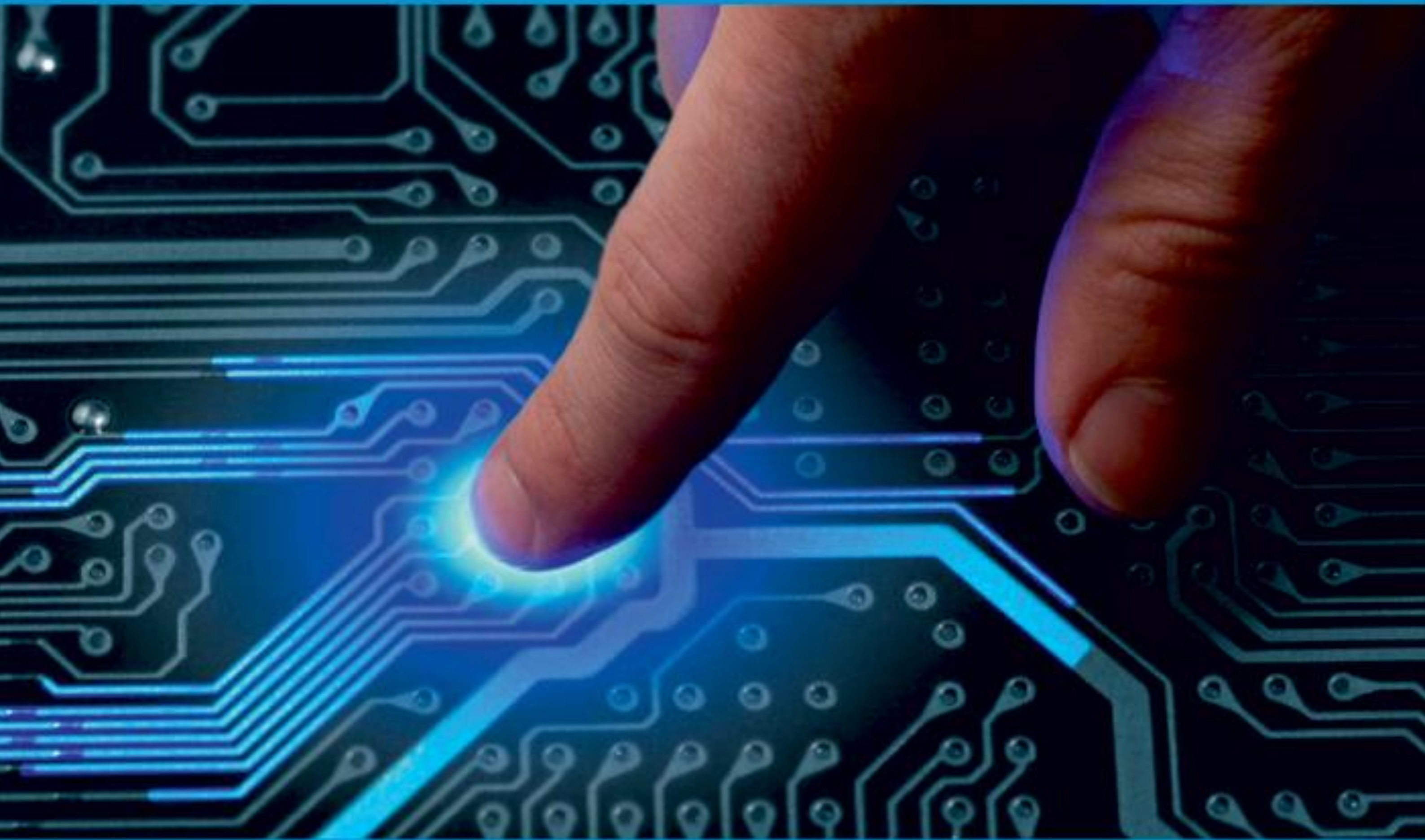
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |

Video Processing Based Tracking and Vehicl Edetection Using Yolo V5

Gaurav Lohakare, Prof Pushpa Tandekar, Prof Vijay Rakhade

Department of Computer Science & Engineering, Shri Sai College of Engineering Technology, Chandrapur, India

Department of Computer Science & Engineering, Shri Sai College of Engineering Technology, Chandrapur, India

Department of Computer Science & Engineering, Shri Sai College of Engineering Technology, Chandrapur, India

ABSTRACT: Automatic Number Plate Detector is also called License Plate recognition using image processing methods .In the proposed algorithm an efficient method for recognition of vehicle number plates has been devised. This system contains of three main parts namely number plate detection, plate character segmentation and character recognition. Number plate detection is the first process that takes place and has been said to have a lot of complications due to vehicle in motion, complex background, distance changes and weather conditions etc. We use YOLO to detect the number plate from the images, the focus of number plate detection is to find the plate region on an image and all the other processes like recognition they all dependent on the detection. Recognition involves using neural network were each character extracted from the image. Our output will include text of the recognized number plate from the given images.

KEYWORDS: Object detection; YOLO v5; Open CV; image processing, Neural Network

I. INTRODUCTION

Vehicle recognition and monitoring are gaining importance in traffic management. However, due to the various sizes of cars, detection remains a difficulty, which directly impacts the accuracy of vehicle counts. The suggested vehicle recognition and counting method first extracts the highway road surface in the image and divides it into a distant regions. A newly developed segmentation strategy in the proposed vehicle identification and counting system first extracts a nd s eparates t he highway road Turface in the image into a distant region and a proximal area; the method is crucial for improving vehicle detection. The aforementioned locations are then sent to the YOLOv5m network to determine the vehicle's kind and position. Finally, the ORB algorithm is utilized to create vehicle trajectories, which may be used to estimate the driving direction of the vehicle and determine the number of distinct cars. Several traffic surveillance recordings from various settings are utilized to validate the suggested meUhodoMoHZ. 5he experimental findings demonstrate that the suggested segmentation approach may give greater detection accuracy, particularly for the detection of little automobile things. In addition, the vehicle detection performance was significantly improved by 99.39 % of m AP compared to the Yolov5 basic. This work has broad practical implications for managing and controlling vehicle objects in traffic scenes.

Deep learning is a branch of machine learning which is based on artificial neural networks. It is capable of learning complex patterns and relationships within data. In deep learning, we don't need to explicitly program everything. It has become increasingly popular in recent years due to the advances in processing power and the availability of large datasets. Because it is based on artificial neural networks (ANNs) also known as deep neural networks (DNNs). These neural networks are inspired by the structure and function of the human brain's biological neurons, and they are designed to learn from large amounts of data.

1. Deep Learning is a subfield of Machine Learning that involves the use of neural networks to model and solve complex problems. Neural networks are modeled after the structure and function of the human brain and consist of layers of interconnected nodes that process and transform data.
2. The key characteristic of Deep Learning is the use of deep neural networks, which have multiple layers of interconnected nodes. These networks can learn complex representations of data by discovering hierarchical patterns and features in the data. Deep Learning algorithms can automatically learn and improve from data without the need for manual feature engineering.
3. Deep Learning has achieved significant success in various fields, including image recognition, natural language processing, speech recognition, and recommendation systems. Some of the popular Deep Learning architectures

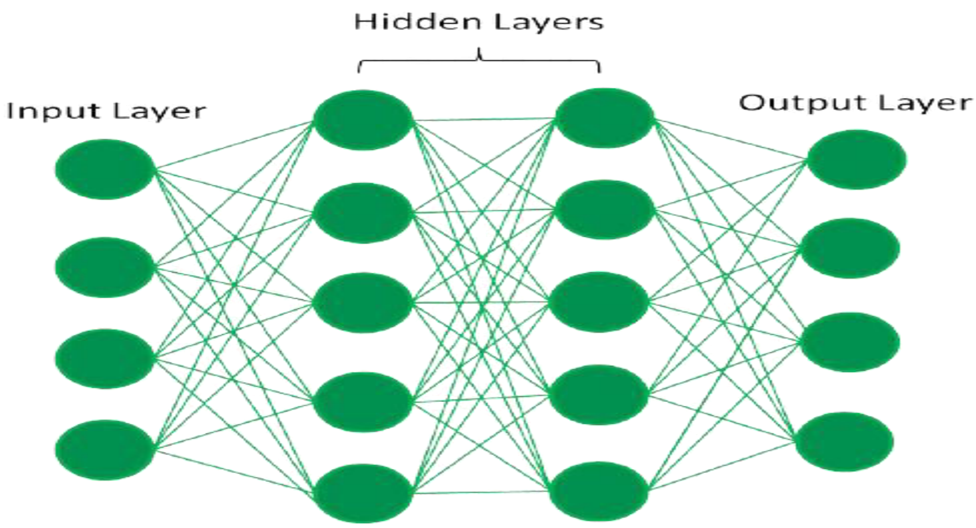
include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Belief Networks (DBNs).

4. Training deep neural networks typically requires a large amount of data and computational resources. However, the availability of cloud computing and the development of specialized hardware, such as Graphics Processing Units (GPUs), has made it easier to train deep neural networks.

In summary, Deep Learning is a subfield of Machine Learning that involves the use of deep neural networks to model and solve complex problems. Deep Learning has achieved significant success in various fields, and its use is expected to continue to grow as more data becomes available, and more powerful computing resources become available.

Artificial Neural Networks

Artificial neural networks are built on the principles of the structure and operation of human neurons. It is also known as neural networks or neural nets. An artificial neural network's input layer, which is the first layer, receives input from external sources and passes it on to the hidden layer, which is the second layer. Each neuron in the hidden layer gets information from the neurons in the previous layer, computes the weighted total, and then transfers it to the neurons in the next layer. These connections are weighted, which means that the impacts of the inputs from the preceding layer are more or less optimized by giving each input a distinct weight. These weights are then adjusted during the training process to enhance the performance of the model.



Fully Connected Artificial Neural Network

Artificial neurons, also known as units, are found in artificial neural networks. The whole Artificial Neural Network is composed of these artificial neurons, which are arranged in a series of layers. The complexities of neural networks will depend on the complexities of the underlying patterns in the dataset whether a layer has a dozen units or millions of units. Commonly, Artificial Neural Network has an input layer, an output layer as well as hidden layers. The input layer receives data from the outside world which the neural network needs to analyze or learn about.

In a fully connected artificial neural network, there is an input layer and one or more hidden layers connected one after the other. Each neuron receives input from the previous layer neurons or the input layer. The output of one neuron becomes the input to other neurons in the next layer of the network, and this process continues until the final layer produces the output of the network. Then, after passing through one or more hidden layers, this data is transformed into valuable data for the output layer. Finally, the output layer provides an output in the form of an artificial neural network's response to the data that comes in.

Units are linked to one another from one layer to another in the bulk of neural networks. Each of these links has weights that control how much one unit influences another. The neural network learns more and more about the data as it moves from one unit to another, ultimately producing an output from the output layer.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |

Types of neural networks

Deep Learning models are able to automatically learn features from the data, which makes them well-suited for tasks such as image recognition, speech recognition, and natural language processing. The most widely used architectures in deep learning are feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs).

Feedforward neural networks (FNNs) are the simplest type of ANN, with a linear flow of information through the network. FNNs have been widely used for tasks such as image classification, speech recognition, and natural language processing.

Convolutional Neural Networks (CNNs) are specifically for image and video recognition tasks. CNNs are able to automatically learn features from the images, which makes them well-suited for tasks such as image classification, object detection, and image segmentation.

Recurrent Neural Networks (RNNs) are a type of neural network that is able to process sequential data, such as time series and natural language. RNNs are able to maintain an internal state that captures information about the previous inputs, which makes them well-suited for tasks such as speech recognition, natural language processing, and language translation.

In simple terms, deep learning is a name for neural networks with many layers.

To make sense of observational data, such as photos or audio, neural networks pass data through interconnected layers of nodes. When information passes through a layer, each node in that layer performs simple operations on the data and selectively passes the results to other nodes. Each subsequent layer focuses on a higher-level feature than the last, until the network creates the output.

In between the input layer and the output layer are hidden layers. This is where the distinction comes in between neural networks and deep learning: A basic neural network might have one or two hidden layers, while a deep learning network might have dozens—or even hundreds—of layers. Increasing the number of different layers and nodes may increase the accuracy of a network. However, more layers can also mean that a model will require more parameters and computational resources.

Deep learning classifies information through layers of neural networks, which have a set of inputs that receive raw data. For example, if a neural network is trained with images of birds, it can be used to recognize images of birds. More layers enable more precise results, such as distinguishing a crow from a raven as compared to distinguishing a crow from a chicken. Deep neural networks, which are behind deep learning algorithms, have several hidden layers between the input and output nodes—which means that they are able to accomplish more complex data classifications. A deep learning algorithm must be trained with large sets of data, and the more data it receives, the more accurate it will be; it will need to be fed thousands of pictures of birds before it is able to accurately classify new pictures of birds.

When it comes to neural networks, training the deep learning model is very resource intensive. This is when the neural network ingests inputs, which are processed in hidden layers using weights (parameters that represent the strength of the connection between the inputs) that are adjusted during training, and the model then puts out a prediction. Weights are adjusted based on training inputs in order to make better predictions. Deep learning models spend a lot of time in training large amounts of data, which is why high-performance compute is so important.

Objectives

Creating a custom vehicle dataset
Training the model to generate an optimized weights file
Detection and classification of vehicles in the frame
Identifying the vehicle using License plate

II. PROBLEM STATEMENT

Vehicle tracking is the process of locating a moving vehicle using a camera. Capture vehicle in video sequence from surveillance camera is demanding application to improve trackingperformance. This technology is increasing the number of applications such as traffic control,traffic monitoring, traffic flow, security etc. The estimated cost using this

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |

technology will be very less. Video and image processing has been used for traffic surveillance, analysis and monitoring of traffic conditions in many cities and urban areas. Various methods for speed estimation are proposed in recent years. All approaches attempt to increase accuracy and decrease cost of hardware implementation. The aim is to build an automatic system that can accurately localise and track the speed of any vehicles that appear in aerial video frames.

III. METHODOLOGY

This project aims to implement a deep-learning based vehicle detection and tracking system using **yolov5** architecture on **pytorch** framework with python. Here dataset is collected online, this dataset is prepared for training process using **Roboflow** online tool. Using this tool images can be annotated and converted to pytorch format for training, this is a free online tool available to anyone. Here the dataset consists of different classes of vehicles such as car, truck etc., it is possible to add upto three classes for this academic project. Once the dataset is prepared Google-Colab is used for training the dataset using ulanalytics yolov5 repository. Next stage would be testing the created model, once the model is optimized the project can be built on the model. License plate recognition can be done using Platerecognizer API. With the development of the new generation of technology, the information-based and data-based smart expressway has been piloted. The smart express-way can improve traffic safety and efficiency. Additionally, the smart freeway allows vehicle–road collaboration by building an efficient communication system between the cloud platform, roadside infrastructure, road users, and big data centers.

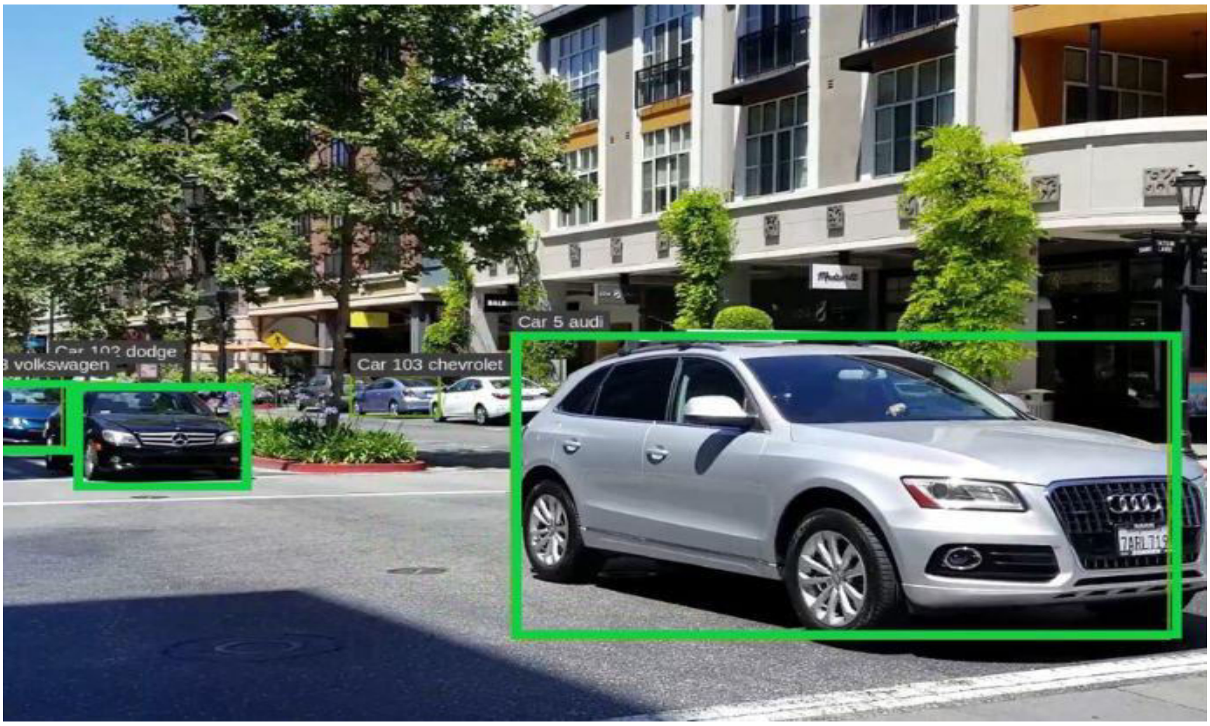
Although the construction of the Chinese expressway network is becoming more intelligent, and comprehensive traffic management technology is improving rapidly, there are still some challenges that need to be solved. The expressway realized the “one network” operation mode of “one pass, one deduction, one notification” [2], and the whole network system adopted segmented billing. The charging mode was changed from weight charging to per-vehicle charging, and the billing mileage was determined by the ETC system and the toll booths according to the driving path. Under the new toll collection system, the expressway toll system faces the problem of evading tolls and difficult recovery. Moreover, compared with urban arterial roads, the expressway has the characteristics of fast speed, large traffic capacity, and a high volume of commercial trucks with dangerous goods. Although the accident rate is relatively low, the harm caused by traffic accidents on the expressway is more serious [3],and the subsequent effects, such as congestion caused by accidents, last longer. Vehicle target detection on the expressway is important for intelligent traffic management and safety monitoring. It is the basis for realizing intelligent and diversified traffic management. Relying on manual supervision is inefficient, and it can only be analyzed after the event has occurred. Some traditional intelligent monitoring systems have a high false-positive rate and a slow speed. The early warning information usually has a high false rate. Since 2006, the rise of deep learning has enabled computer vision technology to develop from manual design features to higher precision and intelligence [4], which also provides technical support for the real-time, full, and efficient use of surveillance videos. Vehicle category detection by monitoring video streams can strengthen the operation and maintenance supervision of expressways and improve vehicle driving efficiency and safety.It is of great significance for ensuring the safety of people’s lives and property and the development of the economy.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |



IV. THE CONCEPT OF DETECTING MOVING OBJECTS IN VIDEOS

Object detection is an enthralling area of computer vision. When we're dealing with video data, it takes on a whole new level. The intricacy increases, but so do the rewards. Using object detection techniques, we can do extremely helpful high-value jobs such as surveillance, traffic control, criminal fighting, etc. Here's an animated GIF to demonstrate the concept: Counting the number of objects, determining the relative size of the items, and determining the relative distance between the objects are all sub-tasks in object detection. These sub-tasks are crucial since they help solve some of the most difficult real-world challenges. Let's have a look at some of the intriguing object detection use cases in real-world applications. Nowadays, video object detection is being used in a variety of sectors. Video surveillance, sports broadcasting, and robot navigation are among the applications. The good news is that the options are limitless regarding future use cases for video object detection and tracking. Here are some of the most fascinating applications :Counting the crowd ,Detection and recognition of car license plates Sports ball tracking ,Robotics. Traffic management

What Exactly Is YOLO?
YOLO is an acronym that stands for You Only Look Once. It is an object recognition algorithm that operates in real time. It is capable of classifying and localizing several objects in a single frame. Because of its smaller network topology, YOLO is an extremely quick and accurate algorithm.

The Architecture Of YOLO

- The YOLO network comprises 24 convolutional layers that are followed by two fully linked layers. The convolutional layers are trained on the ImageNet classification algorithm at half the resolution (224 224 input picture) before being double-trained for detection.
- The several layers minimize the feature set from previous layers, alternate 1 1 reduction layer, and 33 convolutional layers.
- The final four layers are added to train the network to detect objects.
- The last layer forecasts the object class and bounding box probabilities.
- To interact with YOLO directly, we'll use OpenCV's DNN module. DNN is an abbreviation for Deep Neural Network. OpenCV includes a function for running DNN algorithms.
- **Vehicle Detection System And Classification Project Using OpenCV**
- In this project, we will detect and classify cars, HMV (Heavy Motor Vehicle), and LMV (Light Motor Vehicle), on the road, as well as count the number of cars on the road. And the data will be saved in order to examine various automobiles on the road.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

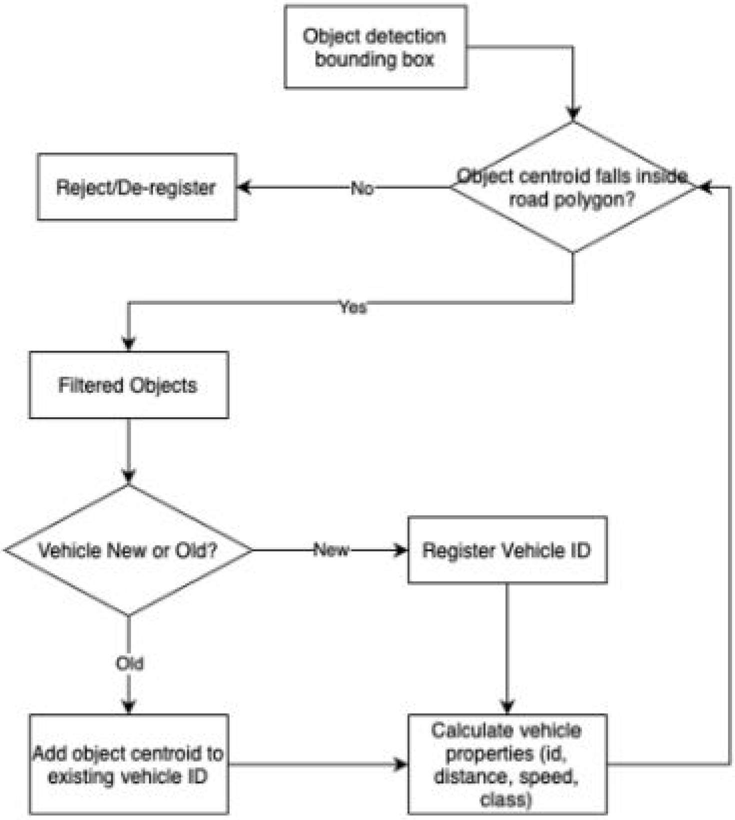
|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |

- To complete this project, we will develop two programs. The first will be a car detection tracker that uses OpenCV to keep track of every identified car on the road, and the second will be the primary detection software.

Tracker

The tracker uses the Euclidean distance to maintain track of an item. It computes the distance between two center points of an object in the current frame and the previous frame, and if the distance is smaller than the threshold distance, it certifies that the object in the previous frame is the same object in the present frame.



- The system begins with login page and then after optioning, taking video or providing recorded video as input to the system of object detection. Then the inputted data is processed into frames for detection. During detection the The Yolo V5 uses the Model or dataset to detect the object in the input data. By using the model after detecting objects the detected objects are classified and represented by using labels and bounding boxes around the detected objects and then it is processed into output

- Input

Using OpenCV we can access camera module and also we can add video files in different formats. Using OpenCV the the real-time video frame is collected from camera lenses. Before collecting the real-time video frame we assigned a Tkinter window prompt which signifies or collects from the user about which dataset or model should be used for detection. Once the input option is received the user option is sent to neural network module and the camera is enabled using OpenCV and starts collecting the video frames from the camera lens.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |

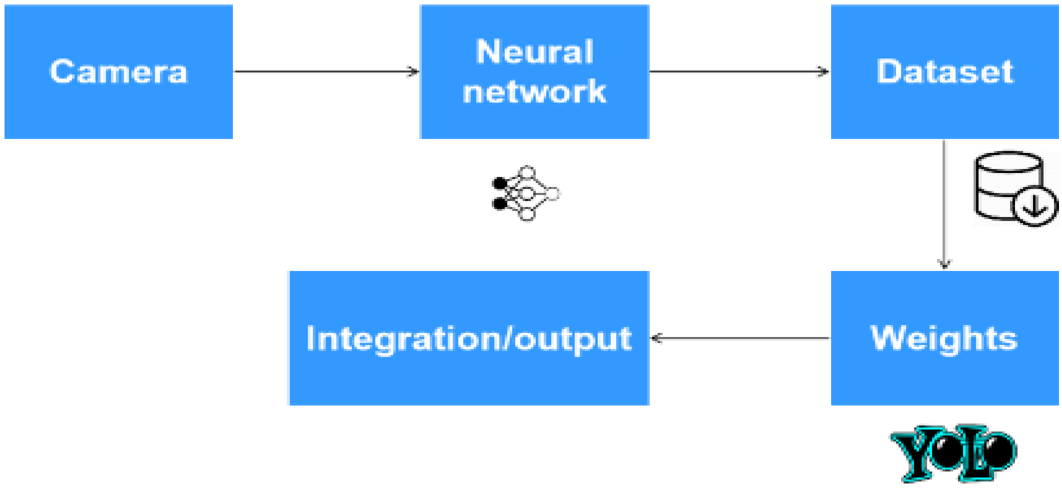


Fig. 2. General working system

B.Neural Network(YOLO V5)
Here using the input received from the camera or video the input data is classified into frames and each frame is sent to yolo detection algorithm with the model which user selected. The model can be a predefined model that is, COCO dataset model or we can create custom models for detection. Once the detection is done, it is bounded with boxes the object is found and is send to output section where the detected frames are collected and are then compressed into output format. Before merging, the detected frames are used for tracking, counting and sorting using OpenCV and also for better results DeepSORT is also

C.Dataset
This field is used for creating custom dataset from raw images in order for creating a custom model which can be used for detection. For this the first thing is used to collect the raw images from various sources and create a dataset. Then from the dataset images the objects must be annotated and labelled from the images. For this Python frameworks like "Labeling" is used for annotating and labelling of the objects. Once the annotating and labeling is done then the dataset is split into train and test images in percentage of 70% for train and 30% for test as it is the general ideal percentage used for training. Once this is done it can send to yolo training algorithm where the dataset can be trained and model can be created using COCO dataset model.

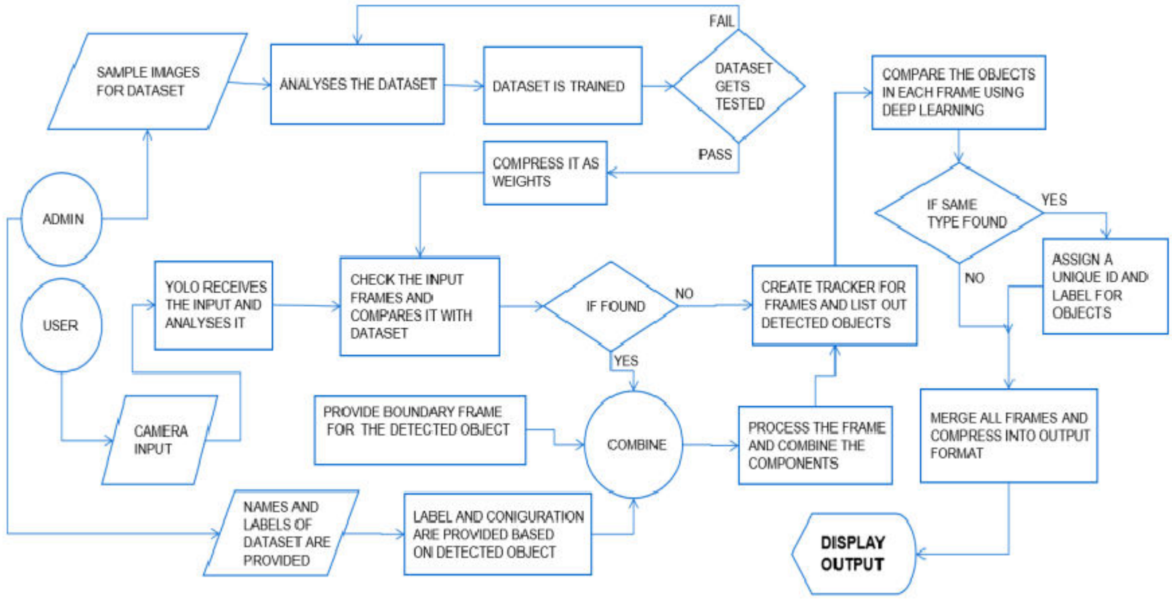
D. Weights / Model
Here the labelled dataset obtained from the framework must be configured with ".yaml" (YAML Ain't Markup Language) extension format file which can be used to append the text label to the algorithm. Once the YAML file is configured it is set up in algorithm and using pytorch the given dataset gets trained using GPU according to the epochs given in algorithm for training and with test dataset the testing of trained model after completion also takes place, predicting the objects in test image. Once the objects are predicted the model is compressed into the yolo model format which is configured using the pre- trained model that is using COCO dataset model. Once this is done the model file with the corresponding test result potted in graphs and texts are written in output folder where the evaluation and testing of the model can be done by using it in a detection algorithm.

E. Integration
Here the above five modules are intersected to make a single system where at front a login page using MySQL can also be set up but it is considered as optional as opencv mostly works under a secure arena . After integrating the modules it is compressed into an executable file which is in ".exe" file format output, which can be used as a MOT application.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |



V. CONCLUSION

This algorithm has revolutionized the field of object detection with its unique approach and impressive speed. Unlike traditional methods that involve separate steps for identifying objects and classifying them, YOLO accomplishes both tasks in a single pass, hence the name 'You Only Look Once'. Our system is successfully recognized about the object's moving and standing activities. This is an essential step for making the legacy data useful for data mining and machine learning. We have used a pre-trained YOLO neural network and used data generated with our own simulator to retrain the network to detect the components we are interested in.

REFERENCES

[1]. Ministry of Transport of the People's Republic of China, Statistical Bulletin of Transport Industry Development 2020. Available
online: https://www.mot.gov.cn/jiaotongyaowen/202105/t20210519_3594381.html (accessed on 9 May 2022).
[2]. Jiangsu Provincial Department of Transport, Framework Agreement on Regional Cooperation of Expressway. Available online:
http://jtyst.jiangsu.gov.cn/art/2020/8/24/art_41904_9471746.html (accessed on 9 May2022).
[3]. Park, S.-H.; Kim, S.-M.; Ha, Y.-G. Highway traffic accident prediction using VDS big data analysis. J. Supercomput. 2016, 72, 2832.[CrossRef]
[4]. Paragios, N.; Chen, Y.; Faugeras, O.D. Handbook of Mathematical Models in Computer Vision; Springer Science & Business Media:Berlin/Heidelberg, Germany, 2006
[5]. Liu, P.; Fu, H.; Ma, H. An end-to-end convolutional network for joint detecting and denoising adversarial perturbations in vehicleclassification. Comput. Vis. Media 2021, 7, 217–227. [CrossRef]
[6]. Lee, D.S. Effective Gaussian mixture learning for video background subtraction. IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27,827–832. [PubMed]
[7]. Deng, G.; Guo, K. Self-Adaptive Background Modeling Research Based on Change Detection and Area Training. In Proceedingsof the IEEE Workshop on Electronics, Computer and Applications (IWECA), Ottawa, ON, Canada, 8–9 May 2014; Volume 2, pp.
59–62.
[8]. Muyun, W.; Guoce, H.; Xinyu, D. A New Interframe Difference Algorithm for Moving Target Detection. In Proceedings of the2010 3rd International Congress on Image and Signal Processing, Yantai, China, 16–18 October 2010; pp. 285–289.
[9]. Zhang, H.; Zhang, H. A Moving Target Detection Algorithm Based on Dynamic Scenes. In Proceedings of the 8th InternationalConference on Computer Science and Education (ICCSE), Colombo, Sri Lanka, 26–28 April 2013; pp. 995–998.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205135 |

[10]. Barnich, O.; Van Droogenbroeck, M. ViBe: A Universal Background SubtractionAlgorithm for Video Sequences. IEEE Trans.Image Process. 2011, 20, 1709–1724. [CrossRef] [PubMed]
[11]. Fang, Y.; Dai, B. An Improved Moving Target Detecting and Tracking Based On Optical Flow Technique and Kalman Filter. In Proceedings of the 4th International Conference on Computer Science and Education, Nanning, China, 25–28 July 2008; pp.
1197–1202.
[12]. Computer Vision-ECCV 2002. In Proceedings of the 7th European Conference on Computer Vision. Proceedings, Part I (Lecture Notes in Computer Science), Copenhagen, Denmark, 28–31 May 2002; Volume 2350, pp. xxviii+817.
[13] Lowlesh Yadav and Asha Ambhaikar, "IOHT based Tele-Healthcare Support System for Feasibility and performance analysis," Journal of Electrical Systems, vol. 20, no. 3s, pp. 844–850, Apr. 2024, doi: 10.52783/jes.1382.
[14] L. Yadav and A. Ambhaikar, "Feasibility and Deployment Challenges of Data Analysis in Tele-Healthcare System," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India, 2023, pp. 1-5, doi: 10.1109/ICAIIHI57871.2023.10489389.
[15] L. Yadav and A. Ambhaikar, "Approach Towards Development of Portable Multi-Model Tele-Healthcare System," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India, 2023, pp. 1-6, doi: 10.1109/ICAIIHI57871.2023.10489468.
[16] Lowlesh Yadav and Asha Ambhaikar, Exploring Portable Multi-Modal Telehealth Solutions: A Development Approach. International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), vol. 11, no. 10, pp. 873–879, Mar. 2024.11(10), 873–879, DOI: 10.13140/RG.2.2.15400.99846.
[17] Lowlesh Yadav, Predictive Acknowledgement using TRE System to reduce cost and Bandwidth, March 2019. International Journal of Research in Electronics and Computer Engineering (IJRECE), VOL. 7 ISSUE 1 (JANUARY-MARCH 2019) ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE).