Hybrid Techniques for Object Detection using Deep Learning

Swetha M S, Muneshwara M S

Abstract*: *Deep learning is a one of the major concept of Artificial Intelligence and Machine learning, which deals with the object detection task. On the other hand, a new targeted dataset is built according to commonly used existing datasets, and two networks called Single Shot Multi box Detector (SSD) and You Only Look Once (YOLO) are chosen to work on this new dataset. Through experimentation strengthen the understanding of these networks, and through the analysis of the results, learn the importance of targeted and inclusive datasets for deep learning. In addition, to this optimize the networks for efficient utilization when integrated with the necessary system or application. Further, explore the applications corresponding to these networks. The implementation includes two major concepts. The first concept is Object detection. Object detection is the process of object recognition and classification. There are several Training sets available online for training an object detection model. But the models are not trained to detect the same object from different geographical regions. The second concept is lane detection and steering suggestion. The model detects using the concept of radius or curvature of the road and also distance of the car from both the lane lines. Using these parameters it also gives steering suggestions such as move right or left by a certain distance. In addition to this it gives the distance and speed attributes of the surrounding objects such as cars, motorcycles, etc.

Finally, the model developed is capable of detecting all the parameters required in order to be integrated and to create a self-driving car and it can be used efficiently in India. Using the above parameters that are obtained from the model the car can navigate through lanes in real-time. Its improved performance is due to the fact that it can detect road specific objects and because it is specifically trained for Indian roads.

Keywords: Single Shot Multi box Detector (SSD), You Only Look Once (YOLO), Convolution Neural Network(CNN),

I. INTRODUCTION

In recent years, with the rapid development of machine learning and deep learning, a number of research areas with respect to these concepts have increased immensely. Along with this there is a continuous improvement of convolution neural networks and with this Computer vision algorithms and architecture has developed immensely. Vision starts with the eyes all things considered happens in the brain. Thus, the principle target of computer vision and AI is detecting the object and their relation, action and intention. In computer science AI research is defined as the study of "wise specialists" which implies any gadget that sees its environment and takes actions that boost its chance of effectively accomplishing its objectives.

Revised Manuscript Received on February 01, 2020.

Swetha M S, Department of Information Science and Engineering, BMSIT&M, Visvesvaraya Technological University (VTU), Bengaluru, India.

Muneshwara M S, Department of Computer Science and Engineering, BMSIT&M, Visvesvaraya Technological University (VTU), Bengaluru, India.

A. Convolution Neural Network:

The dataset is compliant to COCO, Pascal VOC and Image Net. Through experimentation to strengthen the understanding of these networks, and through the analysis of the results, learn the importance of targeted and inclusive datasets for deep learning. In addition, to this optimize the networks for efficient utilization when integrated with the necessary system or application. Fast, accurate algorithms for object detection would allow computers to drive cars without specialized sensors.

B. Self-Driving Car:

We will be using Object detection to implement the application Self-Driving car. To begin, a self-driving car (driverless, autonomous, robotic car) is a vehicle that is equipped for detecting environment and navigating without giving any human input. Self-driving cars can do this using a utilizing an assortment of methods, such as radar, GPS and computer vision. This application uses the more efficient of these two networks and implements detection of the lane curvature and the deviation of the car simultaneously, identifying traffic signals and pedestrians.

If we consider a child's eye, they take one picture in 200 milliseconds. By the age three they would have seen about hundreds of millions of images of the real world. That's a lot of training experience. So given a Object detection model, we require to give that kind of training given to the brain. It should experience training from the dataset with respect to both quality and quantity. Consequently, a Self-Driving car created using these models should be safe to be driven. For this we have to implement and Optimize Object Detection models for a Self-Driving car.

II. LITERATURE SURVEY

This paper explains us about disadvantage of SSD & its limitation over CSSD. Here the proposed framework is CSSD – shorthand for setting mindful single-shot multibox question finder. CSSD is based over SSD, with extra layers displaying multi-scale settings. CSSD utilises two additionally layers convolution layers (DiCSSD) and deconvolution layers (DeCSSD) to detect small object effectively.

CNN uses multilayer wherein each layer is hidden to handle small layer which is useful in processing several neural network. In convolution neural network object detection is done thru shot category detection.

This shot category detection identifies & converts the images into number of frames. These frames are useful in neutral network, which are in different size based on the feature extraction of image processing.

Here images are processed in different angles like Long shot, full shot, medium shot, up shot, waist shot etc. These methods in CNN make it efficient in supervised learning.

MOD is used to solve problem caused while driving in road which has slope and curve and also used to detect road condition from camera by using sensor fusion. The MOD works on image recognition, tracking ID of the object and classification. Combination of detection and classification is done for finding local & global position of object size by using sensor fusion. The camera uses 15 fps to find velocity of the objects. To get more accurate information sensor fusion should be simultaneously processed in a moving vehicle.

In normal process RADAR is used to detect moving object. By using Object confirmation algorithm we can efficiently detect object even when there is bad environmental condition.

OCA is introduced in RADAR perimeter security system which reduces false alarm rate of RADAR. In real world application YOLOv2 system is the basis for the classification of the object thru K mean++ clustering algorithm.

YOLO 9000 is improved version of YOLO as it can detect more than 9000 object categories during object detection. Improvement of YOLO9000 over YOLO recognition strategy is due to enhancement in both unique and strained approach. YOLOv2, is best in class on standard discovery undertakings like PASCAL VOC and COCO. Utilizing a unique multi-scale preparing technique the equivalent YOLOv2 model can keep running at different sizes, offering a simple exchange off among speed and exactness.

This paper considers road safety approach and vehicles movement control using stationary camera. Algorithm uses camera for image processing, detection of number of vehicle movement and detection of road markings. By using this algorithm proper vehicle movement can be ensured, wrong parking of vehicle can be ensured. All these helps in avoiding traffic congestion and accidents. This algorithm can also be used in smart camera system.

III. PROPOSED SYSTEM

The proposed model can detect globally inclusive images and model is created using YOLO. Neural networks and it is data initially trained using the Inclusive Images dataset. Furthermore, it is trained using a newly curated, more representative dataset. A machine learning model should be efficient and accurate even when it learns from imperfect data sources. This model shows an improvement in the accuracy when detecting geographically diverse images especially for Indian roads and implement the application of Self-Driving car.

A.Overall Architecture

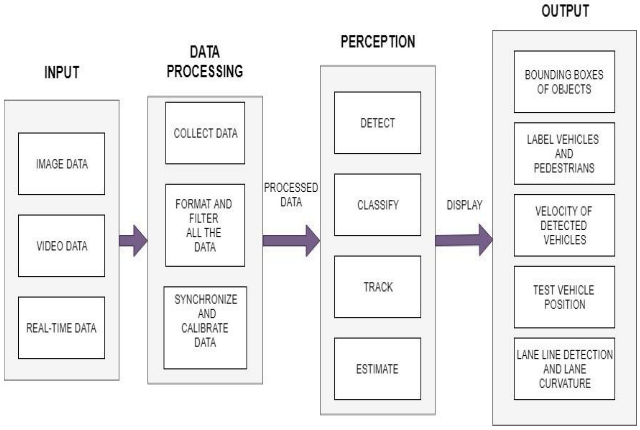


Figure. 1 Overall System Architecture

Input:

The input is the data that has to be given to the system to train. The input is given in three modules that is image data, video data and real time data. Once the input I sgiven the the data processing should be done.

Data processing:

The data processing comprises of collection of data, where the input given is collected, then the formatting of all the data that has been collected and the final step would be to synchronize and calibrate data.

Perception:

Once the data is processed it will be detected first then classified and then for velocity of the vehicles, lane and road curvature it will be tracked first and then estimated.

Output:

Once the perception is done then the output is displayed as the bounding boxes around the objects along with the confidence percentage. The velocity of the car along with the distance from the vehicle is determined

IV. METHODOLOGY

A. Collection of data

1. Open Images Dataset V4

Open Images is a gigantic dataset which contains near 9 million pictures. All pictures accompany marks that were arranged physically by expert annotators. The dataset is separated into the training (9 million pictures), approval (41k pictures), and test (125k pictures) set. Also, it is the biggest existing dataset with object area explanations. These comments have been drawn physically by expert annotators with the end goal to guarantee precision and consistency. The topic in the pictures is differing in nature. There are 8.4 objects per picture by and large in this dataset. To include what tops off an already good thing, information is explained with picture level marks that range a large number of classes

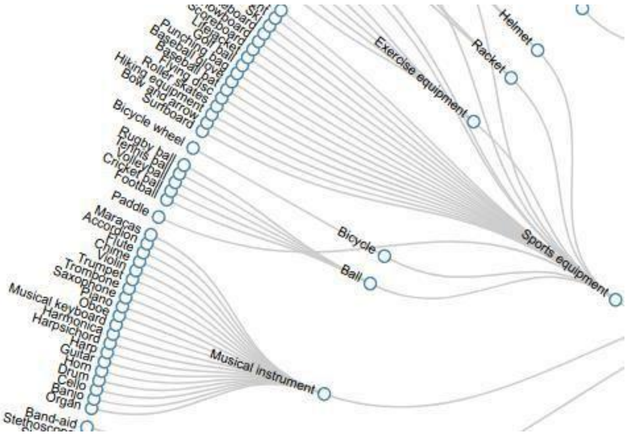


Figure 2. Open Images Dataset V4

2. DETRAC

UA-DETRAC is a testing genuine multi-object detection and multi-object following benchmark. The dataset comprises of 10 hours of videos caught with a Cannon EOS 550D cameras at 24 unique areas at Beijing and Tianjin in China. The videos are recorded at 25 outlines for each second (fps), with goals of 960×540 pixels. There are in excess of 140 thousand casings in the UA-DETRAC dataset and 8250 vehicles that are physically commented on, prompting an aggregate of 1.21 million named bounding boxes of objects.

3. PASCAL VOC

Gives institutionalized picture informational collections to object class acknowledgment and normal set of instruments for getting to the informational indexes and comments. It likewise enables assessment and correlation of various techniques.

4. ImageNet

ImageNet is a picture dataset sorted out as per the WordNet pecking order. Each important idea in WordNet, perhaps portrayed by numerous words or word phrases, is known as an "equivalent word set" or "synset". There are in excess of hundred thousand synsets in WordNet, lion's share of them are things. In ImageNet, we plan to give all things considered 1000 pictures to represent every synset. Pictures of every idea are quality-controlled and human-explained.

C.Image Processing

1. Image Acquisition

Two exceptionally prominent techniques for creating a digital picture are with Digital camera or Flatbed Scanner. Also, here, for training the model we are getting datasets from the previously mentioned sites. For the most part, the picture obtaining stage includes pre-processing, for example, scaling.

2. Pre-processing

The pre-processing of picture goes for specifically evacuating the repetition present in caught pictures without influencing the subtle elements that assume a key job in the general procedure. This includes the accompanying two phases.

3. Image Resizing

Re-sizing of a picture is performed by the procedure of the insertion. It is a procedure which re-tests the picture to decide esteems between characterized pixels. In this manner, resized picture contains pretty much pixels than that of original picture.

4. Filtering

Vulnerabilities are brought into the picture, for example, arbitrary picture clamor, halfway volume impacts and power non consistency ancient rarity (INU), because of the development of the camera. This prompts data loss, SNR gain and debasement of edge and better points of interest of picture. Spatial channels are utilized for commotion decrease.

5. Segmentation

Segmentation of a picture is a procedure of partitioning the picture into homogenous, self-predictable areas comparing to various objects in the picture. It isolates picture into significant areas. Picture can be portioned utilizing fundamental properties of highlights of picture like edge or surface.

Edge detection: Edge is the limit between two locales with moderately unmistakable dim dimension properties. In a nonstop picture, a sharp force progress between neighboring pixels is considered as an edge. In this manner, in the zone of picture processing and computer vision, edges or shapes of pictures give profitable data towards human picture understanding.

6. Feature Extraction

Feature extraction is a low dimension picture processing task which is typically executed as the main activity on a picture. A feature can be characterized as the "important" portion of a picture. Step edges, lines and intersections more often than not pass on the most pertinent data of a picture; henceforth it is essential to recognize them dependably.

7. Representation

The last advance of image processing includes changing over the information to a frame appropriate for computer processing is fundamental.

D.Training and Optimizing the Object Detection Models (YOLOv2 and SSD)

1.Rudimentary Algorithm for Object detection

STEP 1: A model or algorithm is utilized to produce regions of interest or region recommendations. These region recommendations are a substantial set of bounding boxes traversing the full picture (that is, an object localisation segment).
STEP 2: Secondly, visual features are separated for every one of the bounding boxes, they are assessed and it is resolved whether and which objects are available in the proposition dependent on visual features (i.e. an object order segment).
STEP 3: In the last post-processing step, covering boxes are joined into a single bounding box.

2. Training the Algorithm

The Algorithms are prepared utilizing the datasets gotten. In this progression, we will utilize our information to incrementally enhance our model's capacity to foresee objects. The dataset is isolated into train and test set. It's the train set that will be utilized to prepare the model and the test set will be utilized for validation. It is utilized to gauge how well our model has been prepared (that is needy upon the span of your information, the esteem you might want to foresee, input and so forth.) and to assess demonstrate properties (mean blunder for numeric indicators, order mistakes for classifiers, review and exactness for IR-models and so on.).

The validation stage is part into two sections. In the initial segment you simply take a gander at your models and select the best performing methodology utilizing the validation information (=validation) and in the second you gauge the precision of the chosen approach (=test).

3. Implementation of Object Detection for Self-Driving car

Finally, in light of the earlier research both of the model, YOLOv2 or SSD, or a gathering (i.e. Joining the models/consolidating the perspectives/stacking) of the two will be utilized to execute a recreated Self-Driving Car. This vehicle will have the capacity to identify the activity flag and objects in the pathway. Notwithstanding this it will have the capacity to distinguish path ebb and flow and the deviation of the vehicle at the same time.

E. RESULTS AND DISCUSSION

A.Static Images Detection:

Static Images Detection Optimization is a module that deals with the detection of objects in the images from the dataset. The Image is first Detected and the classified. Real time Images Object Detection Optimization is a module that deals with the detection of objects in the real time video that is shot.

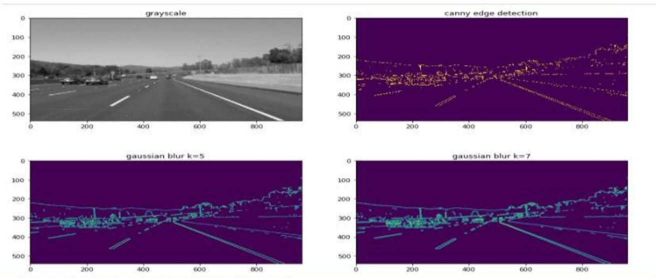


Figure 3. Converting grey scale for edge detection in video

1.Real Time images:

Real Time Inclusive Image is the that module deals with the detection of real time inclusive image in consideration with the Indian Roads.

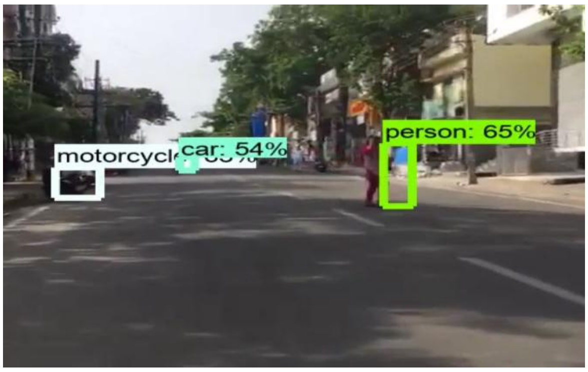
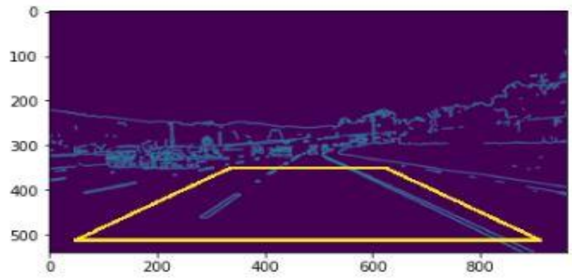


Figure. 4. Indian Roads object detection

2.Lane Detection and Lane Curvature

Lane Detection and Lane Curvature and Steering is the final module in which the system would be able to detect and classify the objects along with the lane curvature in the simulation video of a Self Driving Car.



```
copied = np.copy(blurred)
plt.imshow(region_of_interest(copied,vertices))
plt.show()
```

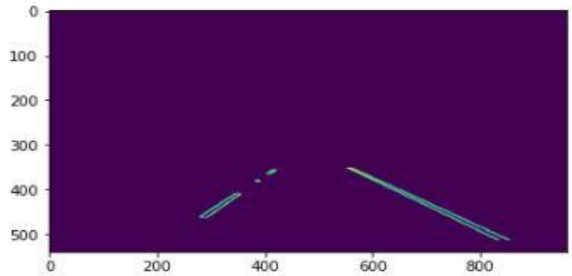


Figure. 4. Detecting the area of Interest

Figure 5. Road Smoothing

Figure 6. Object classification and detection

The final results are successfully detecting and classifying the objects on Indian roads along with the confidence percentage and also determine the speed of the vehicle from the center of the call and the lane detection and along with the radius of the curvature of the road.

Figure 7. Object classification and detection

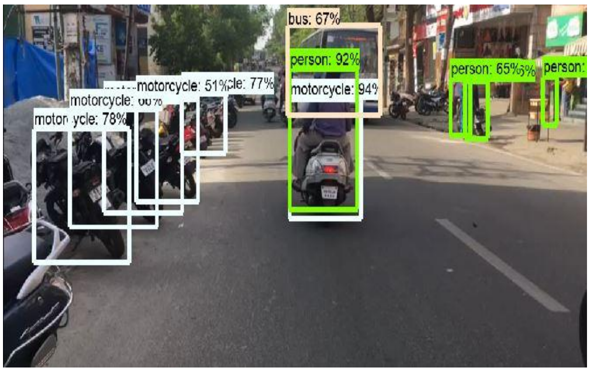


Figure 8. Object classification and detection real time Indian roads.

F. CONCLUSION & FUTURE ENHANCEMENT

The current module detects and classifies the objects from the static images and also is capable of real time object detection. It is specifically capableof detecting road specific objects which makes it faster and more efficient for autonomous cars. It is capable of detecting lanes, lane lines, lane curvature and position of the test vehicle and the traffic. This data can thus be used to provide steering suggestions. What makes this model effectively and efficiently generate data for autonomous cars is that all happens in real time. This model can further be improved by creating a model that can detect inclusive images even when sufficient data is not available to train i**t.** The developed model can be integrated and deployed in hardware system and can be molded according to the necessity. Taking images from multiple cameras and performing object detection can be done and also there is a scope for vehicle to vehicle communication and connected roads.

REFERENCES

1. Wei Xiang, Dong-Qing Zhang, Heather Yu, Vassilis Athitsos " Context-Aware Single-Shot Detector", Winter Conference on Applications of Computer Vision, IEEE,2018.
2. Lihong Yang, Liewei Wang, Shuo Wu"Real-Time Object Recognition Algorithm Based on Deep Convolutional Neural Network", International Conference on Cloud Computing and Big Data Analysis, IEEE, 2018.
3. Jinwoo Kim,Yongbon Koo,SungHoon Kim," MOD: Multi-camera based Local Position Estimation for Moving Objects Detection", International Conference on Big Data and Smart Computing, IEEE, 2018.
4. Deokkyu Jung, Jeong-Woo Son, Sun-Joong Kim "Shot Category Detection based on Object Detection Using Convolutional Neural Networks", International Conference on Advanced Communications Technology Vis. (ICACT), Feb 11-14, 2018.
5. Boris A. Alpatov, Pavel V. Babayan, Maksim D. Ershov" Vehicle Detection and Counting System for Real-Time Traffic Surveillance", 7thMediterranean Conference on Embedded Computing, June 11-14, 2018
6. Joseph Redmon, Ali Farhadi," YOLO9000: Better, Faster, Stronger", Conference on computer Vision and Pattern Recognition, IEEE, 2017.
7. Chengcheng Ning, Huajun Zhou, Yan Song , linhui Tang " INCEPTION SINGLE SHOT MULTIBOX DETECTOR FOR OBJECT DETECTION", Proceedings of the IEEE International Conference on Multimedia and Expo Workshops Vis. (ICMEW) July 10-14, 2017.
8. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi" You Only Look Once: unified, Real-Time Object Detection",International Conference on Computer Vision and pattern Recognition, IEEE, 2016.
9. D. Biswas, H. Su, C. Wang, J. Blankenship, A. Stevanovic, “An automatic car counting system using OverFeat framework,” Sensors, vol. 17, 2017, pp. 1-13.
10. K. Sundbeck, “Car detection with and without motion features,” Master’s Thesis, 2016. S. Choudhury, S.P. Chattopadhyay, T.K. Hazra, “Vehicle detection and counting using Haar feature-based classifier,” Industrial Automation and Electromechanical Engineering Conference, 2017, pp. 106-109.
11. P. Babayan, N. Shubin, “Line detection in a noisy environment with weighted Radon transform,” Image Processing: Machine Vision Applications VII, 2014, pp. 1-6
12. V.A. Vittikh, P.O. Skobelev, “The multi-agent models of interaction in demand-resource networks,” Automatica and Telemechanica, 2003, pp. 177-185.
13. P. Babayan, N. Shubin, “Detection of curved lines and estimation of their parameters on images,” Proceedings of 6th Mediterranean Conference on Embedded Computing (MECO), 2017, pp. 201-204.

AUTHORS PROFILE

Prof. Swetha M S received B.E. and M. Tech from VTU in 2008 and 2013 respectively. During 2014 to present affiliated to BMS Institute of Technology and Management as Assistant Professor at Department of IS&E, actively involved in research area like Wireless Sensor Network, MANET’s and Cloud Computing. Several papers published in reputed Journals and conferences. Currently, pursuing PhD in Computer Science under VTU.

Prof. Muneshwara M S received B.E. and M. Tech from VTU in 2005 and 2012 respectively. During 2006 to present affiliated to BMS Institute of Technology and Management as Assistant Professor at Department of CS&E, actively involved in research area like Distributed Network security and Cloud Computing. Several papers published in reputed Journals and conferences. Currently, pursuing PhD in Computer Science under VTU.