Deep Learning Techniques in Autonomous Vehicle Environment: A Comparative Study

[1]Ann Mary Martin, [2]Pooja Das, [3]Anoop Thomas, [4]Ajai V Babu

[1,2]M.Tech Scholar, [3,4]Asst. Professor, Rajagiri School of Engineering and Technology, Kakkanad, India. [1]annmarymartin96@gmail.com, [2]poojaanildas@gmail.com, [3]anoopt@rajagiritech.edu.in, [4]ajaiv@rajagiritech.edu.in

Abstract: Road transportation has brought enormous benefits both to society and to individuals by facilitating easy access to a wide economic service. The increase in motorization coupled with expansion of road networks has brought adverse effects. The important reasons for these traffic problems are mainly because of the environment, over-speed, violation of traffic rules, lack of visibility, fatigue of drivers and so on. Vehicles are more prone to accidents at bad environmental conditions and obstacles in the roads . To reduce these accidents detection of Obstacles in the road environment is mandatory. For that the latest technologies can be incorporated with the vehicle for making it smarter. The detection of obstacles, preceding vehicles as well as oncoming vehicles can be done using sensors like camera and ultrasonic sensors. Using the collected data from the sensors, proper object detection can be done. There are several object detection frameworks existing today like R-CNN, SSD, SVM, YOLO etc. These are the image processing tools developed based on the theory of ML, DL and AI. To detect the high mobility vehicles and obstacles, a proper object detection framework should be selected by comparing the characteristics of each machine learning technique. Different types of deep learning techniques for object detection are compared. Since the YOLO and SSD showing more improved characteristics a detailed comparison is required to choose a better one. So, by comparing the pros and cons of each version of YOLO and SSD, the best among those could be selected.

Keywords — YOLO, SSD, Object Detection, Deep Learning (DL), Artificial Intelligence (AI), Machine Learning (ML), Sensors.

I. INTRODUCTION

We couldn't imagine a world without vehicles and transportation. That much vehicles have a significant role in human life. It makes several benefits to each individual as well as society by making the transportation easier and time efficient. Roadways are a crucial means of transportation. They provide people and vehicles to commute a wide range of locations and help them to reach their destination without any hiccups. Road transportation is the process of transporting goods or people from one destination to the other via roads. Several evolutions have taken place in the field of automobile industry. The invention of steam powered vehicles was one of the most significant inventions in this automobile field. After that there were several modifications given to this field and presently the field is at a scenario where a vehicle is capable of driving on its own. The automobile field is advancing day by day.

Even though these advancements came into picture, normal people will definitely go for an IC engine vehicle because of its easy availability and low cost. But in the case of an IC engine vehicle, it is having lots of environmental problems and fuel consumption is also high in IC engine vehicles. There comes the relevance of the electric vehicle which are also having the same problems with IC engine, but comparatively low and engine could be replaced with the motor. The increase in motorization coupled with expansion of the road network has brought several traffic problems, such as road accidents. In order to properly solve urban traffic problems and overcome the existing disadvantages, such as the lack of enough vehicle information and the low accuracy of vehicle information retrieval, intelligent transportation was strongly developed. The emergence of Electric Vehicles and Autonomous Vehicles could reduce these adverse effects to a certain extent, since it is more concerned about safety. Tesla cars have revolutionized the electric vehicle industry. In fact, these electric vehicles have received much attention in every corner of the world. These cars are environmentally friendly and have the potential to deliver a superior performance to the users. Along with Electric vehicles, Autonomous vehicles also have a great relevance in the present scenario.

An autonomous vehicle is a system that is capable of sensing its environment and moving without human input. Over the

 DOI : 10.35291/2454-9150.2021.0227

last few years, autonomous vehicles have attracted a lot of attention in both research and industrial domains and have witnessed a perpetual evolution since then. Autonomous vehicles have received increasingly significant attention due to vehicle safety, performance, traffic efficiency and energy saving. Nowadays autonomous vehicles have become a concrete reality that pave the way for a future system where computers take over the art of driving.

Autonomous vehicles should be instantaneous, accurate, stable, and efficient in computations to produce safe and acceptable traveling trajectories in numerous urban to suburb scenarios and from high-density traffic flow to high-speed highways. In real-world traffic, various uncertainties and complexities surround road and weather conditions, whereas a dynamic interaction exists between objects and obstacles and tires and driving terrains. If there is a large number of objects in the camera data then it will be difficult for an autonomous vehicle to recognize the objects and it will create real-time issues and the process will be time consuming. In order to overcome such issues, object detection learning techniques are adopted in autonomous vehicles. At present, Autonomous vehicles have the capability to drive, slow down, stop, and even park by themselves. Google is one of the companies to start the prototyping of fully self-driving vehicles. These vehicles are equipped with advanced sensors such as ultrasonic, RADAR, LIDAR and cameras with 360 degrees of view. The purpose is to overcome obstacles, collision avoidance, and ensure a safe driving of the vehicle especially in congested areas. For instance, in 2018, a crash caused the first death of a pedestrian caused by a fully autonomous car. Therefore, it is essential to devise more advanced and complementary solutions to further improve the capabilities of these vehicles. Therefore, in order to overcome such failures different object detection models are adopted. Using the latest technologies like Machine Learning, Deep Learning and Artificial Intelligence this image processing can be done easily. Deep learning, also known as a deep neural network, is a method used to process raw data and automatically find the representation of features needed to do classification or detection. It is cited in that YOLO uses a single step to predict the whole image, classification, and localization at once. The whole image features are used to predict classes and bounding boxes simultaneously.

This work will overview the main object detection learning techniques using image processing for autonomous vehicle applications. The purpose is to investigate the performance of machine learning solutions as well as deep learning algorithms based on the accuracy and the timing-process learning algorithms based on the accuracy and the timing-process. We start by presenting the architecture of the three approaches: namely the machine learning SVM model, the deep learning You Only Look Once (YOLO) approach, and the Single Shot Multibox Detector (SSD) deep learning model.

II. LITERATURE SURVEY

Autonomous vehicle technology is the biggest innovation in the automotive industry that has been seen in the past century. In recent times, the race towards making Autonomous vehicles or self-driving cars has taken a sudden growth as several automotive companies are investing large amounts of capital in it. Autonomous cars are equipped with an autonomy system where the architecture of this autonomy system is divided into the perception system and the decision-making system. This section is an overview of a literature survey that has been made for the implementation of this paper.

Object detection is one of the important software components in the next generation of autonomous vehicles. Convolutional machine learning approaches for object detection mainly suffer from low detection rate. Modern algorithms are mostly based on artificial neural networks, such as the YOLO algorithm, which solves the problem without accuracy and precision losses. [4] The third version of the YOLO algorithm is used for the detection of different types of traffic participants. The network is trained for object classes like car, truck, pedestrian and so on. The effectiveness of the YOLOv3 in the variety of driving conditions like bright and overcast sky, snow, fog, and night are observed and evaluated.

Autonomous technology must provide a high safety level on road networks to decrease the number of accidents caused by human errors. Automated object and obstacle detection is one of the main research tasks that must be undertaken and a number of vision-based learning techniques have been designed to improve the vehicle detection capabilities and reduce the shortcomings of other sensors such as the LIDAR system that have shown poor results during severe weather conditions. [6] Here is a study of main learning models for video-based object detection specifically that can be applied with autonomous vehicles. The focus is on machine learning solutions, namely, the Support Vector Machine (SVM) algorithm and two deep learning solutions like the You Only Look Once (YOLO) and the Single Shot Multibox Detector (SSD) methods.

The earlier Machine Learning algorithms which break down each problem into separate modules and solve them individually. Nowadays requirement of detection algorithm is to work end to end and take less time to compute. [5] Real Time detection and classification of objects from video records provide the foundation for generating different kinds of analytical aspects such as the amount of traffic in a particular area over the years or the total population in an area. To overcome these issues, YOLO based detection and classification approach for improving the computation and

 DOI : 10.35291/2454-9150.2021.0227

processing speed and at the same time efficiently identify the objects in the video records.

For the detection of high mobility vehicles at night there are many robust image processing tools which are developed based on the theory of Machine Learning, Deep Learning and Artificial Intelligence. [1] The object detection framework YOLO is used to detect and track a large number of high mobility vehicles. Different versions of YOLO have been developed and compare the detection speed and detection accuracy.

Real-time detection and classification of on-road objects is necessary to analyze the vehicle environment in order to provide to the driver valuable information so that he can be notified of potential hazards and for better assistance. [7] A single shot detector which uses VGG-16 as its base network and 6 feature layers afterwards. The grid cell of each feature layer is responsible for detecting objects in a specific size using default bounding boxes at specific aspect ratios. The first three feature maps use six default bounding boxes at each grid cell, while the next three ones use only four. Also, YOLO performs its detections using two intermediate fully connected layers, and performs detections on a single scale feature layer.

[8] There is an end-to-end deep learning-based system for multi-object detection, depth estimation, localization, and tracking for realistic road environments. For the object detection module, an effective detector based on YOLOv3 was proposed. The feature extraction is performed by DarkNet with 53 layers. It has several implementations on different Deep Learning Python libraries. They have all been tested in order to compare their computation times. [10] Deep learning techniques have several applications in fields like Surveillance, Military, Transportation, Medical, and Daily Life. Number of factors are affecting the detection performance of each deep learning technique. Wide range of object categories, limited storage capacity, computational power etc. are few of them.

The environmental condition is also important in driving an autonomous vehicle. [9] Poor road conditions like cracks and potholes can cause inconvenience to passengers, damage to vehicles, and accidents. Detecting those obstacles has become relevant due to the rise of the autonomous vehicle. This issue can be solved with the help of deep learning–based technique You Only Look Once version 2 (YOLOv2) detector and propose a deep convolutional neural network (CNN) based on YOLOv2 with a different architecture and two models. Proposed architecture is able to obtain a significant increase in performance. [2] Based on the background of the convolutional neural network (CNN), one of the best CNN representatives You Only Look Once (YOLO), which breaks through the conventional CNN family's tradition and develops a fully new way of solving object detection with simple and high performance. Its fastest speed has achieved the exciting unparalleled result and achieved more performance when compared with Faster R-CNN. Additionally, compared with the latest most advanced solution, YOLOv2 achieves an excellent efficiency in speed and accuracy as well as an object detector with strong generalization ability which represents the full image. [3] There is a system for vehicle detection and tracking from the high-resolution video input. It detects the object and identifies the object by comparing its features with the features of objects stored in the database. If the features match, then the object is tracked using the YOLOv3 framework. The object detection from a video input, which further helps to evaluate detection rate on real time systems also.

III. OBJECT DETECTION

There are different methods for detecting, recognizing and classifying objects from different backgrounds. There are different types of object detection learning techniques using image processing for autonomous vehicle applications. The purpose is to investigate the performance of machine learning solutions as well as deep learning algorithms based on the accuracy and the timing-process. Fast R-CNN, Faster R-CNN, Histogram of Oriented Gradients (HOG), Region-based Convolutional Neural Networks (R-CNN), Region-based Fully Convolutional Network (R-FCN), Single Shot Detector (SSD), Spatial Pyramid Pooling (SPP-net), YOLO (You Only Look Once) are some of the commonly used Deep Learning object detection frameworks. Among these most recently developed and commonly used for real-time applications are SSD and YOLO frameworks. YOLO is again subdivided into different versions. These approaches are adopted to guarantee the performance in terms of accuracy, speed of frame process, real time processing, loss at minimal rate.

1.1 SINGLE SHOT MULTIBOX DETECTOR (SSD)

Single shot multibox detector (SSD) is designed for object detection in real time. The approach of this algorithm is according to the feed-forward convolutional neural network (CNN) followed by a non-maximum suppression filter. In the case of faster R-CNN, it uses a region proposal network to make the boundary boxes and using those boxes, it classifies objects. In SSD the whole process runs at 7 frames per second. The main disadvantage with SSD is the reduced accuracy. To cover up this drop in accuracy, few improvements like multi-scale features and default boxes are applied by SSD. These improvements make the accuracy and speed of SSD to increase by using lower resolution images. So thus, SSD achieves the real-time processing speed and accuracy.

1.1.1 ARCHITECTURE-SSD MOBILENET

The SSD architecture is a single convolution network that tries to predict bounding box locations and classify these locations in that particular image. Thus, SSD can be trained

 DOI : 10.35291/2454-9150.2021.0227

end-to-end. The SSD network consists of a base architecture (Mobile-Net) followed by different convolution layers. Using SSD, we need to take only one single shot to detect more than one objects within that particular image. Hence, SSD is far better in speed when compared with other CNN techniques.

1.2 YOU ONLY LOOK ONCE(YOLO)

YOLO is a fully Convolutional Neural Network (FCN). It is also known as YOLO VERSION 1(YOLOv1). It contains only convolutional layers and is invariant to the size of the input image given. YOLO means that an image can predict what the objects are and localization of those objects at one glance. It is very simple to develop and can be trained directly on full images unlike other detections tools. In YOLO the input image is converted into a grid containing SXS cells and each of these cells are used for the prediction of N bounding boxes enclosing each object in the image. For every bounding box, the object belonging to a particular class is detected with a score showing the confidence level, that is the accuracy of detection.

1.2.1 ARCHITECTURE OF YOLO V1

YOLO is a simple detection framework. YOLOv1 is an extremely fast object detection technique that processes images in real-time as indicated by several studies. A single convolutional neural network which predicts more than one bounding box at the same time and class probabilities for all those bounding boxes. Unlike other detection techniques, YOLO sees the entire image during training and test time as a result of which it can encode information about a particular class as well as their looks.

YOLOv1 architecture has 24 convolutional layers followed by 2 fully connected layers. YOLOv1 uses features from the entire image and predicts bounding boxes simultaneously. The whole image is mainly divided into SXS grids and each grid produces B bounding boxes and their confidence scores. These confidence scores reflect how accurate the model is that the box contains an object and also how confidently it thinks the box is that it predicts. Each bounding box consists of 5 prediction parameters such as x, y, w, h, and confidence. The (x, y) coordinates indicate the center of the box relative to the bounds of the grid cell. The width and height are predicted relative to the whole image. As per the architecture, the initial convolutional layers of the network extract features from the image. Then fully connected layers predict the output probabilities and coordinates.

ADVANTAGES OF YOLO V1

It is very much fast and used for real time applications. Its global trainable system eases the optimization, since it is more generalized.

LIMITATIONS OF YOLO V1

YOLO imposes strong spatial constraints on bounding box predictions. This spatial constraint limits the number of nearby objects that the model can predict. The model struggles with small objects that appear in groups. Since YOLOv1 is having this many disadvantages, an alternate 2nd version of YOLO (YOLOv2) is developed with better performance.

1.2.2 ARCHITECTURE OF YOLO V2

The use of YOLOv2 in the proposed algorithm has the advantages of decreased computational costs, increased speed, and increased mean average precision compared to YOLOv1. Batch normalization and Anchor-Boxes are the techniques used to preprocess the input data, which greatly improve the performance.

In object detection, we have to predict the location and the shape of an object, not only the classification of the image. In the case of YOLOv2, the output is a 3-dimensional array. Particularly, the shape of output of YOLOv2 is *13x13xD*, where D changes depending on the number of classes of object to detect (D=5 for single class). The first two-dimensional array (*13x13*) is called grid cells and in total 169 grid cells will be there. YOLOv2 is trained on different architectures such as VGG-16, Google Net, and Darknet-19. The reason for choosing the Darknet architecture here is its lower processing requirement than other architectures. For detection purposes, we replace the last convolution layer of this architecture and instead add three *3 x 3* convolution layers every 1024 filters followed by *1x1* convolution with the outputs we need for detection.

ADVANTAGES OF YOLO V2

It is faster than all other detection systems. It can run at a variety of image sizes to provide a smooth tradeoff between speed and accuracy.

LIMITATIONS OF YOLO V2

The precision for small objects is less for YOLOv2 and the localization errors are present in YOLOv2. So, a better version of YOLO is needed to meet the requirements.

1.2.3 ARCHITECTURE OF YOLO V3

YOLOv3 uses a variant of Darknet, which is having a 53-layer network trained on Image Net. For the task of detection, 53 more layers are stacked onto it, which gives 106 layer fully convolutional architecture for YOLOv3. This is the reason behind the reduced speed of detection of YOLOv3. The most important feature of YOLOv3 is that it makes detections at three different scales, that is the output is precisely given by downsampling the dimensions of the input image by 32, 16 and 8 respectively. In YOLOv3 the layer is defined as the ratio by which it down samples the input given. The feature map developed by this kernel has the same

 DOI : 10.35291/2454-9150.2021.0227

width and height of the earlier feature maps and has detection attributes along the depth.

ADVANTAGES OF YOLO V3

Detections at different layers help address the issue of detecting small objects, a frequent complaint with YOLOv2. The up sampled layers concatenated with the previous layers help preserve the fine-grained features which help in detecting small objects. The *13 x 13* layer is responsible for detecting large objects, whereas the *52 x 52* layer detects the smaller objects, with the *26 x 26* layer detecting medium objects. YOLOv3 with the darknet-53 network, which increases the accuracy of object detection by expanding networks. YOLOv3 is better in real time object detection when compared to YOLOv2 and it is providing faster and accurate output. The average precision for small objects improved, it is now better than other versions.

1.3 COMPARISON OF YOLO TO OTHER DETECTION SYSTEMS

Object detection is a difficult task in computer vision. Detection mostly initiated by extracting a set of properties from input images given. Then, classifiers are used to identify objects in the feature space. These classifiers are run either in sliding window fashion over the whole image or on some subset of regions in the image. We can compare the YOLO object detection system to other object detection systems, projecting the similarities and dissimilarities. Static features, classify regions, predict bounding boxes for high scoring regions, etc. have to be extracted to get an overview about the detection frameworks. The system replaces all of these complicated parts with a single CNN. The network performs feature extraction, bounding box prediction, non-maximal suppression, and contextual reasoning all simultaneously. In the place of static features, the network trains the features inline and optimizes them for the detection task. The unified architecture leads to a faster, more accurate model than other detection systems. R-CNN and its variants use region proposals to find objects in images while YOLO uses the sliding windows. Each stage of this complex pipeline must be precisely tuned independently and the resulting system is very slow. Every grid cell proposes the bounding boxes and scores. Anyway, the system gives the spatial constraints on the grid cell proposals that helps to increase more than one detection of the same object. Other Detectors like Fast and Faster R-CNN are actually speed up the R-CNN framework by sharing computation and using neural networks to propose regions instead of Selective Search. While they offer speed and accuracy improvements over R-CNN, both still fall short of real-time performance. YOLO is a general-purpose detector that learns to detect a variety of objects simultaneously. Unlike R-CNN, YOLO trains CNN to predict regions of interest instead of using Selective Search. YOLO predicts both bounding boxes and class probabilities for different objects of more than one class in a particular image. When comparing with other real time systems also, YOLO is a better option in several aspects. YOLO is trained using VGG-16. This model is more accurate but also significantly slower. It is useful for comparison with other systems that use VGG-16 but since it is slower than real-time, other faster models have to find out. Different versions of YOLO are developed by the researchers and they are still focusing on the updated versions which makes the detection more accurate and precise to cope up with the real time situations.

1.4 POTHOLE DETECTION

Pothole detection using YOLO Deep learning is a method used to process raw data and automatically find the representation of features needed to do classification or detection. It is cited in that YOLO uses a single step to predict the whole image, classification, and localization at once. The whole image features are used to predict classes and bounding boxes simultaneously.

Pothole detection in autonomous vehicles is used to detect the irregularities in the road surface such as depressions and hollow spaces. Pothole detection is about sensing the road surface ahead of an autonomous vehicle, warning and navigating the vehicle in the best possible method. For the detection we use YOLOv2, enabling it to create a more accurate assessment on detecting potholes on the road surface, since it shows a satisfying result in pothole detection. Thus, this technique has a high opportunity to be developed and implemented as a tool for effective negative obstacle detection. In the pothole detection method, various studies have been carried out with different approaches, using a laser, vibration sensor and image-based detection, both two dimensional and three-dimensional using a stereo camera. Recently, object detection using end-to-end deep learning has been reported to outperform traditional methods. The development of advanced image processing techniques, easy image interpretation, and the availability of low-cost camera acquisition devices helped in the development of image-based pothole detection. Here, YOLOv2, being one of the new methods in the field of object detection, is chosen as a pothole detection method.

IV. SIMULATION

The simulation for object detection is done using different object detection frameworks. The platform used is Python (Google Collaborations). A comparison is required to find which framework is better for the real-time application. So that object detection is performed using 3 commonly used frameworks such as SVM, SSD and YOLO.

3.1 Object Detection

The Images used for the detection are mainly of Indian roads and Indian vehicles. Thus, the performance of different object detection frameworks can be compared using the test images taken from different Indian Environments. The main criteria for comparison are to find the better object detection

 DOI : 10.35291/2454-9150.2021.0227

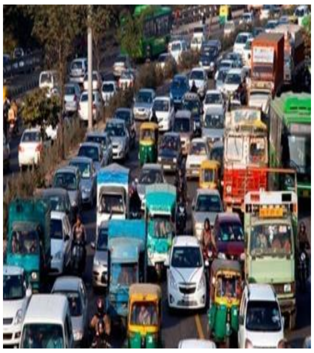
International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-02, MAY 2021

framework with high performance and accuracy rate at real time environmental conditions.

3.2 *Pothole Detection*

Deep learning model YOLOv2 detects potholes from the test images that are given as inputs. The model has been trained with 150 images having 568 potholes labels. The accuracy of the model will increase by training the model with a larger size dataset.

Detection can be done using different test images shown below [Refer Fig.1].



(i)

(ii)

(iii)

(iv)

(v)

(vi)

Fig.1. Sample images for testing

V. RESULTS

From the simulation results it is visible that YOLOv3 is more performed than YOLOv2 and SSD, even though SSD and YOLOv3 are showing almost similar characteristics. In the case of SSD MobileNet v2 the detection is showing a way of identifying only nearby objects with high accuracy. That is, SSD is useful in a condition where nearby objects are crucial when compared to far objects. Even though SSD is detecting nearby objects or vehicles, the detection accuracy is almost similar to the situations where the real time detection is needed [Refer Fig.2(i) and Fig.2(ii)]. The main limitation of the SSD framework is that it is showing less performance in situations like dark, night, foggy and rainy conditions. During such situations the detection becomes more difficult and accuracy is reduced. During night conditions, if the light from the headlight is enough for the detection but a compromise with accuracy has to be made.

YOLO makes the object detection simpler and highly efficient. The detection in the YOLO model is of higher accuracy with the ability to detect even small objects in real-time when sudden driving decisions have to be made. It identifies all objects in an image along with their class label and bounding boxes. The performance of YOLOv2 is very much lesser when compared to SSD's performance. Unlike SSD, YOLOv2 detects more images than SSD in a particular frame. But the accuracy or confidence level of detection is less. The objects which couldn't be detected by SSD, especially some small objects are easily detected and located by the YOLOv2 framework [Refer Fig.3(i) and Fig.3(ii)]. During the environmental conditions like dark, rainy or foggy atmosphere the detection rate is more for YOLOv2 when compared to SSD. But the accuracy of YOLOv2 is not very high as well as all the objects are not caught up by the processing system.

(i)

(ii)

Fig.2. Output of SSD

(i)

(ii)

Fig.3. Output of YOLOv2

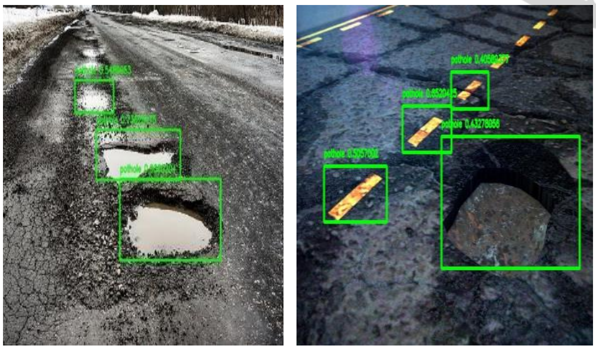
 DOI : 10.35291/2454-9150.2021.0227

Unlike other detection systems YOLOv3 is showing much higher performance when compared to other two frameworks. In normal visible light, the detection accuracy is higher and the number of objects detected also has a very drastic increase. This is mainly because of the architecture used in YOLOv3. For detecting far away objects in a real time system, YOLOv3 is the better option. For each nearby object its confidence level is more than 90% [Refer Fig.4(i) and Fig.4(ii)]. In a dark, rainy or foggy atmosphere, this framework is showing very high precision and accuracy in identifying objects, both nearby as well as far away. In some images the YOLOv3 has detected the objects which couldn't even be identified by other algorithms. In the night condition the detection becomes a little bit difficult because of the darkness.

(i) (ii)

Fig.4. Output of YOLOv3

Simulation results show that accuracy of YOLOv3 on detecting objects is more than YOLOv2 and almost similar to SSD, even though SSD showing some limitations due to its architecture. So as per the findings, YOLOv3 is showing better performance when compared to others, but it is somewhat similar to SSD because of its accuracy. So, it's the decision of the user to choose which algorithm from SSD and YOLOv3 for the application.



(i) (ii)

Fig.5. Pothole Detection

In Fig.5(i) and Fig.5(ii) potholes are detected, recognized and classified. It can be concluded that a model with YOLO neural network succeeded in detecting potholes for asphalt pavement images. Result shows a better detection rate using the applied architecture YOLOv2. Therefore, deep CNN can be applied to detect potholes with promising results in accuracy and speed, even considering the much smaller size of real world, labelled pothole data. With more training data, the results will substantially improve.

VI. FUTURE WORKS

Real-time is particularly important for object detection models that operate on video feeds, such as self-driving cars. The other advantage of real-time object detection models is that they are small and easy to work by all developers. A higher version of YOLO is being developed by the researchers to meet those properties. Studies related to YOLOv4 and YOLOv5 are ongoing. The Results of those advanced YOLO networks can give which version is more accurate and precise. This can make a revolution in every field where image processing is necessary.

VII. CONCLUSION

In this work, we have investigated the performance of different deep learning models in detecting objects for real-time autonomous vehicle applications. As object detection is the most important part in autonomous driving, different types of object detection frameworks are compared and evaluated in-depth. According to several studies, machine learning solutions, in general, are not suitable for real-time object detection. Deep learning solutions such as YOLO and SSD provide effective vehicle detection results. Deep learning techniques can be implemented to complement other technologies in detection of vehicles and obstacles in the road environment. YOLO is very fast due to the fact that it looks at the image once but provides worse performance than SSD, which employs multi-scale features. Nevertheless, the SSD still suffers from a lower speed in treating frames. Depending on the application objectives, YOLO should be employed for extremely real-time processing due to its rapidity in treating the frame while SSD can be employed for its high accuracy in detecting small objects. Since the YOLO framework is more advantageous than other frameworks, YOLO itself is selected for object detection at real-time. A detailed comparative study conducted between the different versions of YOLO (i.e., YOLOv1, YOLOv2, YOLOv3) to get the suitable one for particular application. YOLO neural network succeeded in detecting potholes which is one of the major causes of road accidents. The result [Fig.5(i), 5(ii)] shows satisfactory detection rate using the applied architecture of YOLOv2. From the comparison, YOLOv3 is found as a better framework for the real-time application since it is faster in detecting images with high accuracy and precision than other YOLO versions.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Anoop Thomas and Mr. Ajai V Babu for his constructive comments and insightful suggestions that helped in improving the practicality and quality of this paper. We also would like to thank our

 DOI : 10.35291/2454-9150.2021.0227

institution for their support to carry out this project and providing with valuable materials related to this topic.

REFERENCES

[1] Tung Lam Pham, Huy Nguyen, Hoan Nguyen, Van Hoa Nguyen, Van Bui, Yeong Min Jang," Object Detection Framework for High Mobility Vehicles Tracking in Night-Time", *ICAIIC* 2020.

[2] Juan Du," Understanding of Object Detection Based on CNN Family and YOLO", *IOP Conf. Series: Journal of Physics*: Conf. Series 1004 (2018) 012029.

[3] Shaikh Shakil Abdul Rajjak, Abdul Kadir Kureshi" Object Detection and Tracking using YOLO v3 Framework for Increased Resolution Video", International *Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-9 Issue-6, April 2020

[4] Aleksa Corovic, Velibor Ilic, Sinisa Duric, Malisa Marijan, and Bogdan Pavkovic," The Real-Time Detection of Traffic Participants Using YOLO Algorithm ", *26th Telecommunications forum TELFOR 2018* Serbia, Belgrade, November 20-21, 2018.

[5] Arka Prava, Abhiraj Biswas, Mohana," YOLO based Detection and Classification of Objects in video records", 2018 3rd *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology* (RTEICT-2018), MAY 18th & 19th 2018.

[6] Mehdi Masmoudi, Hakim Ghazzai, Mounir Frikha, and Yehia Massoud," Object Detection Learning Techniques for Autonomous Vehicle Applications ", 2019 *IEEE International Conference on Vehicular Electronics and Safety (ICVES).*

[7] Seyyed Hamed Naghavi, Hamidreza Pourreza, "Real Time Object detection and Classification for Autonomous Driving", *8th International Conference on Computer and Knowledge Engineering* (ICCKE 2018).

[8] Antoine Mauri, Redouane Khemmar, Benoit Decoux, Nicolas Ragot, Romain Rossi, Rim Trabelsi, Rémi Boutteau, Jean-Yves Ertaud and Xavier Savatier, "Deep Learning for Real-Time 3D Multi-Object Detection, Localisation, and Tracking: Application to Smart Mobility*"* *www.mdpi.com/journal/sensors.*

[9] Lim Kuoy Suong, Kwon Jangwoo, "Detection of Potholes Using a Deep Convolutional Neural Network" *Journal of Universal Computer Science*, vol. 24, no. 9 (2018).

[10] Lubna Aziz1, Sah bin Haji Salam, Sara Ayub, "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review" *Department Computer Engineering, BUITMES, UTM Malaysia, School of Computing, School of Electrical Engineering, UTM Johor Malaysia*, 2017.

 DOI : 10.35291/2454-9150.2021.0227