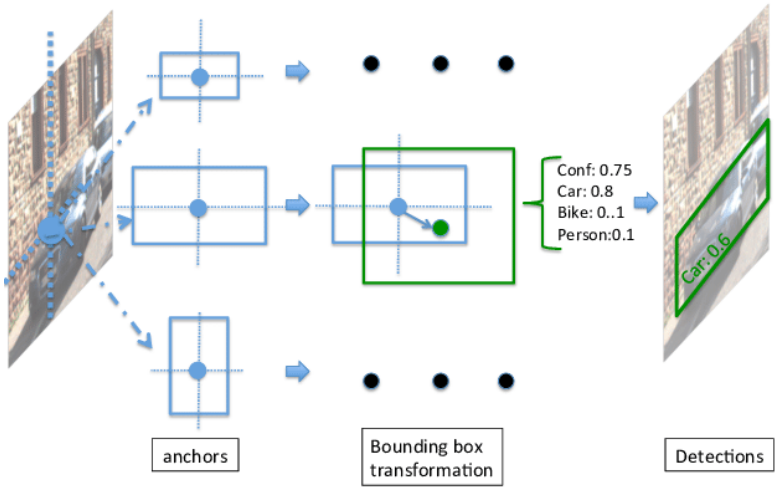

Figure 2.1 : Predict k bounding boxes YOLOv2 used the idea of Anchor boxes

Place values are predicted by YOLOv2 in relation to the surface of the grid cell's location. The fundamental reality is certain to fall anywhere from 0 to 1. For every cell, the network estimates five bounding boxes. For every surrounding box, it predicts five directions: tx, ty, tw, th, and to. The forecasts should match if the cell is shifted by (cx,cy) from the image's upper left corner and the perimeter of the prior (establish package) has the following dimensions: pw, ph.

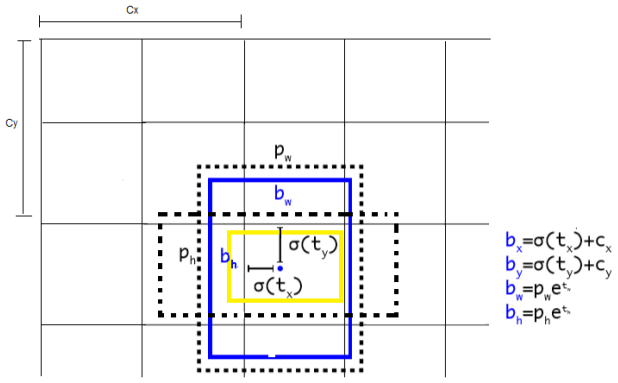

Figure 2.2 : Bounding box with dimension priors & location prediction

Then imagine just the box of blue. Unlike YOLO, which provides the estimated blue unit to that grid cell alone, YOLOv2 allocates the blue box to one of the anchoring boxes, which will be the one which has with the highest IoU with the ground truth box.The blue box is assigned to the outline of the grid and the anchoring area by YOLOv2 using the overhanging parameters.

The formula above can be used to determine the expected cell based on the explanation before:

Pr(object) * IoU(b, object) = sigmoid(t0)

Then imagine just the box of blue. Unlike YOLO, which provides the estimated blue unit to that grid cell alone, YOLOv2 allocates the blue box to one of the anchoring boxes, which will be the one which has with the highest IoU with the ground truth box.The blue box is assigned to the outline of the grid and the anchoring area by YOLOv2 using the overhanging parameters.

The formula above can be used to determine the expected cell based on the explanation before.

4. Comparison with others system: YOLOv2 represents a cutting-edge detection technique that outperforms other algorithms in a range of sets. Additionally, it can be used with different image sizes to offer an adequate compromise among precision and rapidity.

2.4 YOLO version3 algorithm analysis:

The Yolo v2 and v3 models differ slightly. Compared to Yolo v2, it is very precise as well as reliable. For every enclosing package, the network predicts four directions: tx, ty, tw, and th. If the box defining the previous has dimensions Pw, Ph, and the cell itself is offset from the upper left corner of the image by (cx,cy), the forecasts should match:

bx = sigmoid(tx) + cx
by = sigmoid(ty) + cy
bw = pw.e^(tw)
bh = ph.e^(th)

The bounding box prediction: Additionally, YOLOv3 uses logistic regression analysis for predicting an item score for every box that borders. This ought to be the case if the preceding boundaries box overflows an elementary object by crossing over the preceding bounding box. For instance, the one before it surpasses the primary ground truth object (which has the most IOU) by surpassing all other limit box priors, while preceding two overlaps another ground truth article. Each ground truth object is given a single bounding box previously by the system. There will be no loss for geometry or category projections when the boundary box prior is not applied to a factual object; rather the object is affected.

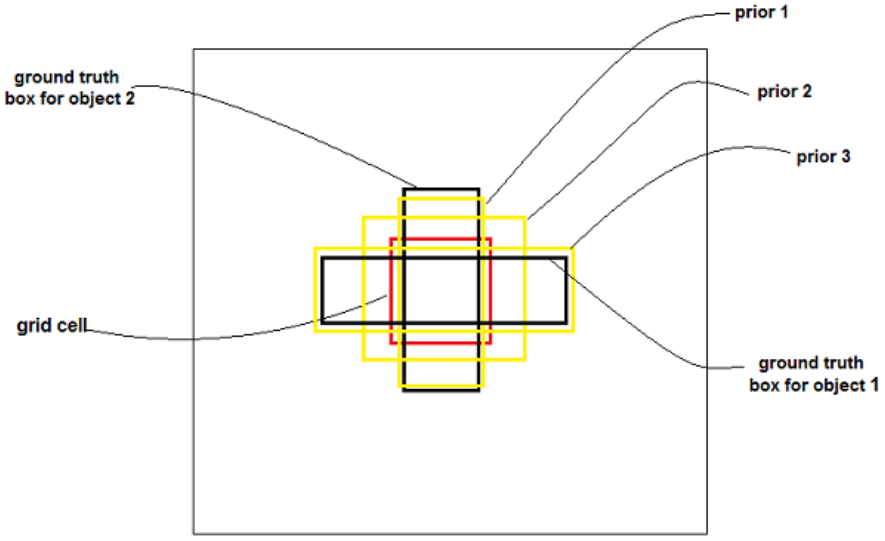

Figure 2.3: YOLO v3 bounding box prediction using ground truth value

We disregard the assumption whenever the box with the most IOU overlaps the actual element by beyond a certain amount. The minimum requirement in this portion is 0.5.

Prediction of multiple classifications: A component may have several labels on certain records, such as the Open Source Vision Set; for instance, it may be labelled as both a woman and someone.

It contains a large number of repetitive labels. It is assumed that every box has exactly one class when using a softmax technique for class anticipating, but this is frequently not the case (as in Open Image Set). The data has been superior modelled using a multilabel technique.

Small objects detection: YOLO had trouble handling little items. However, YOLOv3 performs better for small objects, and this is because it uses quicker connections. We may extract more particulars from the prior map of features by using this association approach. But YOLOv3 does not perform more effectively on larger objects than the previous iteration.

Predictions Across Scales: YOLOv3 estimates compartments at every scale, and are shown in the above image, in contrast to YOLO and YOLO2, whose only forecast results at the layer that follows.

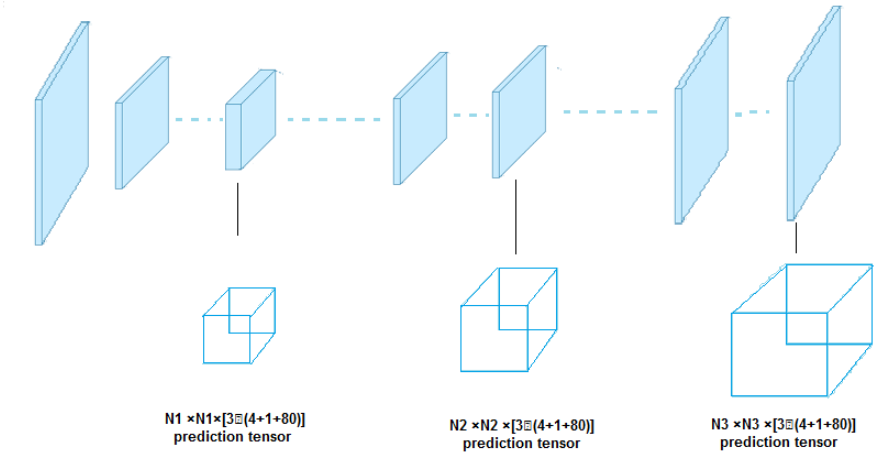

Figure 2.4: A simple diagram for YOLOv3 network

Comparison with other versions: The YOLOv3 framework, the architecture incorporates three anchor boxes at each scale, which significantly enhances its ability to predict three bounding boxes for each grid cell. This multi-box prediction capability is crucial for accurately detecting objects of varying sizes and shapes. This limitation helps streamline the detection process and reduces potential confusion in object localization.

Although YOLOv3 is still faster than other identification algorithms, its performance decreases noticeably as the IoU threshold rises (IOU = 0.75), suggesting that it has more difficulty aligning the shapes with the thing being detected as YOLOv2.

2.5 Scope of the Problem:

Machine learning's reach is always expanding. However, classification complex is rising quickly as well as it. In this day and age, the work uses a system for vehicle detection that combines machines learning along with selected retrieval. When compared to other outdated techniques, this sort of detection provides a naturalised architecture without sacrificing effectiveness in Unnamed Aerial Video (UAV) or images.

By simply altering the initial division value, required proposals size, and limit in the particular search algorithm, in relation to the image's initial resolution and the categorisation of vehicles identified, this resolved approach can be extended to an unidentified drone video or image. The suggested method's topology is sufficiently transparent to run on low-end computers. Since this approach is considerably simpler to put into practice, it may also be simpler to enhance the system for a variety of uses.

The concept of machine learning is not unique. Because contemporary computer systems consume a lot of information and consciousness, the system is experiencing a renaissance. The technique makes it very simple to make adjustments in the area of moving and motionless item recognition and identification from different angles.

Challenges:

The reliability and precision of current video-based devices may be reduced due to their sensitivity to environmental factors like weather and luminosity. The primary difficulties with vehicle identification and section implementations are covered in the following section. [7-10]:

1. **Cameras Motion:** Video processing can be difficult when there is any activity in the sensor's field of view and the footage was taken by brittle or vibrating cameras. blurred motion in video vision is typically the result of this phenomenon, making detection and decisions more difficult. By temporally de-blurring a single motion blur kernel for the entire image, motion blur can be prevented. Environmental factors like breezes, storms, or other emergencies are the cause of this occurrence.

2. **Camouflage:** Generating accurate recognition and selection is a difficult task because some cars' views are identical to the still landscape (such as a grey car travelling on a road in poor light). This is particularly crucial for tracking applications. Thus, identify is an exceptionally difficult challenge for behavioural differencing techniques. There are several geometric morphological techniques that can be used to reduce this difficulty.

3. **Congested Traffic Conditions:** Two common occurrences in transportation photographs of cities and metropolitan areas are vehicle occlusion and traffic congestion.

When a vehicle passes behind other vehicles in relation to a road, vehicle occlusion can happen at any time video. Both partial and impartial forms of blockage may have an impact on the underlying frame enumeration process. In situations where a vehicle recognition system relies on motion data, the result may be significantly impacted by congestion.

4. **Different Formation of Vehicles:** Because here are so numerous different kinds of transportation, some can be incorrectly labelled if a detection system is developed that classifies vehicles based on visual cues like headlights and borders. The most prevalent problem in supervised learning detection of vehicles is this one.

5. **Low Light Conditions:** Cars cannot be easily identified by their apparent forecasts in darkness, such as tunnels or inclement weather; only the headlights can be identified in these conditions. Therefore, additional areas that are either dark or only faintly exposed to light will seem to be muddy in colour. The processes of perception and identification may be impacted by this lack of visual outlooks. The recognition of lamp pairs to be regarded as a separate car is the other difficulty in this situation. Binary image conversion with a suitable threshold figure can be used in these kinds of settings. Vehicle headlights and poor lighting, which can produce persistent noise, can also make detection more difficult in nighttime videos.

6. **Noise**: As noise levels rise, video signals typically become distorted. Video surveillance background removal techniques must contend with such distorted signals impacted by various noises, including compression artefacts and sensor distortion. Both the identification and selection tasks can be impacted by noise.

CHAPTER 3

METHODOLOGY

3.1 Vehicle Classification:

Vibrating items are identified from the provided video material. These things that move are categorised into the appropriate classes using the YOLOv3 object detection-based paradigm. The third object detection in the You Only Look Once algorithm is called YOLOv3. It is better at detecting things and has improved accuracy with numerous processes. It is built using the Darknet-53 technology for the classifier model. The architecture of the artificial neural network is displayed in Table 1.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
	Convolutional	32	1 × 1	
1×	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
	Convolutional	64	1 × 1	
2×	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
	Convolutional	128	1 × 1	
8×	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
	Convolutional	256	1 × 1	
8×	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
	Convolutional	512	1 × 1	
4×	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Table 1. **Darknet-53.**

Features:

1. **Bounding Box Predictions:** Since YOLOv3 is one network, the classification and objectiveness losses must be determined independently from the exact same network. YOLOv3 uses the logistic regression method to estimate the accuracy score, where 1 denotes full bounding box agreement over the ground truth object. For a single ground truth object, it will only predict one bridging box beforehand; any inaccuracy in this will result in both recognition and identification loss. Other bounding box priors that have an objectiveness level higher than the threshold but lower than the best one would also exist. Those mistakes won't occur for classifier loss; they will only occur for the identification loss.

2. **Class Prediction:** Individual logistic classifications are used by the YOLOv3 algorithm for every class variation in a standard soft-max layer. This is done in order to create a classification with several labels system. Each box uses multilabel classification to anticipate which classes within it might include.

3. **Predictions across scales:** The method YOLO predictions cases at three distinct scales to aid recognition on a variety of dimensions. Following that, employing a technique similar to feature triangle associations, values are gathered from each scale. With the aforesaid strategy, YOLOv3 acquires the capacity to make better predictions at different scales. There are three limiting box priors per scale, for a total of eight surrounding box prior knowledge, after the boundary box priors produced using dimension groupings are validated into three categories.

4. **Feature Extractor**: The sophisticated Darknet-53 design, featuring 53 convolutional layers of data, is used by YOLOv3. Compared to YOLOv2, this structure is deeper and includes bypass or leftover connections. For this reason, it has more power than Darknet-19 and is more efficient than ResNet-101 along with ResNet-152.

3.2 Vehicle Detection:

The YOLOv3 model is used for identifying the cars. Violation cases are examined when the vehicles have been identified. In the user-provided video look over, a traffic queue moves across the road. The traffic signal is red, as indicated by the vertical line. Each vehicle that violates the traffic line in red is in contravention.

The circumference of the discovered objects is green. A violation occurs when a car violates an illuminated light in red. When an offence is detected, the vehicle's surrounding perimeter turns red.

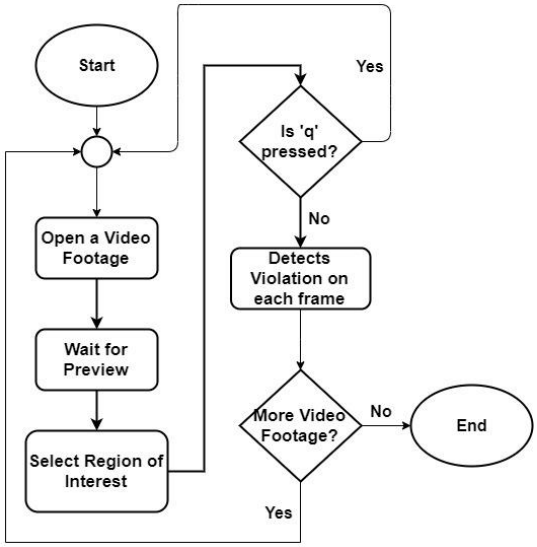

Figure 3.1: Flow-diagram of violation detection using each video frame

The perimeter of the box and that vehicle's distinctive identification number are used to identify the automobiles from each video frame. To choose the area of passion, a line is first drawn. Any object that crosses the line is flagged as illegal and is indicated with a red bounding box. Following detection, each vehicle is caught by a second sensor and broadcast to a different windows screen.

3.3 YOLO v3 network:

YOLO is a fully convolutional network (FCN) since it exclusively uses convolutional layers. The task's weighted file contains convolutional properties that allow the CNN to function and finish the execution of the application process.

The convolution weights begin loading when the program launches, and once all of the weights have been loaded, the software outputs a picture. Visual input is then provided to test the application's functionality.

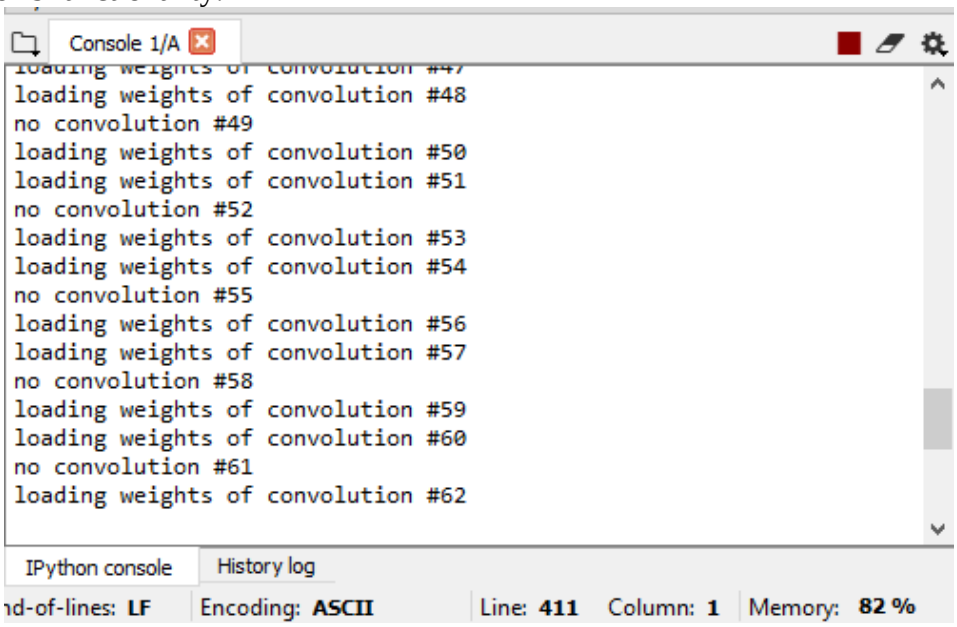

Figure 3.2 : Loading weights of convolution while running the program

3.4 Invariant to the size of input:

YOLO is independent from the image's input size. Given a number of issues, we might want to maintain a consistent input size in actuality.A number of these issues is that any of the files must have the same resolution if we wish to process them in batches. This is because the graphics processor (GPU) may handle the data in batching in parallel, which increases productivity. Concatenating several photos into a big batch requires this.

The step of the neural network is the method by which a picture is down-sampled by a parameter.For instance, an input image of size 390 × 390 will produce a product of size 13x13 if the network's diagonal step is 3o.

The forward motion of any layer of data in the system is typically similar to the factor that makes the layer's produce larger than the system's initial picture.

3.5 Analyzing an Example:

A cell in a grid is responsible to recognise a substance if its centre or halfway is inside the grid square.

The final two dimensions related to the form (19, 19, 5, 85) encoding will be flattened or simplicity's sake. Thus, (19, 19, 425) is the Deep CNN's outputs.

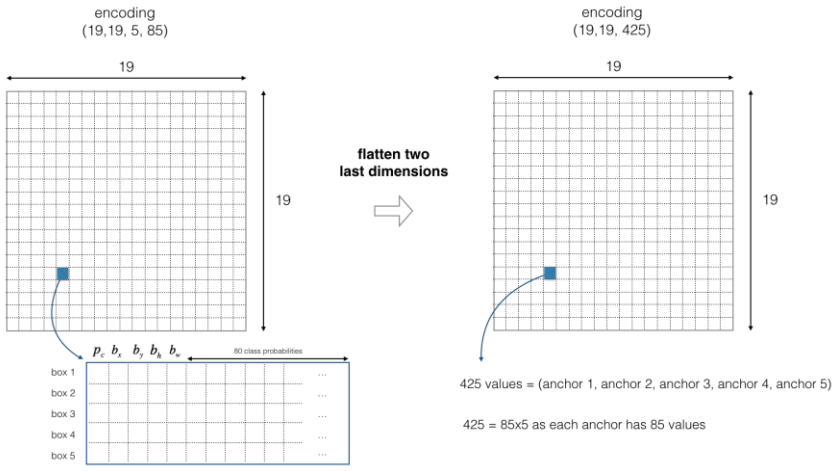

Figure 3.3: Detection of object by centering into grid cell

3.6 Output Prediction:

We are going to find the subsequent bits by product for each box (of each cell) and derive a probability that has a specific class.

A total of 80 probability classes exist. It is then multiplied by the surrounding box pixel to find the anticipated boxes. The boxes that have been identified are bx, by, bh, and bw.

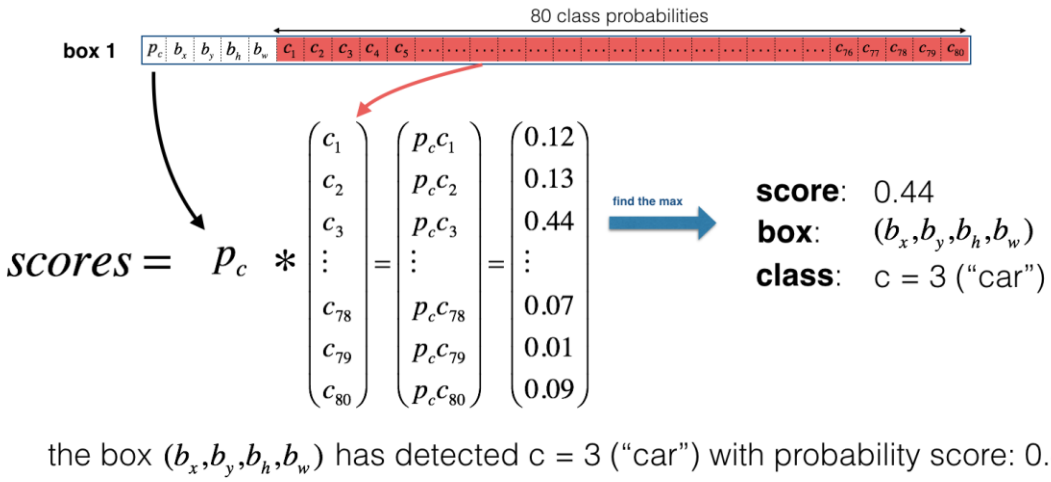

Figure 3.4: Detected image with probabilistic bounding box

Ultimately, a probability value of 0.44 shows it is an automobile, as the detected value of c = 3 fits with one likelihood classification.

During the process, the sigmoid algorithm is used to alter the network output. The most crucial components for the outcome are the grid cell centres and actual boundaries.

CHAPTER 4

VEHICLE VIDEO DATASETS

Despite the widespread installation of video surveillance systems on roads around the world, traffic photos are rarely made public because of copyright, anonymity, and safety considerations. Three groups can be distinguished from the perspective of collecting pictures in the traffic image collection.: [11]

1. Visual data recorded by the automotive camera.
2. Images recorded by the observation camera.
3. Visual data collected from cameras that are not designed for surveillance.

4.1 Surveillance camera datasets:

Typical roadside surveillance cameras record film to various locations, at various times, and with various circumstances. The information sets are categorised here by vehicle type, such as buses, cars, trucks, motorcycles, bicycles, people, etc.

1. **Data sets of roads:** Data sets have been assembled from various surveillance cameras operating in two unique areas.

Figure 4.1: Datasets of roads from surveillance camera

2. **Datasets of buses**: The bus records are gathered from both forward and reverse viewpoint regions using a different surveillance camera.

Figure 4.2: Datasets of buses from surveillance camera

3. **Data sets of motorbike & cars:** Cars as well as motorbikes are among the other significant vehicles that are most frequently seen on surveillance footage. To identify automobiles, two different video databases' worth of film are gathered.

Figure 4.3: Datasets of cars and motorbikes

The pictures have a resolution of 1920*1080 and are in RGB format. The collection includes vehicle items with significant scale variations because we tagged the more compact objects in the immediate traffic area.

Training records, labelled datasets, and testing data sets are the three groups into which the data is separated. The training data sets are used in this research.

CHAPTER 5

IMPLEMENTATION

A variety of software and Python tools are utilised to complete the project. The steps involved in carrying out this project are listed below:

5.1 Anaconda Platform:

It should be installed first. The simplest method for performing Python/R data science and machine learning on Linux, Windows, and Mac OS X is the open-source Anaconda Distribution. It is an accepted preferred method of training, testing, and development on a single system, and it has more than 19 million subscribers globally.

5.2 Anaconda Navigator:

The Anaconda package comes with a desktop graphical user interface (GUI) called Anaconda Navigator that enables users to operate conda-packages, environments, and networks as well as run programs without the need for command-line procedures. Navigator may install packages in an environment, run them, and change them after searching for them on Anaconda Cloud or in a local Anaconda directory. It is compatible with Linux, macOS, and Windows.

5.3 Jupyter Notebook:

Navigator comes with the programs by definition. Programming in many languages is made possible by Jupyter Notebook's ability to connect to multiple kernels. The Jupyter Notebook comes pre-installed with the I-Python kernel. There are currently 49 Jupyter-compatible kernels for a variety of computer languages, including Python, R, Julia, and Haskell, as of the 2.3 release [12–13] (October 2014).

5.4 Spyder:

Spyder is a cross-platform, open source development platform (IDE) for Python scientific programming. Numerous well-known packages in the scientific Python stack are integrated with Spyder.

5.5 Installation of packages: In order to finalize the project, several Python 3 packages have been installed on the Anaconda platform. The specific packages are listed below:

1. NumPy
2. SciPy
3. MatPlotlib
4. I-Python

5. Keras&
6. Tensor flow backend

5.6 Computer Vision:

This project uses OpenCV, an open source library of computer vision and machine learning software, for image processing purposes. The car classifiers with darknet-53 is implemented using a Tensor-Flow backend.

5.7 YOLOv3:

This project leverages a specific algorithm from the Open Computer Vision library, which is an integral part of the esteemed You Only Look Once (YOLO) family of object detection models. In particular, we are utilizing version 3 of this algorithm, which has gained recognition for its remarkable speed and accuracy in executing real-time object detection tasks.

CHAPTER 6

EXPERIMENTAL RESULTS & ANALYSIS

Following are the final results, which are primarily the output scenarios. We outline the approaches' performance testing in the following section. The vehicle object collection created in the vehicle video dataset portion was used for our experiments. High quality highway footage were used in our experiment to identify and choose automobiles. It is crucial to complete all required steps in a sequential manner. Weight Convolution, Region of Interest, GUI, and, lastly, bounding box detection.

6.1 Weight Convolution:

The yolo.weight file serves as the foundation for the convolution. The loading weights are complicated after the Tensor-Flow backend has started up.

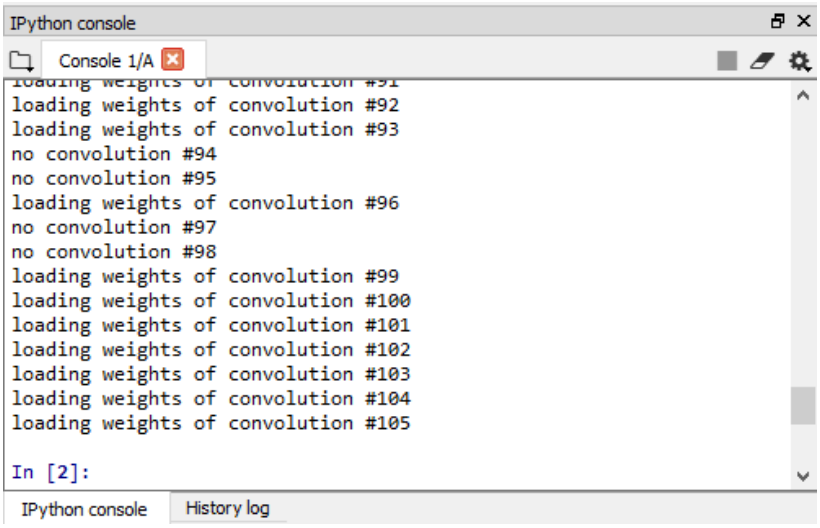

Figure 6.1: Convolution of weighted value

6.2 Graphical User Interface:

This section is designed to facilitate the straightforward acquisition of input images or videos specifically for the purpose of detection. It serves as a crucial starting point in the workflow. Once the convolution of weights has been successfully completed, the detection process will commence.

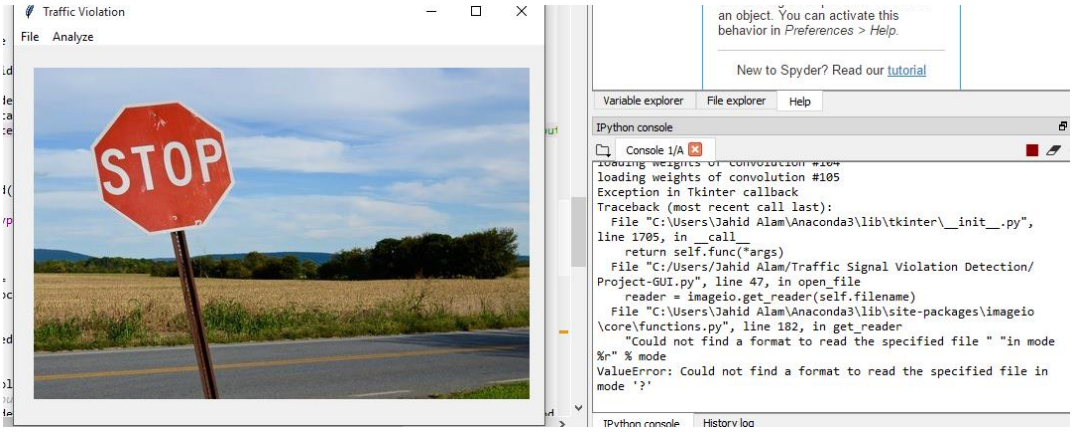

Figure 6.2: Initial GUI output

Each of the software's features are available through the graphical user interface. The program is used for debugging and management. Code editing is not required for any kind of management. For instance, we may use the open item to open any video clip that we need to. The File & Analyse option is shown in the representation. which will be used to take the next step.

6.3 Region of Interest:

Following the aforementioned process, we input a video file and begin the analysis by looking at the leftmost portion of the input video. After that, select the area of interest from the list of areas indicated by the blue line.

Whenever a vehicle breaches the line, this is the crucial point where the infraction is discovered. On the road, build a traffic line. The administrator's designed traffic line will serve as a traffic signal line. We must choose the "Region of interest" item from the "Analyze" option in the above picture in order to activate the line drawing capability. The system administrator must next choose two locations to draw a line that indicates the traffic signal.

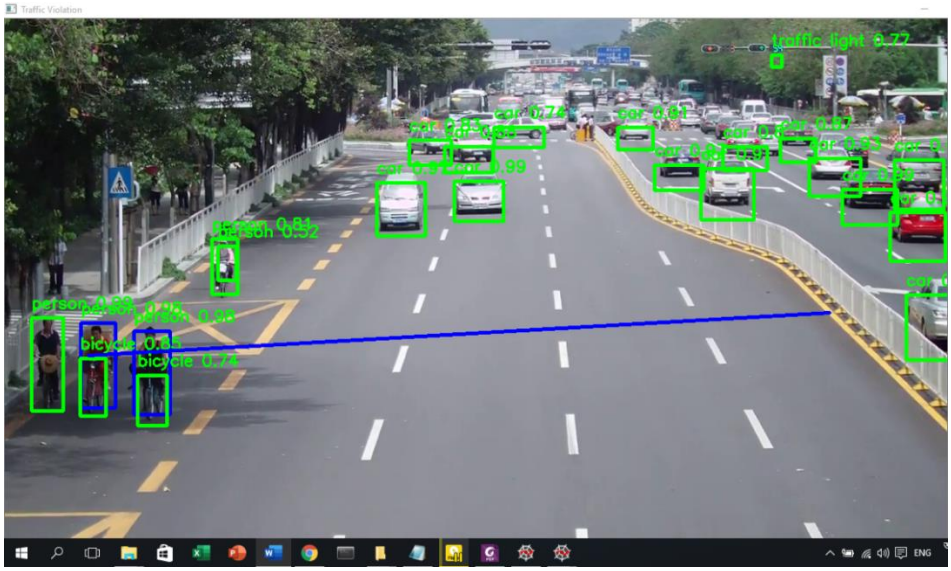

Figure 6.3: Region of interest (blue line) output

6.4 Detection of various vehicle:

It is the end result of using several camera inputs recognise a vehicle in context. Here, we use the entire process that we previously outlined to illustrate the detection and selection of different vehicle kinds. Using the box framework, it includes several categories of vehicles and objects, such as buses, automobiles, motorcycles, bicycles, and men (both drivers and passengers).

1. **Output of multiple vehicles:** It is difficult to identify more than one vehicle at once. The process becomes more complicated since a large number of box borders and labelled data are needed to do this.

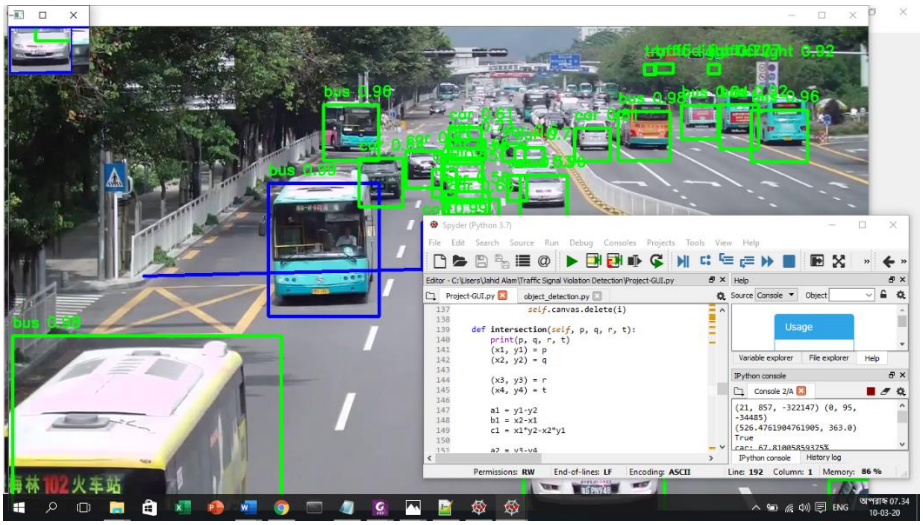

Figure 6.4: Multiple vehicles detection and selection

2. Output of cars:

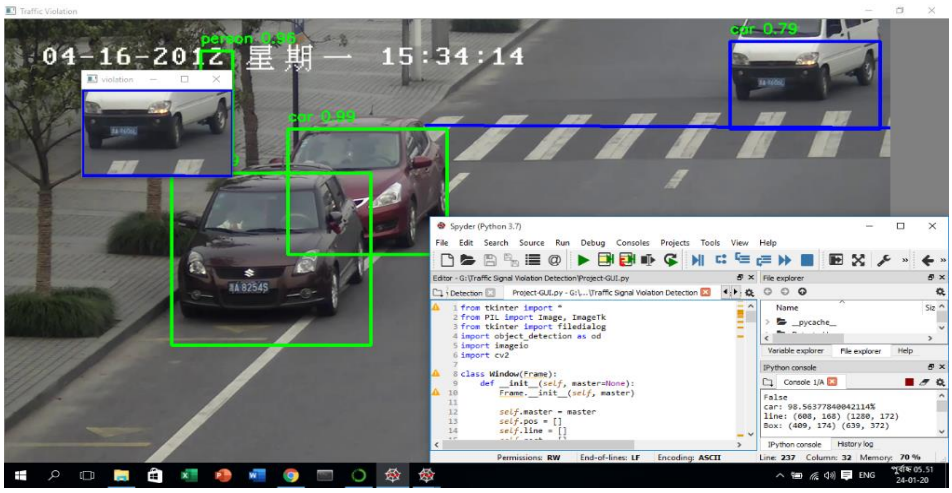

Figure 6.5: Car detection and selection

3. Output of pedestrians and bicycles:

Figure 6.6: Pedestrians and bicycle detection

4. Output of motorbikes:

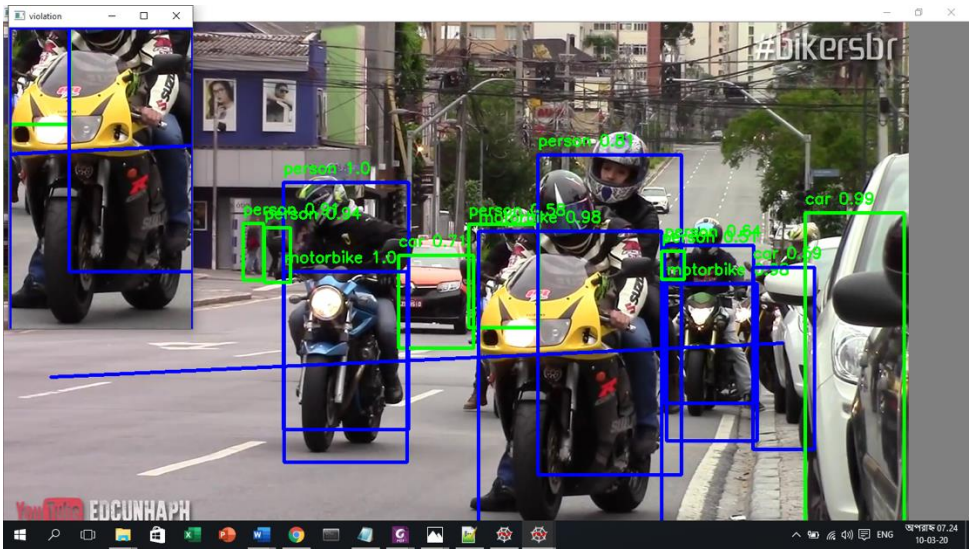

Figure 6.7: Motorbikes detection and selection

6.5 Analysis:

Once the entire project is finished, it is simple to explain that a moving object will be identified from any angle and under all circumstances. Perhaps it's low lighting, or perhaps during emergencies, the camera may be bright and the footage may be distorted. Multi-object tracking and other advanced vehicle object identification applications are also crucial ITS tasks [14]. Detection-Based Tracking is used in the majority of multi-object tracking techniques (DBT). The detection system will be created quickly by completing all of these training and detection tasks. The entire process will then perfectly achieve the anticipated outcome. As a result, it can be said that research on deep convolutional network techniques has replaced previous approaches for vehicle object detection. There will be more analysis.

CHAPTER 7

CONCLUSION & FUTURE SCOPES

7.1 Conclusion:

A project's future is largely determined by its ability to effectively cover its entire area. In this study, we endeavored to examine all relevant aspects of the region in question. We undertook an analysis of several datasets while accounting for various difficult environmental conditions to yield optimal results. The final output demonstrates the successful detection and classification of all moving objects, especially vehicles. This detection process is instrumental in uncovering effective solutions to illegal activities. The operation of this mechanism is characterized by a high degree of confidence and accuracy on congested highways, which greatly benefits law enforcement agencies. Its ability to function effectively in high-traffic situations ensures that violations are detected without delay, allowing for prompt action. Although this technology is just beginning to evolve, its future potential is substantial, with ongoing research and development likely to yield even more sophisticated capabilities. In summary, we believe it will lead to a marked decrease in traffic violations, enhancing the safety of both drivers and pedestrians. By fostering a culture of compliance with traffic laws, this mechanism could play a crucial role in reducing accidents and improving overall road safety.

7.2 Future Scopes:

1. Future research in this sector will focus on enhancing the accuracy of vehicle count and speed measurements.

2. The calculation of both the velocity and acceleration of a moving object can be accomplished with relative ease.

3. This feature will serve as a fundamental element in the reduction of both traffic jams and noise pollution.

4. The execution of this project could lead to the removal of vehicle obstruction and the optimization of traffic management practices.

5. The implementation of the project's extended format greatly simplifies the process of tracking high-speed vehicles and determining the appropriate lanes for each one.

6. The ability to counteract illegal activities is instrumental in fostering the creation of automated vehicles that promise greater safety in the near future.

REFERENCES

[1] H. Chung-Lin and L. Wen-Chieh, "A vision-based vehicle identification system," in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, pp. 364-367 Vol.4.

[2] Z. Wei, et al., "Multilevel Framework to Detect and Handle Vehicle Occlusion," Intelligent Transportation Systems, IEEE Transactions on, vol.9, pp. 161-174, 2008.

[3] N. K. Kanhere and S. T. Birchfield, "Real-Time Incremental Segmentation and Tracking of Vehicles at Low Camera Angles Using Stable Features, "Intelligent Transportation Systems," IEEE Transactions on, vol. 9, pp. 148-160, 2008.

[4] N. K. Kanhere, "Vision-based detection, tracking and classification of vehicles using stable features with automatic camera calibration," ed, 2008, p. 105.

[5] G.Ou, Y. Gao and Y. Liu, "Real Time Vehicular Traffic Violation Detection Traffic Monitoring System," in 2012 IEEE/WIC/ACM, Beijing,China, 2012.

[6] X. Wang, L.-M. Meng, B. Zhang, J. Lu and K.-L.Du, "A Video-based Traffic Violation Detection System," in MEC, Shenyang, China, 2013.

[7] Shaikh, S.H.; Saeed, K. and Chaki, N. (2014) "Moving Object Detection using Background Subtraction," SpringerBriefs in Computer Science, IS SN: 2191-5768.

[8] Lan, J.; Li, J.; Hu, G.; Ran, B. and Wang, L. (2014) "Vehicle Speed Measurement based on Gray Constraint Optical Flow Algorithm",International Journal for Light and Electron Optics, 125 (1), pp.289-295.

[9] Dueholm, J.V.; Kristoffersen, M.S.; Satzoda, R.K.; Ohn-Bar, E.; Moeslund T.B. and Trivedi, M.M. (2016) "Multi-perspective Vehicle Detection and Tracking: Challenges, Dataset, and Metrics," Proc.Of the IEEE 19th International Conference on Intelligent Transportation Systems, Rio de Janeiro, pp. 959-964.

[10] Cho, S.; Matsushita, Y. and Lee, S. (2007) "Removing Non-Uniform Motion Blur from Images," Proc Of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, pp. 1-8.

[11] Luo, Z. (2018). Traffic analysis of low and ultra-low frame-rate videos, Doctoral dissertation. Université de Sherbrooke.

[12] "Reviews for spyder". apps.ubuntu.com. Retrieved 9 February 2014

[13] "Seznámení s Python IDE Spyder | Fedora.cz". fedora.cz. Archived from the original on 20 August 2013. Retrieved 9 February 2014.

[14] Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., Kim, T.K.(2014). Multiple object tracking: A literature review. arXiv preprint arXiv:1409.7618.