Evolution in Electrical and Electronic Engineering Vol. 4 No. 2 (2023) 11-18

EEEE

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/eeee
e-ISSN : 2756-8458

Automatic Detection for Moving Car on the Road Using YOLOV5 Algorithm

Fakhrurazzi Rosli[1], Mohd Norzali Hj Mohd [1]*

[1]Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/eeee.2023.04.02.002
Received 15 January 2023; Accepted 05 May 2023; Available online 30 October 2023

Abstract: A camera vision system is a technology that utilizes image processing and video surveillance techniques to gather real-time and video input data. This technology is capable of detecting and classifying objects in real-time, whether through taking pictures or recording videos. The older versions of YOLO, such as YOLOv2 and YOLOv3, may have limitations that can lead to errors in data collection. These errors can cause inconvenience for both data collectors and road users, which can negatively impact the reputation of the system. This work aims to create a camera vision system that incorporates advanced algorithms to improve the accuracy of object detection and tracking. Specifically, the system utilizes the algorithm of YOLOv5 and Deep SORT to detect and track various classes of objects with a high degree of accuracy, thus reducing the risk of errors in data collection. This will enhance the performance of the system, and also make it more reliable for its intended purpose. The primary objective of this work is to develop a robust and efficient camera vision system that can be utilized in a wide range of applications. The system will be designed to improve safety and efficiency and will be suitable for tasks such as traffic monitoring, by providing accurate and real-time information about the traffic situation. For object detection on YOLOv5, a high accuracy of at least 90% is ideal, along with a low false positive rate. The model should be able to detect and classify objects in real-time with minimal latency. It should be able to recognize a diverse range of objects and perform well under various conditions, such as different lighting, camera angles, and object sizes. Additionally, the model should be able to handle occlusions and multiple objects in the same frame with high precision, providing accurate object boundary boxes and confidence scores.

Keywords: Detect, YOLO, Real-Time, Accuracy

1. Introduction

The impact of technology on our daily lives has been significant, with smart systems and artificial intelligence making many tasks easier and more efficient. Among the technologies that have been developed, camera vision systems stand out for their ability to detect and classify objects in real-time,

Corresponding author: norzali@uthm.edu.my

using image processing and video surveillance techniques. These systems are particularly useful for tracking the movement of detected objects, which allows them to monitor the position of these objects throughout the duration of the video stream [1]. By analyzing visual data, camera vision systems can provide valuable insights and information.

This work aims to demonstrate the importance of using a camera vision system for the detection and classification of vehicles passing through a road. The work is commissioned by Sena Traffic Sdn Bhd, which has provided reference materials and video data. The focus of the work is to analyze the types of cars passing through the road and use the data obtained from the camera vision system to improve road conditions and reduce the risk of accidents. However, object detection can be challenging in computer vision due to factors such as lens distortion, lighting variations, changes in scale, sudden movements, partial occlusions, motion blur, object warping, and background noise [2]. These issues can lead to occlusion and loss of information, as well as distorted or blurry images. Nevertheless, object detection is a vital component of the camera vision system, allowing for real-time data analysis, identification of vehicle types, road flow conditions, and vehicle speeds. Visual tracking helps to make the data collected by the camera vision system more accurate and reliable.

Object detection and tracking systems are utilized in various fields, such as military, medical, and security monitoring, to enhance the precision of object identification. However, traditional methods like Content-Based Image Retrieval and filtering techniques may not provide accurate results and can be expensive [3]. Moreover, older object detection models like YOLOv2 and YOLOv3 have limitations in detecting small objects, which can result in errors in data collection, especially in traffic monitoring systems. To overcome these issues, this work aims to develop a camera vision system using advanced algorithms like YOLOv5 and Deep SORT for object detection and tracking. This algorithm will enable the system to detect and track objects accurately, including various classes, minimizing the risk of errors in data collection. The primary objective is to create a reliable and efficient camera vision system that can be used in traffic monitoring systems, improving safety and efficiency while reducing manpower requirements. With the successful implementation of this work, the camera vision system can be utilized in various applications, enhancing the accuracy of object detection and tracking.

2. Materials and Methods

In this work, the primary objective is to implement the YOLOv5 algorithm for object detection and tracking. YOLOv5 is an advanced object detection algorithm that combines speed and accuracy. It outperforms earlier versions like YOLOv3 and YOLOv4, thanks to various optimizations and improvements. While previous studies have mostly used these older models, this work will utilize YOLOv5s - the latest version that offers improved accuracy and performance. YOLOv5m balances accuracy and speed, YOLOv5x is faster and more accurate for high-end GPUs, while YOLOv5s is optimized for resource-constrained devices and real-time applications. The collected dataset will be utilized to train the model by feeding it into the algorithm and adjusting the internal parameters to ensure accurate identification and classification of objects within the data. Once the training is complete, the model will be tested and validated, and ultimately deployed for use in the object detection and tracking system.

Figure 1 shows the process of creating an object detection model involves several steps. The numerous numbers need to be collected as a dataset of images that contain the objects for the model able to make a detection. These images should be divided into three sets: a training set, a label file, and a verification set. The training set will be used to train the model, while the label file will contain the labels for each image in the training set, such as the names of the objects present in the image. The verification set will be used to evaluate the model’s performance. Proper labeling is crucial to the success of the model, as it allows the model to learn what different objects look like and how to classify them accurately. The labeled dataset is then checked for diversity to ensure that it contains examples of all the different object classes the model needs to detect. Once this is complete, the dataset, including the training set, label file, and verification set, is uploaded to Google Drive for easy access from Google Collaboratory, a cloud based Jupyter notebook environment.

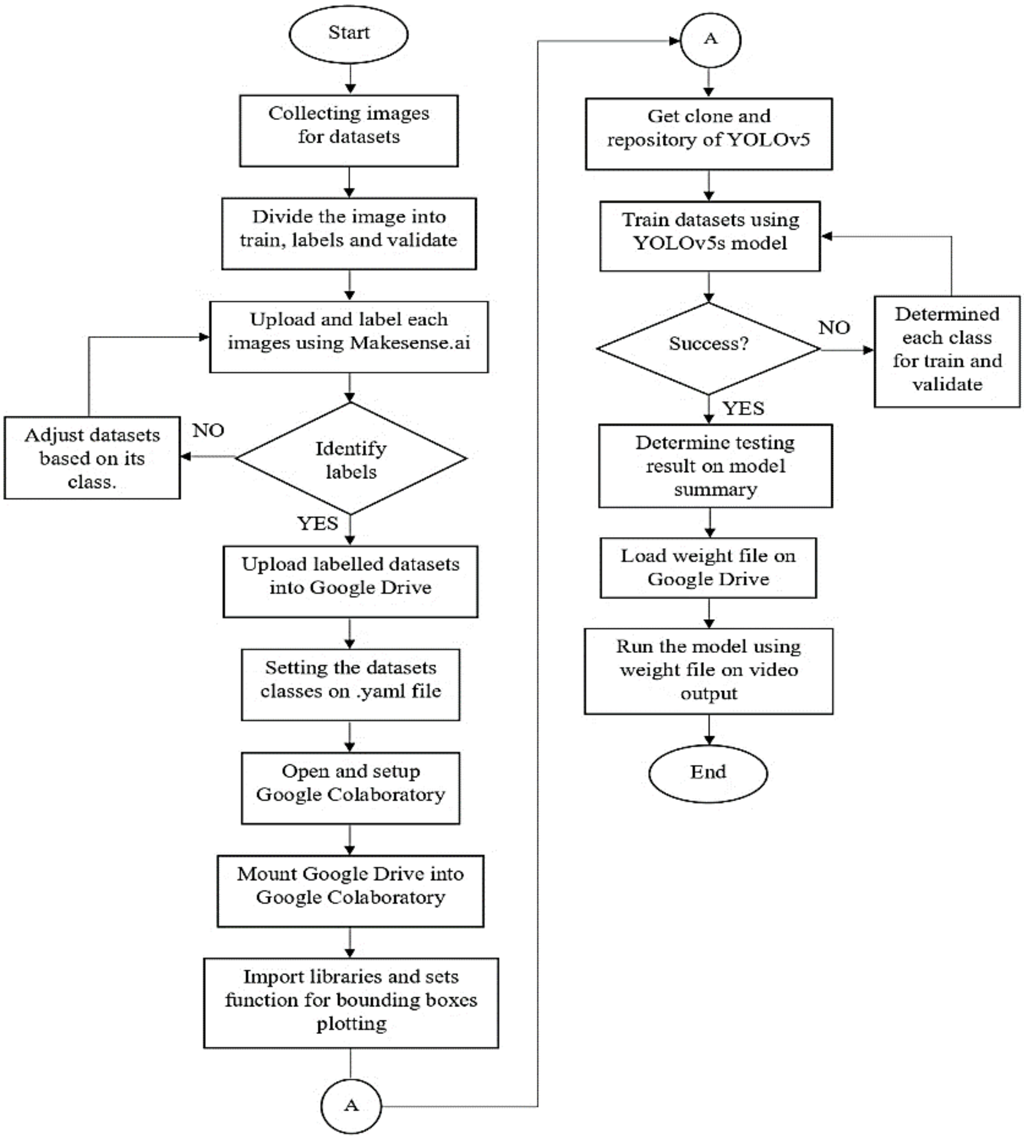

Figure 1: Flowchart of the work

In Google Collaboratory, the data for coco128.yaml is set by specifying the path to the directory containing the dataset on Google Drive using the "data" field in coco128.yaml. It is also essential to import the necessary libraries and functions for plotting bounding boxes around detected objects. A bounding box is a rectangle that encloses an object in an image and helps locate it. To train the model, the YOLOv5 clone and repository model will be placed into Google Collaboratory. The YOLOv5 model is used to train the dataset, which involves using the training set to teach the model to recognize and classify different objects accurately. Once the training is complete, the trained weight file or the trained model is saved to Google Drive.

To use the trained model for object detection on new images or videos, the weight file is loaded, and the model is run on the desired video. The model will process the video and identify any objects present, drawing bounding boxes around them as it goes. This model is also capable of detecting objects in an image and video.

3. Results and Discussion

The graph in Figure 2 shows the train and validation losses, as well as the precision and recall metrics and the mAP metric for a machine learning model. The train loss measures how well the model is able to make predictions on the training dataset, while the validation loss measures how well the model generalizes to new data. The graph shows that the validation loss remains stable or decreases over time, indicating that the model is not overfitting to the training dataset. The precision and recall metrics show the model's ability to correctly identify objects of interest, with high values indicating better performance. The mAP metric is a measure of the overall performance of the model on the dataset, with higher values indicating better performance.

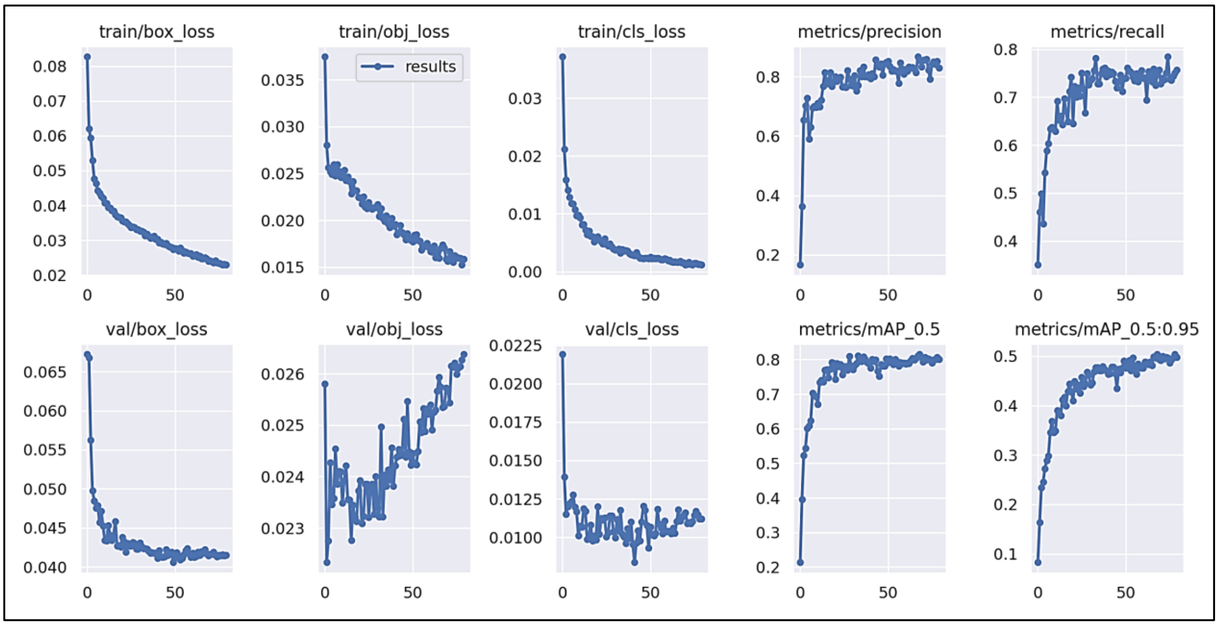

Figure 2: Result for datasets training

In the field of machine learning, people often use precision, recall, precision-recall, and the F1 curve as metrics to measure performance. Precision is a measure of how well a model can predict what will happen. It is found by dividing the number of true positives by the number of positives that were predicted. Recall is a way to measure how well a model can find all the instances of a certain class. Precision-recall is a mix of precision and recall. It is used to figure out how well precision and recall compare. The F1 score is a metric that considers both precision and recall. It is found by taking the harmonic mean of the precision and recall scores. In the case of YOLO, the F1 curve can be used to judge how well the model works on object oriented. Figure 3 shows the data for each curve.

A confusion matrix is a tool for assessing the effectiveness of a classification algorithm, such as an object detection model. It comprises a table that displays the number of accurate and inaccurate predictions made by the model for each category. The confusion matrix offers insight into the model's performance in identifying different types of objects. It provides information about the number of times the model correctly and incorrectly identified each class, with the rows indicating predicted classes and columns indicating actual classes. Through matrix analysis, the strengths and weaknesses of the algorithm can be identified, including the categories in which the model excels at detecting and the areas where it faces challenges. In the context of YOLO object detection, the confusion matrix is useful in evaluating the model's accuracy in recognizing various types of objects.

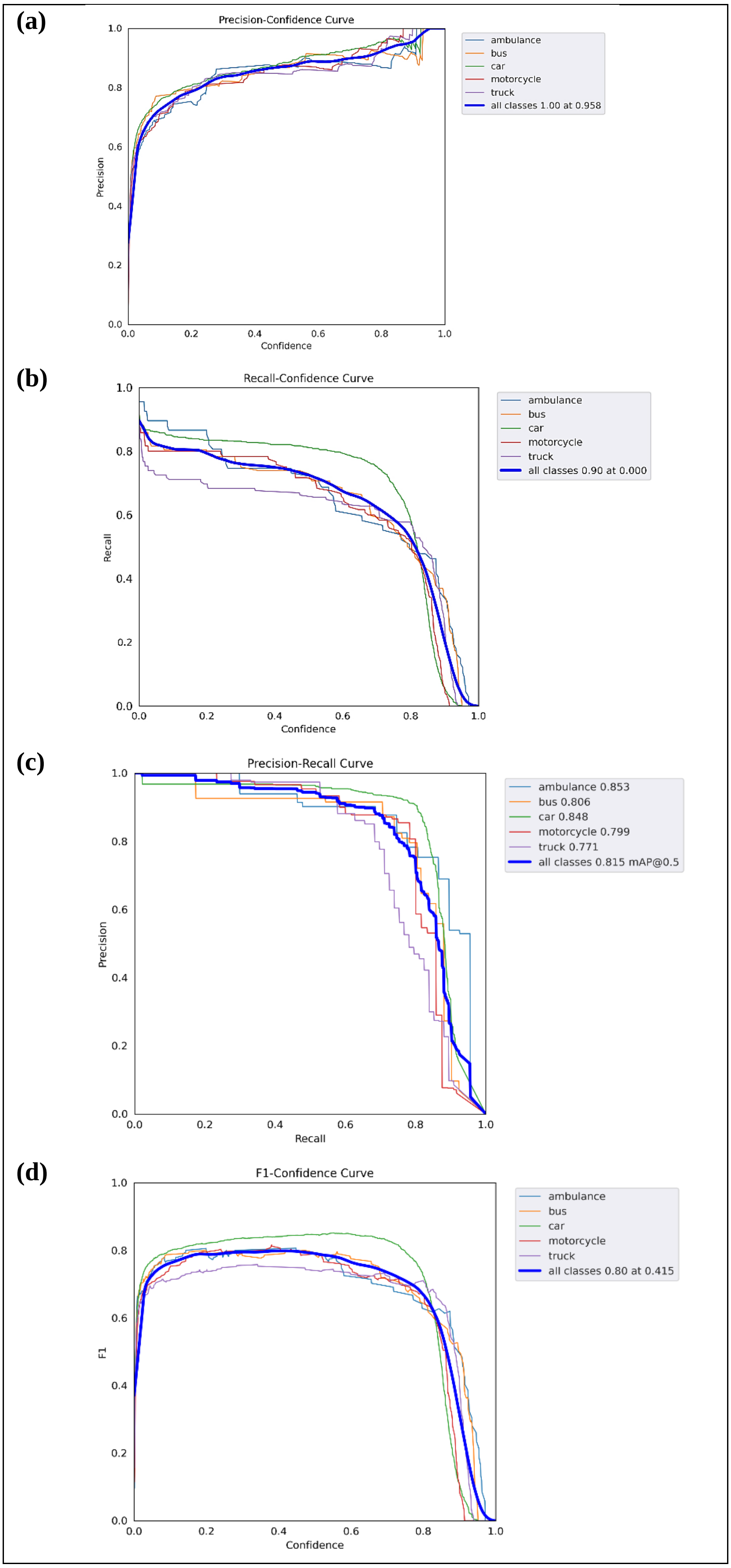

Figure 3: Result for (a) precision, (b) recall, (c) precision-recall, and (d) the F1 curve

Table 1 is a representation of how accurately a classification model is performing. Each row in the table represents a class, and each column represents the model's prediction. The values in each cell represent the accuracy of the model's predictions. This information can be used to evaluate the model's performance and make improvements where necessary.

Table 1: Confusion matrix for TP, FP, TN, and FN

Class	Prediction	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
Ambulance	Ambulance	0.75	0.25	0.00	0.00
Bus	Bus	0.80	0.20	0.00	0.00
Car	Car	0.84	0.16	0.00	0.00
Motorcycle	Motorcycle	0.78	0.22	0.00	0.00
Truck	Truck	0.68	0.32	0.00	0.00

In order to interpret a confusion matrix, it's important to understand its layout. The matrix consists of rows and columns, with each row representing a predicted class and each column representing an actual class. The cells within the matrix display the count of predictions made by the model for each combination of predicted and actual class. The true positives (TP) indicate the number of times the model correctly predicted the actual class, while true negatives (TN) represent the number of times the model correctly predicted a class that is not the actual class. False positives (FP) indicate the number of times the model incorrectly predicted the actual class, while false negatives (FN) indicate the number of times the model incorrectly predicted a class that is not the actual class. From this confusion matrix in Figure 4, the average accuracy of the system for each class is around 70%.

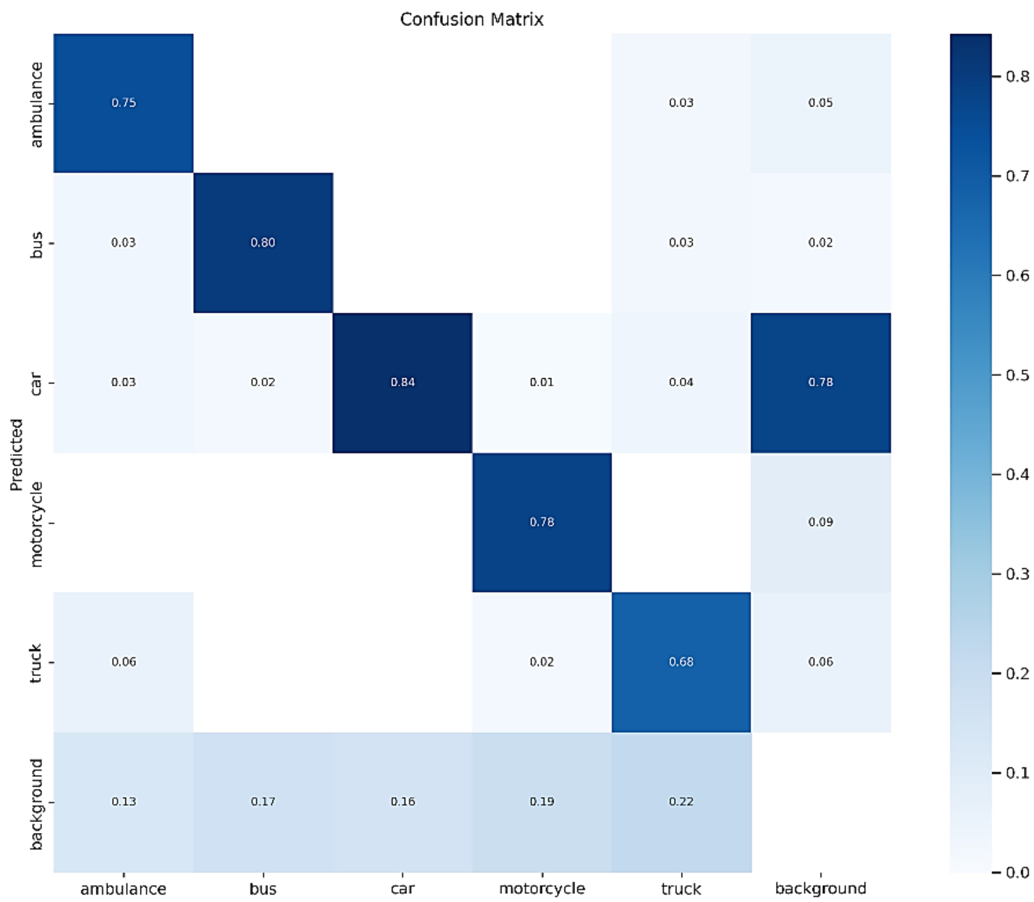

Figure 4: Result of confusion matrix

This work utilizes two distinct video samples, which have been recorded and retrieved from the internet. These samples are utilized to evaluate the effectiveness of the object detection system, with the aim of ascertaining if the detection accuracy is affected by various scenarios. The analysis will present the detection results in the form of images or visualizations derived from the video samples. This will enable an objective evaluation of the system's performance, gauging its ability to detect objects within the videos. The testing of the model system using various video samples is imperative to ensure its robustness and its capability to handle different input types. The utilization of video samples obtained from the internet is crucial as it enables the simulation of real-life conditions, allowing for an accurate evaluation of the system's performance in diverse situations.

The goal is to determine the ability of the object detection system to accurately identify objects in the selected video samples. The assessment will consider the system's ability to identify objects under varying environmental conditions, such as changes in lighting, object size, and distance. The analysis will also evaluate the system's efficiency in detecting multiple objects simultaneously, and its ability to accurately track the movement of the objects within the video. The usage of video samples in evaluating the object detection system is critical to determine the system's effectiveness and robustness.

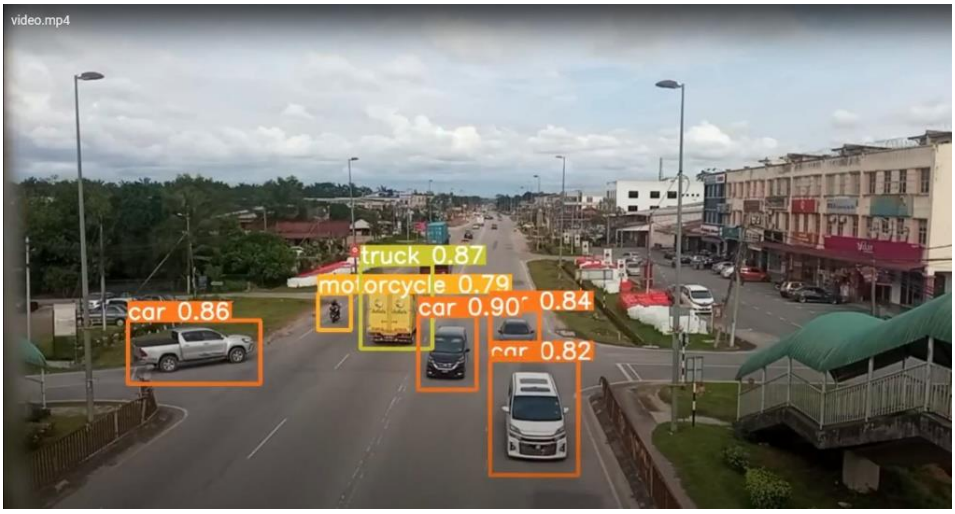

Figure 5: Snapshot from the recorded video

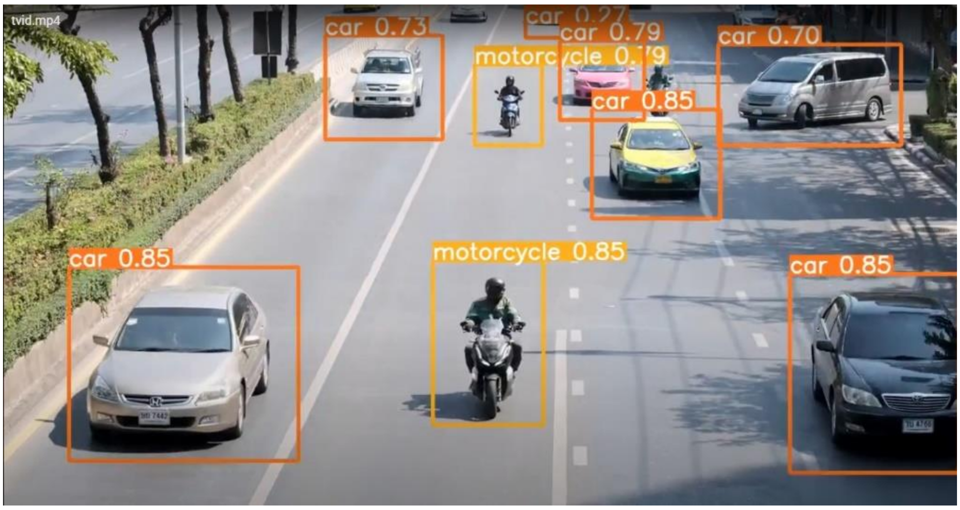

Figure 6: Snapshot from the downloaded video

Based on the evaluation of the recorded and downloaded videos, the object detection system appears to be performing exceptionally well, with a high number of detections for each class. As shown in Figure 5 and 6, the system has shown an impressive level of accuracy in detecting and tracking objects, even when they are in motion or distant. Furthermore, the model can differentiate between different types of objects effectively, as demonstrated by its ability to accurately categorize the detected cars into their respective classes. These findings highlight the robustness and adaptability of the YOLOv5 model, as it can accurately detect objects in diverse scenarios and contexts.

4. Conclusion

In this work, the YOLOv5 algorithm was selected for object detection due to its high accuracy in identifying cars on the road. However, it is worth noting that the accuracy of the detection system is also influenced by the quantity and quality of the datasets used, as well as the number of training sessions or epochs during the training process. To train the custom dataset on the cloud, we utilized Google Colaboratory's GPU, which offers excellent performance. The dataset was first uploaded to Google Drive and then mounted to Google Colaboratory for use in training. After installing the required packages and libraries, the custom model was trained on the dataset, with the architecture defined, compiled, and evaluated using a validation set. The results were then analyzed using a PR curve, mAP metric, and confusion matrix to assess the performance of the YOLOv5 algorithm for object detection.

The results showed that the YOLOv5 algorithm was able to achieve an average accuracy of 80% in identifying cars on the road. This level of accuracy is impressive for a detection object model and suggests that the YOLOv5 algorithm is an effective choice for car detection and tracking. Overall, the work demonstrates the importance of using appropriate algorithms and datasets for accurate detection and tracking of cars on the road, which can ultimately contribute to improving road safety.

Acknowledgement

This research was made possible by the Faculty of Electrical and Electronics Engineering, University Tun Hussein Onn Malaysia for its support in terms of the acquisition of additional materials for the development of this works.

References

[1] T.L. Dang, G. T. Nyuyen and T. Coa. “Object Tracking Using Improved Deep Sort YOLO V3 Architecture,” ICIC Express Letter, vol. 14, no. 10, pp. 961–969, 2020, doi: 10.24507/icicel.14.10.961.

[2] A. Bathija and G. Sharma, G. “Visual Object Detection and Tracking Using YOLO and SORT Visual Object Detection and Tracking Using YOLO and SORT,” Intern. J. of Eng. Research & Technology (IJERT), vol. 8, no. 11, pp. 705-708, Nov. 2019.

[3] K. Madasamy et al. “OSDDY: Embedded System-Based Object Surveillance Detection System with Small Drone Using Deep YOLO,” EURASIP Journal on Image and Video Processing, vol. 2021, no. 1, 2021, doi: 10.1186/s13640-021-00559-1.