WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

LANE AND OBJECT DETECTION IN SELF DRIVING CARS USING DEEP LEARNING

Palisetti Sai Geetha

Department of Computer Science and Engineering
GMR Institute of Technology, Rajam, Indias
saigeethapalisetti@gmail.com

ABSTRACT :

Nowadays, one of the most emerging trends in the current world is Self-driving cars which are getting huge responses. Self-driving cars are also called Autonomous cars and driverless cars. Automation in the world has been progressing eventually in such a way that it is transforming the world with Cars and Trucks that Drive us, instead of us driving them. These will reduce the human input to drive for long hours and can able to sense the environment around it. These automotive technologies also have the potential to improve equity, accessibility, and traffic congestion. It can be able to operate without any involvement of humans. An Autonomous vehicle makes use of cameras, sensors, radar, and Artificial Intelligence (AI) to navigate the distance between the destinations. These are particularly introduced to avoid accidents that are caused due to sleeplessness of the driver or any other factors. With the help of the YOLO (You Only Look Once) algorithm, we can navigate cars throughout the distance with ease and safety. YOLO is an effective real-time object recognition algorithm that works based on regression. YOLO uses Convolution Neural Network to detect the objects where the prediction of an entire image is done in a single runtime. There are many versions of YOLO namely YOLO V2, YOLO V3, YOLO V4, and YOLO V5, for all these versions YOLO is the base. Automated driving systems have the potential to improve efficiency and convenience.

Keywords: Self-Driving Cars, Deep Learning, YOLO Algorithm, Artificial Intelligence, Convolutional Neural Network

1. INTRODUCTION

Currently, everyone in the present generation is fond of traveling, that too with their own vehicle. Most of us are using cars rather than two-wheelers because a car has the capacity to hold a group of people. Whenever anyone tends to travel long distances for hours due to sleeplessness they have to stop somewhere and take some rest and then they can start again. By using Self-driving cars one can able to overcome this and able to reach their destination by conserving time and human input. For the past 5 years self-driving cars are the most trending and everyone is willing to buy these cars based on automatic driving, reducing of physical strain. One more benefit for choosing these cars are if a person doesn't know driving also one can able to make these cars as an opportunity to learn how to drive and to fulfil their dream. Most of the companies are showing a large interest to produce these types of cars as their value is very much high in the customer's view. Many famous companies like Tesla, Audi, Uber, Ford, Alphabet etc. A well-known application of multiple object detection is in self-driving cars, where the algorithm not only needs to detect the cars but also pedestrians, motorcycles, trees, and other objects in the frame and then draw bounding boxes around each of the detected objects.

These self-driving cars are mainly introduced to reduce road accidents which are happening at a high rate and are one of the main causes of to increase death rate in the world. It is estimated that more than a million people die due to road accidents and other people suffer from major injuries. These will lower the crashes caused due to human mistakes in different incidents. In these cars there is a perception system that is used for object detection, localization, and the surrounding things which are static as well as moving and also tracking of lanes and traffic signals accurately. Many algorithms are used to generate customized outputs in these cars mainly to detect objects apart from the weather conditions, some of the algorithms used are the YOLO algorithm, CNN etc.

2. LITERATURE SURVEY

In this paper they introduced a method named Convolution Neural Networks (CNN) which is a deep learning approach. CNN is used for image classification and recognition because of its high accuracy. U19-Net is used and it is compared with U-Net. As U19-Net is an encoder-decoder deep model which is based on the deep layers on VGG19 model. Small U-Net ,Full U-Net and U19-Net are compared and the scores of training, validation and testing sets are noted. They used R-CNN which is called Region Based Convolution Neural Network which is faster than CNN. Simply R-CNN is an extension of CNN which focus on object detection mainly where as CNN is used for normal image classification.[1]

This paper proposes a method called Faster-YOLO, which is able to perform object detection in real-time The YOLO core idea is to use the entire image as the input to the network and directly return the bounding box coordinates and class probabilities in the output layer. The method faster YOLO is based on YOLOv2 and YOLOv3 which has characteristics of end-to-end operation and directly predicts the bounding box and object class. With this method we can simplify the network structure and decrease computational and memory requirements.[2]

Dist-YOLO came into existence because normal YOLO algorithm can't able to calculate the distance of the object. YOLO is only used for object detection and in order to know the distance between the car and other objects Dist-YOLO is used. After YOLO arrival, many of researchers implemented various versions of this namely YOLOv2, YOLOv3, YOLOv4, YOLOv5 etc.DistYOLOv3 is used to perform accurate and fast calculations. While knowing the distances there are two parameters involved the first one represents the detected distance and the second one shows the ground-truth distance.[3]

This paper describes the role of Deep Learning in self-driving cars. They also mentioned about different advantages of using self-driving cars and the ability to decrease collisions and death rate. Deep learning (DL) and the factors that make DL a powerful technique in computer vision. It compares normal cars and self-driving cars and the challenges involved in it. Key features of autonomous cars and various sensors used to achieve requirements. Various deep learning like CNN, YOLO, RNN are used for scene perception and object detection.[4]

This describes Open CV , which means Computer Vision. OpenCV (Open Source Computer Vision Library) is an open-source computer vision and machine learning software library. IPs (Indoor Positioning Network) which is used to provide real-time coordinates.[5]

The frequency of accidents caused by human mistake is currently relatively high, but with the introduction of autonomous driving, it can be drastically reduced. This work aims at assisting in the field of autonomous driving by helping recognize things with the use of deep learning techniques. Object identification through computer vision is one of the major prerequisites and a huge element of autonomous driving. A cutting-edge algorithm called YOLO (you only look once) was utilized in research to detect various items that emerge on the road and classify them using bounding boxes into the categories to which they belong. The YOLO v4 weights are used to train our model specifically to recognize the objects .[6]

The use of computer vision for feature extraction and object detection in real-time is the most challenging aspect of training. The subject of enhancing image segmentation algorithms has been the subject of extensive pertinent study. In order to improve the effectiveness of autonomous vehicles, it is suggested that convoluted neural networks be used with spatial transformer networks and real-time lane detection. Vehicles will be trained using the depth of the neural network, and during the testing phase, they will learn to make decisions based on the training data. The vehicle will be able to react swiftly in the event of sudden changes to the environment in order to minimize damage or potential threat to lives. A self-driving automobile needs to be capable of more than just lane detection.[7]

3. METHODOLOGY

In this methodology, after analyzing various papers from the recent years we are going to know about

¬ Convolution Neural Networks
¬ Architecture of CNN
¬ You Look Only Once algorithm
¬ Intersection Over Union
¬ Activation Functions

Convolution Neural Networks:

Neural networks are artificial systems that were inspired by biological neural networks. These systems learn to perform tasks by being exposed to various datasets and examples without any task-specific rules. The idea is that the system generates identifying characteristics from the data they have been passed without being programmed with a pre-programmed understanding of these datasets. Convolutional networks are used for alternating between convolutional layers and max-pooling layers with connected layers (fully or sparsely connected) with a final classification layer. The learning is done without unsupervised pre-training. Each filter is equivalent to a weights vector that has to be trained. The shift variance has to be guaranteed to dealing with small and large neural networks. This is being resolved in Development Networks.

CNNs are a class of Deep Neural Networks that can recognize and classify particular features from images and are widely used for analyzing visual images. Their applications range from image and video recognition, image classification, medical image analysis, computer vision and natural language processing. The term 'Convolution" in CNN denotes the mathematical function of convolution which is a special kind of linear operation wherein two functions are multiplied to produce a third function which expresses how the shape of one function is modified by the other. In simple terms, two images which can be represented as matrices are multiplied to give an output that is used to extract features from the image. CNN's are the class of Deep Neural Networks that can recognize and classify particular features from images. These are widely used for analysing visual images. CNN provides better results than classical statistical models for image segmentation and object classification tasks.

There are three types of layers in CNN which are:

Convolution Layer:

This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size MxM. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MxM). The output is termed as the Feature map which gives us information about the image such as the corners and edges. Later, this feature map is fed to other layers to learn several other features of the input image.

Pooling Layer:

In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon method used, there are several types of Pooling operations.

Fully Connected Layer:

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.

Architecture

There are mainly two parts in CNN architecture:

1. A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction.
2. A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.

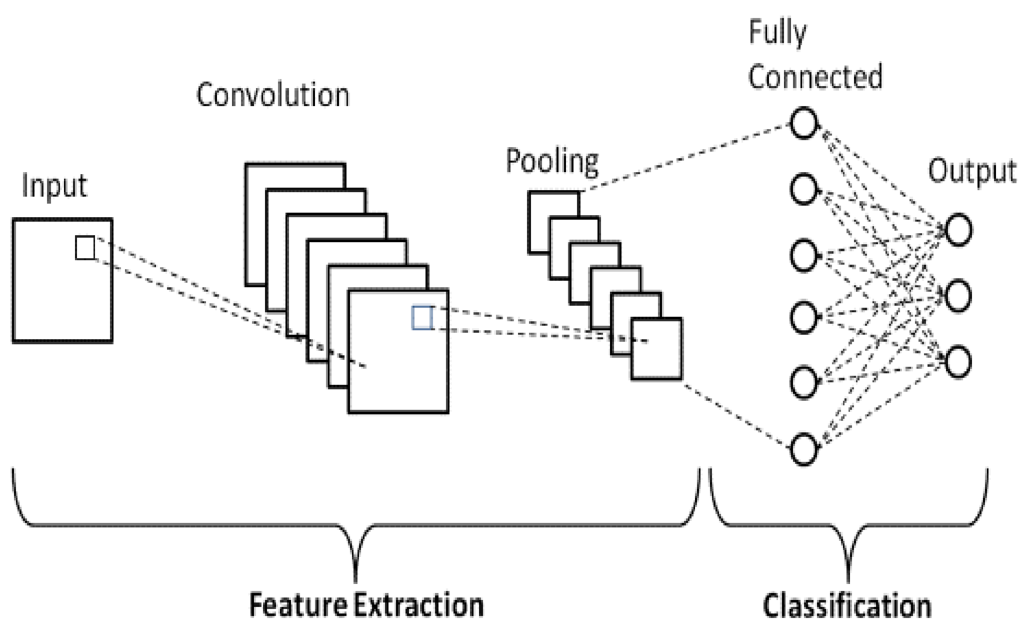

Figure 3.1: Architecture of CNN

You Look Only Once:

YOLO is an object detector that uses a single convolutional neural network for both classification and localization of the object using bounding boxes.

For object detection an image or a real time video is taken as an input using CNN. In this example, the image is divided as grids of 3x3 matrixes. The image can be divided into any number grids, depending on the complexity of the image. Once the image is divided, each grid undergoes classification and localization of the object. The objectness or the confidence score of each grid is found. If there is no proper object found in the grid, then the objectness and bounding box value of the grid will be zero or if there found an object in the grid then the objectness will be 1. In the grade (1, 1) the confidence score of the box is zero and for the grade (9, 9) the confidence score for the box is one.

Figure 3.2: Grid 1 bounding box and class values

Consider the above example, an image is taken and it is divided in the form of 3x3 matrixes. Each grid is labelled and each grid undergoes both image classification and objects localization techniques. In localization problem, the output y should have a bunch of real numbers .Pc – Represents whether an object is present in the grid or not. If present pc=1 else 0. (bx,by), bh and bw are mid point, height and widths respectively of the bounding box. In the grid (1, 1), there exists no proper object so the pc value is 0. And rest of the values of the grid (1, 1) doesn't even matter because there exist no object for detection. So, rest of the values were represented as

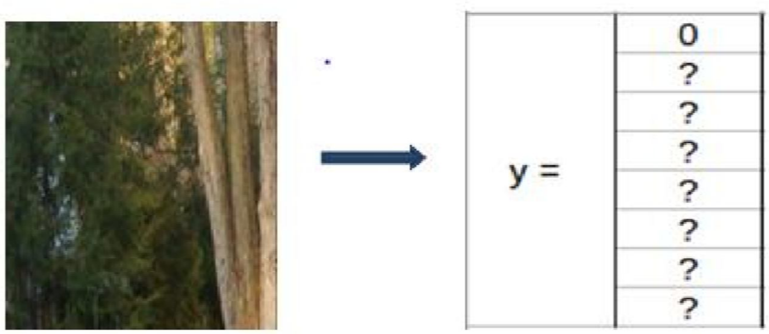

Figure 3.3: Bounding box and Class values of grid 1.

Consider a grid where we can detect an object. Clearly 5th and 6th grid of the image contains an object. For an instance consider the 6th grid, it is represented as (2, 3). In the notation of figure 4, 1 indicates the presence of an object. And bx, by, bh, bw are the bounding boxes of the object in the 6th grid (2, 3). The object detected in that grid is a car so the classes are (0, 1, 0). The matrix for the grid (2, 2) will be little similar with different bounding boxes values for bx, by, bh, bw depending on the objects position in the considered grid.

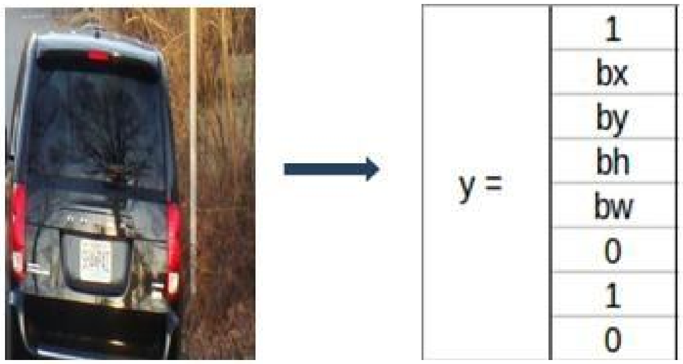

Figure 3.4: Bounding box and Class values of grid 6.

ANCHOR BOX:

By using Bounding boxes for object detection, only one object can be identified by a grid. So, for detecting more than one object we go for Anchor box. Anchor boxes are a set of predefined bounding boxes of a certain height and width. These boxes are defined to capture the scale and aspect ratio of specific object classes you want to detect and are typically chosen based on object sizes in your training datasets. During detection, the predefined anchor boxes are tiled across the image.

The network predicts the probability and other attributes, such as background, intersection over union (IoU) and offsets for every tiled anchor box. The predictions are used to refine each individual anchor box. You can define several anchor boxes, each for a different object size. Anchor boxes are fixed initial boundary box guesses. The network does not directly predict bounding boxes, but rather predicts the probabilities and refinements that correspond to the tiled anchor boxes. The use of anchor boxes enables a network to detect multiple objects, objects of different scales, and overlapping objects.

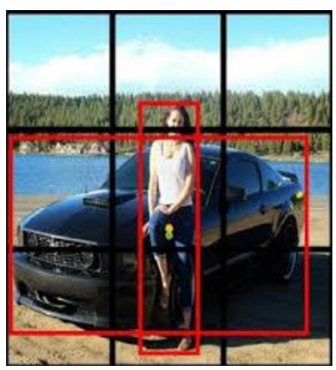

Figure 3.5: An example image for anchor box

Consider the above picture, in that both the human and the car’s midpoint come under the same grid cell. For this case, we use the anchor box method. The red color grid cells are the two anchor boxes for those objects. Any number of anchor boxes can be used for a single image to detect multiple objects. In our case, we have taken two anchor boxes.

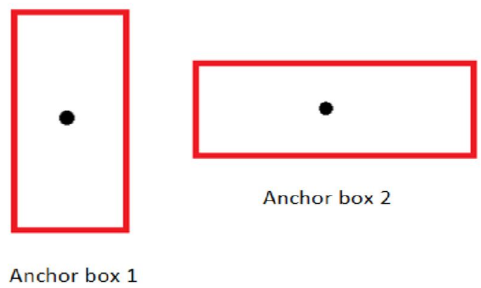

Figure 3.6: Anchor boxes

The above figure represents the anchor box of the image we considered. The vertical anchor box is for the human and the horizontal one is the anchor box of the car.

Intersection Over Union:

In IoU, it will take the actual and predicted bounding box value and calculates the IoU of two boxes by using the formulae, IoU = Area of Intersection / Area of Union. If the value of IoU is more than or equal to our threshold value (0.5) then it's a good prediction. The threshold value is just an assuming value.

Activation Functions:

Finally, one of the most important parameters of the CNN model is the activation function. They are used to learn and approximate any kind of continuous and complex relationship between variables of the network. In simple words, it decides which information of the model should fire in the forward direction and which ones should not at the end of the network. It adds non-linearity to the network. There are several commonly used activation functions such as the ReLU, Softmax, tanH and the Sigmoid functions. Each of these functions have a specific usage. For a binary classification CNN model, sigmoid and softmax functions are preferred an for a multi-class classification, generally softmax us used.

4. RESULTS AND DISCUSSIONS

This paper presents the comparison of various algorithms to identify and localize objects based on accuracy, time, and parameter values with varying sizes of the input image. U19-Net model is compared with an U-Net network in the same detection tasks for vehicles and pedestrians. Here for obtaining accurate results we are comparing full U-Net, small U-Net and U19-Net.
Summary of detection scores within networks

Vehicle detection scores :

Network	Training	Validation	Testing
Small U-Net	74.38	70.34	70.24
Full U-Net	95.99	85.61	86.44
U19-Net	94.50	86.31	87.08

Pedestrian detection scores :

Network	**Training**	**Validation**	**Testing**
Small U-Net	66.86	54.89	53.98
Full U-Net	89.97	75.39	72.25
U19-Net	88.88	77.81	78.18

By utilizing a deep network architecture U19-Net performed better than full U-Net and small UNet in these experiments with the added benefit that training can be done end-to-end. However after this various algorithms like Faster R-CNN Inception v2, Faster R-CNN REsNet50 are used for comparing U19-Net with various CNN Networks. But as we compared the CNN and all other algorithms with YOLO algorithm , it is the best algorithm for object detection till date.

Algorithm	**Accuracy**
YOLO	83.6%
CNN	75.5%
R-CNN	70.4%
Faster R-CNN	72.1%

5.CONCLUSION

By Comparing all the algorithms, The CNN,U19-Net, YOLO, Dist-YOLO, R-CNN, Faster-YOLO etc. The accuracy of detecting the lane and objects is high in YOLO algorithm as compared to others YOLO algorithm rule is easy to create and it can be trained directly on raw images.
Therefore, for lane and object detection in self-driving cars using YOLO algorithm is giving effective results of 83.6% accuracy.

References :

1. Cervera-Uribe, A. A., & Méndez-Monroy, P. E. (2022). U19-Net: a deep learning approach for obstacle detection in self-driving cars. *Soft Computing*, *26*(11), 5195-5207.
2. Vajgl, M., Hurtik, P., & Nejezchleba, T. (2022). Dist-YOLO: Fast Object Detection with Distance Estimation. *Applied Sciences*, *12*(3), 1354
3. Gupta, A., Anpalagan, A., Guan, L., & Khwaja, A. S. (2021). Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array*, *10*, 100057.
4. Newman, J., Sun, Z., & Lee, D. J. (2020, October). Self-Driving Cars: A Platform for Learning and Research. In *2020 Intermountain Engineering, Technology and Computing (IETC)* (pp. 1-5). IEEE.
5. Yin, Y., Li, H., & Fu, W. (2020). Faster-YOLO: An accurate and faster object detection method. *Digital Signal Processing*, *102*, 102756.
6. Sarda, A., Dixit, S., & Bhan, A. (2021, February). Object detection for autonomous driving using YOLO [You Only Look Once] algorithm. In 2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV) (pp. 1370-1374). IEEE

7. Kanagaraj, N., Hicks, D., Goyal, A., Tiwari, S., & Singh, G. (2021). Deep learning using computer vision in self driving cars for lane and traffic sign detection. International Journal of System Assurance Engineering and Management, 12(6), 1011-1025.
8. Sahib, F. A. A., Taher, H. B., & Ghani, R. F. (2021, May). Detection of the autonomous car robot using Yolo. In Journal of Physics: Conference Series (Vol. 1879, No. 3, p. 032129). IOP Publishing.
9. Newman, J., Sun, Z., & Lee, D. J. (2020, October). Self-Driving Cars: A Platform for Learning and Research. In 2020 Intermountain Engineering, Technology and Computing (IETC) (pp. 1-5). IEEE.
10. Yin, Y., Li, H., & Fu, W. (2020). Faster-YOLO: An accurate and faster object detection method. Digital Signal Processing, 102, 102756
11. Simhambhatla, R., Okiah, K., Kuchkula, S., & Slater, R. (2019). Self-driving cars: Evaluation of deep learning techniques for object detection in different driving conditions. SMU Data Science Review, 2(1), 23
12. Farag, W. (2018). Recognition of traffic signs by convolutional neural nets for self-driving vehicles. International Journal of Knowledge-based and Intelligent Engineering Systems, 22(3), 205-214.
13. Kulkarni, R., Dhavalikar, S., & Bangar, S. (2018, August). Traffic light detection and recognition for self driving cars using deep learning. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-4). IEEE.
14. Nugraha, B. T., & Su, S. F. (2017, October). Towards self-driving car using convolutional neural network and road lane detector. In 2017 2nd international conference on automation, cognitive science, optics, micro electro-mechanical system, and information technology (ICACOMIT) (pp. 65-69). IEEE.
15. Rajasekhar, M. V., & Jaswal, A. K. (2015, August). Autonomous vehicles: The future of automobiles. In 2015 IEEE International Transportation Electrification Conference (ITEC) (pp. 1- 6). IEEE.