European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 4, 2020

REAL TIME OBJECT DETECTOR FOR VISUALLY IMPAIRED USING OPEN

[1]V.Balaji, [2] S. Kanaga Suba Raja , [3] C.J. Raman , [4] S.Priyadarshini , [5] S.Priyanka[5], [6] S.P.Salai kamalathai
[1,2,4,5,6] *Easwari Engineering College, Chennai,* [3] *St. Joseph's College of Engineering,* Chennai,

ABSTRACT
The goal of the present project is to model an object detector to detect objects for visually impaired people and other commercial purposes by recognizing the objects at a particular distance. Available old techniques for object detection needed large training data it takes more time and its quite complicated and it's a difficult task. Hence, in this paper, we propose computer vision concept to convert object to text by importing the pre trained dataset model from the caffemodel framework and the texts are further converted into speech. This system supports multiple object detections on a single screen, that is it helps in real time object detection. This paper involves the idea and methodology and system architecture to implement the system along with the obtained intermediate results and performance analysis of the tools used in the proposed system. This system can further be implemented in any portable devices for the visually impaired people to detect objects at a particular distance from them.

Keywords: real time object detection, Caffemodel framework, Mobilenet SSD, Deep Neural Network

1. INTRODUCTION

Object detection is used in many scenario. Conventional methods of these object detection depends on huge amount of datasets and it also takes large amount of time to train these datas. Training of small or unseen objects is a more challenging task. Human brains and visual systems are more accurate and faster in detecting objects in real time and has conscious thoughts in detecting obstacles. Due to the availability of large amount of data and with more advanced technologies and better working algorithms, classificarion and detection of multiple objects in the same frame has become easy with high accuracy.

With more advancements in technologies, object detection can also be done using pre trained models available. One such model is the Mobile net which is being imported from the cafemodel framework of open computer vision. This helps in real time object detection and thus reduces the computational time of the overall system since training of data has been done already. When building object detection networks, existing network architecture is normally being used, such as Very Deep Convolutional Network(VGG) or Residual Neural Network, which are then implemented inside the object detection network. But the problem with these networks are, they are large sized network.

Network architectures like the above mentioned are difficult to be implemented in constrained devices due to their large size and more computational complexities. To overcome this problem, we use Mobilenet which can be used in constrained devices like smart phones. MobileNets has streamlined architecture with depth-wise separable convolutions to build a better and improved deep neural networks. Mobile Nets vary from traditional Convolutional Networks through the usage of depth wise separable convolution. The general idea behind depth wise separable convolution is seen in two stages:

1. *3×3* depth wise convolution.
2. *1×1* point wise convolution.

This helps in reducing the number of parameters used and are also much resource efficient.

2. RELATED WORKS

Visually impaired people face a lot of problem. This paper is to provide an assistance for visually impaired. This system combines techniques that are based on sensors along with computer vision concept to result an economically feasible solution. This system uses an algorithm known as novel obstacle based on image depth information and fuzzy logic. Through these techniques, visually impaired people would be able to avoid the front obstacles by detecting the objects in front of them. This system helps in assisting 6 visually impaired people indoor and outdoor. The hardware requirements of the system includes camera, GPS, Wifi, microphone, compass, gyroscope and microcontroller. At software level, it requires multisensory data and computer vision approaches.

The system provides a 100% obstacle avoidance with 96% detection accuracy. It helps in safe traversal and gives a high performance. The system is considered to be more reliable, simple, wearable and economically accessible.
This system has got few disadvantages also which comes into note. The system would not be compatible when it comes to cases like detecting large objects like walls, doors since their sizes play a vital role as they may not be detected in the frame. So there might be difficulties in detecting by finding the difference between backrounds and foregrounds. Therefore results can be obtained at the best way usig the ultrasonic sensors. Apart from thee, ultrasonic sensors may give a good result in terms of reliability, efficiency in obstacle detection. This helps in measuring the distance of obstacle along with the computer vision techniques thereby increasing the accuracy. [1]

This paper aims in independent movement of people with visual impairment. This is done by designing a portable constrained assistance for the visually impaired people to detect the traffic light. The system is designed based on the algorithm known as AdaBoost. This approach is faster and robust in object detection. This system is enhanced along with a flexible parallel architecture on the field programmable gate array (FPGA) platform. The main images and weak classifier's confidence are calculated in parallel. The weak classifiers parameters are trained using AdaBoost algorithm using MATLAB software and then configured on FGPA platform. Fgpa is found top be more flexible and consumes less power.

Experiments show that this system will be able to detect the traffic lights in the video at the rate of about 30 frames per second. This task is done on a real time and hence it detects the position of the traffic light in each of the frames in the video in real time and the system is also considered to be an efficient one.
Though the system is efficient, it has got few limitations that it takes large time to train the weak classifiers. The future work should be designed by improving the designed system by introducing more types of weak classifiers, in order to achieve a higher detection rate and accuracy. [2]

To provide navigation and orientation, this paper proposes a an electronic device known as NavCane. This device helps in obstacle free path for the visually impaired people both indoor and outdoor. This device sends prior information about the obstacle to the person using it without any information overload and the information is sent to the person through actile and auditory methods. It has many components like ultrasonic and wet floor sensors, GSM and GPS modules, a gyroscope, a vibrator motors radio-frequency identifier, a global system for positioning the module and batteries. This system is assessed by 80 visually impaired people and was successful at various scenarios.

Unlike other electronic devices, this NavCane device detects obstacles present in known indoor settings also. It is considered to be a low power consuming embedded device and a low power system. Analysis says that this NavCane improves the obstacle-free performance much more than the white cane. [3]

This paper aims in developing fast deep neural network for the purpose of real time video object detection by concentrating more on knowledgebale training of data and predicted area of interests. This system develops a framework for training of datasets by deep neural network using limited samples along with

cross network knowledge projection which helps in improving the performance and also reduces the computational complexity. Here the training process is regulated by learning projection matrix by projecting the knowledge and visual representation of the teacher-level network from its intermediate layer to the intermediate layer of the child level network. This paper focuses on the following:

1. This system proposes an architecture for transferring the data knowledge from large pre-trained teacher level network to thinner child level network.
2. It develops a fast method to find out the candidate regions of interest which contain the targetted objects.
3. It establishes an analytical model to calculate the support regions at each convolution layer and integrates this along with the existing object detection framework using deep convolutional neural networks.

Experiments are done to prove that this system reduces network computational complexity 16 times and improves the network performance at a significant margin. [4]

Object detection in real time is very difficult on an embedded device due to limited memory and computational power. Therefore this paper proposes a design without reducing detection accuracy. In this work, a light weight object detection method called as Mini-YOLOv3 . For the Darknet based backbone, depth seperable convolutions and point-wise group convolutions to reduce the size of the parameter in the network. The process of first reducing the dimensions and then increasing the dimension is adopted along with the residual structure. Boundary effect of point wise group convolution is suppressed by channel shuffle. A Multi-Scale Feature Pyramid (MSFPN) is proposed for fusing 3 feature maps in the backbone to generate base feature. Finally, Mini YOLOv3 is tested with MS-COCO dataset and found that Mini YOLOv3 gives more accuracy and ½ the detection time and achieves a high speed and very competitive performance in embedded applications. [5]

You Only Look Once, also known as YOLO is one of the well known methods for fast real time object detection that is being popularly used. For better results in object detection in a particular environment, Tinier-YOLO concepts is used and proposed in this paper for minimizing the size of the system for better detection results, accuracy and performance. To overcome all the problems faced with Tiny-YOLO, network performance, detection speed and accuracy is focused. One of the main challenges while introducing fire modules in Tiny-YOLO V3 is to find out the total number of these modules and their places in the system. Then it is also important note that difficulties arises in the connectivity style between the modules to obtain better results in terms of detection, accuracy and performance. This system adopts a dense connections among the fire modules so that the movement features could be strengthened and also to ensure flow of information between the network is at the maximum. But further, the model size cannot be reduced because if it is reduced, then detection accuracy may be affected on a large scale. Therefore, the pass through component is used in this system in order to face the problem of reduction in model size. These pass through layers helps in combining the feature maps extracted from the front layers to obtain the proper features. Finally, to reduce the computational cost factors, this system proposes to remove batch normalization from the modules along with the overall performance. [6]

3. OVERVIEW OF THE PROPOSED SYSTEM

Real time object detection is widely used Open CV concept that mainly concentrates on identifying, detecting objects belonging to different classes in the input image. Detecting the objects along with its unique features and backround details is done using different techniques and methods.. One among them is creating bounding boxes so that pixels of all the aspects of the objects are found and detection is done accordingly with certain computations.

The proposed system's goal is to detect objects that are shown in front of it and convert the detected object to speech. Object detectors can be made as a portable device that can be fitted or carried anywhere

to detect and recognize objects. This system involves importing a pre trained model from the caffemodel framework which is a deep learning framework. The caffemodel framework has Mobilenet model which contains pre trained objects which is being imported for object detection. The model uses Deep Neural Network (DNN) algorithm for classifying the objects and detection them. The system is trained in such a way that it takes objects which as confidential level above 50%. This helps in better detection of multiple images in a real time.

A deep neural network (DNN) is a type of, Artificial Neural Network which has got multiple hidden layers between the input and output layers. It is similar to other network models such as shallow ANNs. DNNs can model complex non-linear relationships. The main usage of a neural network is to receive a set of inputs that is in the form of images or datasets, perform progressively all the complex calculations on them, and give output to solve real world problems like classifications or detections.The system requires a portable device which integrates the code, camera and sensors required for object detection. The main working model runs in python code in the laptop and the output is obtained through the web camera on the laptop or any personal device. Before running the code, the necessary packages must be installed and the pre-built model should be properly imported. The datasets are present in the model and its completely trained . In case of using any portable device for object detection, then ultrasonic sensors are required for sensing the object at a distance, camera to capture the object and the code is accessed through cloud.

4. SYSTEM DESIGN

Firstly, the pre trained Mobile net model is imported from cafemodel framework It is preprocessed to obtain detection of objects with confidence more than 50%. In this system the object is being captured first, then it undergoes prediction to detect the objects that features the trained objects obtained from the Mobile net. Further after detection of objects, it is converted to text which is then converted into speech using google gtts. The audio for the speech is obtained through the speech receiver.

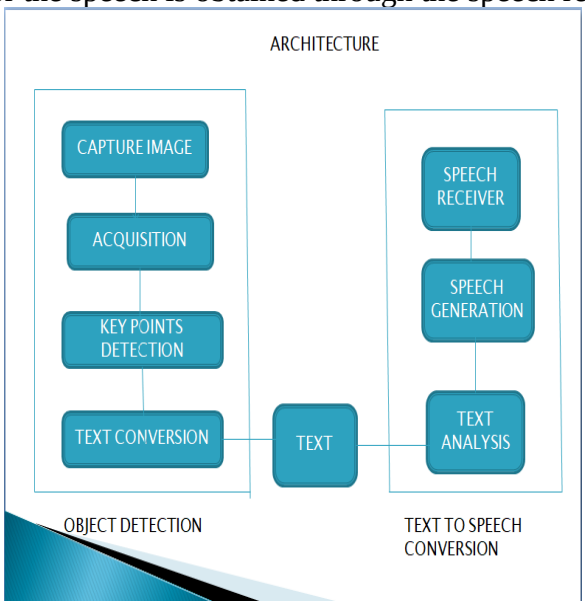

Fig.1 Architectural diagram of real time object detector

5. TOOLS AND METHODOLOGY

The proposed system involves downloadingthe Mobile net model from the caffemodel framework and the importing the pre-trained model along with the necessary packages. The Mobile net model has about 30 trained datasets which would be useful in our system to detect the object. Caffemodel object detection API is a framework or model used for creating deep learning networks that solve object detection problem.

Here, Deep Learning Object Detection is being focused since Caffemodel uses Deep Learning for computation.

Some necessary Prerequisites are as follows;

1. Python 3.5
2. Caffemodel framework
3. OpenCV
4. Numpy 1.14

Download the pre-trained model and python codes to look at how to detect and recognize objects in camera feeds. To build our deep learning-based real-time object detector with OpenCV , need to access the webcam in an efficient manner and apply object detection to each frame. The web camera is used as the default device which is pre-installed in our system . Commonly, laptop would be the webcam device. The live video stream using OpenCV is started and sent. As OpenCV can access the device webcam , output video frame is seen for any detected objects. Detection speed depends on the speed of computer's CPU or GPU resources. The sample or example screenshot result of applying deep learning is shown below in fig.2 and fig.3.

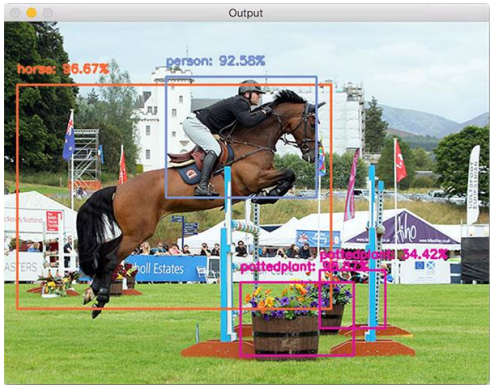

Fig.2 example image 1

Fig.3 example image 2

Steps Involved in implementing the single shot object detection algorithm.

1. The initial step is to load the pre-trained object detection network using Open CV's deep neural network module.
2. This passes the input images through the network and obtain the output bounding box of each object in the image.

3. Now we write the code to print the name of the detected object, their confidence scores and other specifications that are needed in the output.
4. Finally, we look at the output of MobileNet Single Shot Detector for our real time input images.

5. INTERMEDIATE RESULS

After running the code in python, the result will be obtained in the command prompt. Initially, the pygame will be loaded then the model that is being used in this system will be loaded (fig.4).

Fig.4 Screenshot of model being loaded

Then the video stream will be ready to detect the objects in front of it, that is the laptop's web camera will turn on automatically to capture the objects (fig.5)

Fig.5 Screenshot of video streaming

Then the system tries to give the object's name in the form of speech. Therefore it tries to enable the speech part that is done by google gtts (fig.6)

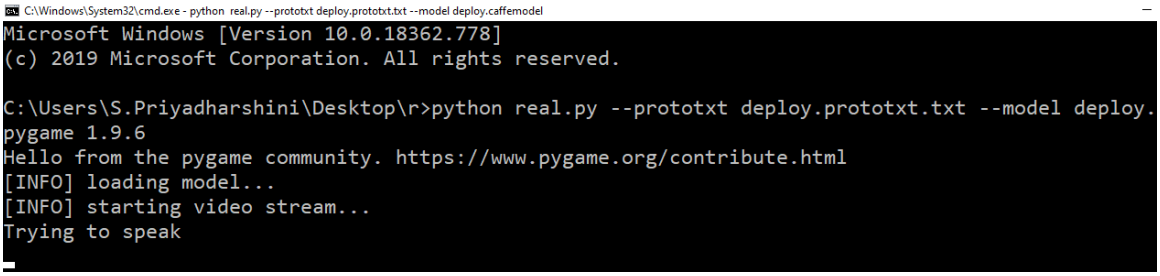

Fig.6 Screenshot of speech start

Finally it displays the name of the detected object along with the voice that tells the object name. The object that is being detected will be displayed along with its confidence level

6. RESULT AND PERFORMANCE ANALYSIS

The system uses Mobile net SSD model for training the datasets. Single Shot object detection ie SSD detects multiple objects in a single frame by taking a single shot. It comprises of two parts:

1. Initially, it extract feature maps, and
2. Apply convolution filters to detect the objects present.

SSD is developed by Google researcher teams to balance between different object detection methods or algorithms like YOLO and RCNN. There are two models of SSD available in specific

1. SSD300: In this model the input size is constant and is taken as 300×300 which is used in images with least resolutions. This model has faster processing speed but it is less accurate than SSD512
2. SSD512: In this model the input size is again constant and is taken as 500×500. This model is a better one than SSD300 as it is used in images with high resolutions and it is also found to have more accuratcy rate than other models.

SSD is faster than other models because it takes single shot for object detections wherein the R-CNN model takes two, one for generating region proposals and the other for object detection. YOLO is one of the open-source methods available for object detection and is used to recognize objects in images or videos quickly whereas in SSD (Single Shot Detector) a convolutional network comes into action which runs on input image only one time and generates a feature map. SSD is a better option as detection of objects can be done on a real-time video and the correctness is also very exactly seen. Therefore, SSD is considered to be a healthier recommendation.

We are using MobileNet-SSD model which is a caffe implementation of MobileNet-SSD detection network which has got its pre-trained weights.. When both the MobileNet architecture and the Single Shot Detector (SSD) framework are combined, it results in a faster and more efficient deep learning-based method for object detection. SSD with MobileNet provides the best accuracy within the fastest detectors. Resnet or VGG or alexnet are similar to Mobilenet but the size of the network is large and involves more computational complexities. But in Mobilenet the architecture is simple.

The different results after the executions are shown below through screenshots. Here few sample objects like car, bottle are shown infront of the laptop camera and the output is obtained in real time.

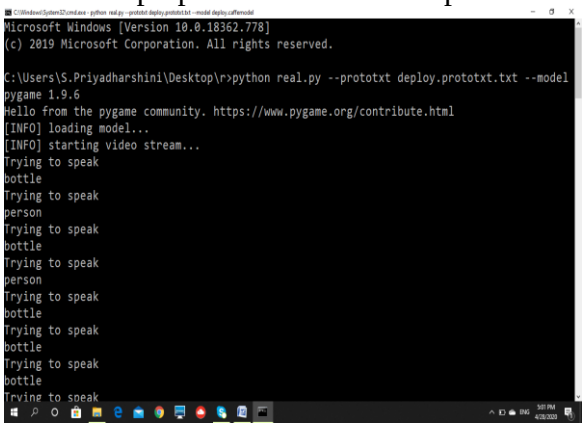

Fig.7 Sample output

Fig.7 shows the execution of the model in command prompt.

Table.1 Detection results

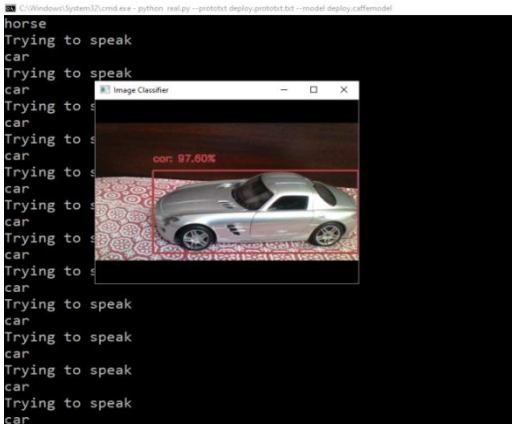

Fig.8 Execution result 1

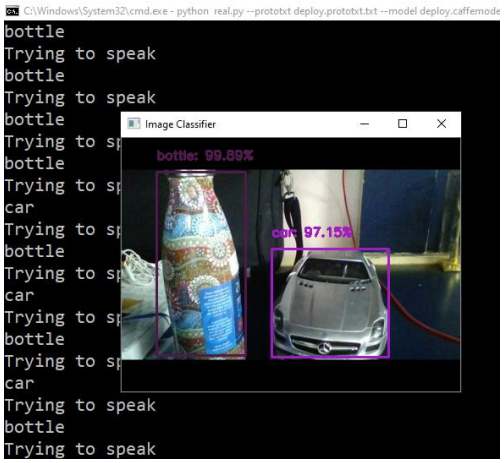

Fig.9 Execution result 2

Fig.8 and Fig.9 are examples showing how multiple images are being detected along with its confidence level on a single screen using the implemented system.
Detection results of the trained MobileNet-SSD algorithm for sample 6 objects are shown in table 1.
Accuracy of the proposed system: The proposed system shows upto 95% accuracy in detection of objects. The model is trained in such a way that it detects objects correctly with the extracted features and boundary boxes.

Name of the input	No of inputs given	No of times correctly detected	Classification Accuracy in %
Person	30	30	100%
Car	40	38	95%
Bottle	20	18	95%
Chair	30	27	96%
Potted plant	20	18	90%
Sofa	30	28	98%

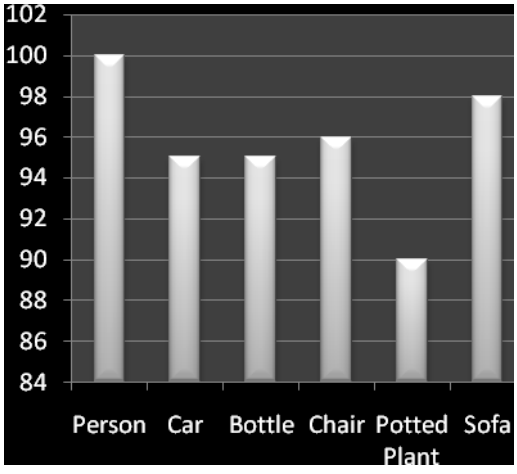

Fig.10 Sample graph of objects vs accuracy

Limitations:

1. The system might find it little difficult to detect objects when there are some backround disturbancs or if the images are blurred.
2. Mobilenet SSD is less effective in accuracy while detecting smaller objects.
3. It would be easy for the system to detect the objects if its properly captured in the camera since confidence percentage might change if the object that has to be detected is not captured properly.

7. CONCLUSION

The above system implements object detection method by detecting objects with high level confidence. Use of pre trained model is much more efficient since it reduces the computational time and computational costs and gives a more efficient system. In such pre trained model, we load the file, then preprocessing is done. After this process the input is given in real time and the output is obtained. This system helps in detecting multiple objects in real time. Therefore separate training of images are not necessary since the pre trained model includes all datasets and training of datasets as well. At the end, the detected objects names are displayed as text which is then converted to speech. Since it has an extra voice feature, this system would be useful for visually challenged or deaf and dumb people. Therefore, it has a wide variety of applications.

This system could further be developed by bringing it into any electronic devices and a better system can be developed to avoid background disturbances. Mobilenet architecture should be improved in future to detect objects with less confidence level and objects that are smaller with much more higher accuracy.

8. REFERENCES

[1] WafaM.Elmannai,KhaledM.Elleithy. “A Highly Accurate and Reliable Data Fusion Framework for Guiding the Visually Impaired”. IEEE Access 6 (2018) :33029-33054. [1]

[2] Xue-Hua Wu,RenjieHu,Yu-Qing Bao. “Parallelism Optimized Architecture on FPGA for Real-Time Traffic Light Detection”. IEEE Access 7 (2019) :178167-178176. [2]

[3] Qi-Chao Mao,Hong-Mei Sun,Yan-Bo Liu,Rui-Sheng Jia. “Mini-YOLOv3:Real-Time Object Detector for Embedded Applications”.IEEE Access 7 (2019) :133529-133538. [5]

[4] ZhenchaoOuyang,JianweiNiu,YuLiu,MohsenGuizani. “Deep CNN-Based Real-Time Traffic Light Detector for Self-Driving Vehicles”.IEEE Access 19 (2019):300-313.

[5] Wei Fang,LinWang,Peiming Ren. “Tinier-YOLO:A Real-Time Object Detection Method for Constrained Environments”.IEEE Access 8 (2019) :1935-1944. [6]

[6] VidulaV.Meshram,KailasPatil,VishalA.Meshram,Felix Che Shu. “An Astute Assisstive Device for Mobility and object Recognition for Visually Impaired People”.IEEE Access 49 (2019) :449-460. [3]

[7] WenmingCao,JianheYuan,ZhihaiHe,ZhiZhang,Zhiquan He. “Fast Deep Neural Networks With Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection”.IEEE Access 6 (2018): 8990-8999. [4]

[8] QiankunLiu,BinLiu,YueWu,WeihaiLi,Nenghai Yu. “Real-Time Online Multi-Object Tracking in Compressed Domain”.IEEE Access 7 (2019): 76489-76499.

[9] Meimei Gong,Yiming Shu. “Real-Time Detection and Motion Recognition of Human MovingObjects Based on Deep Learning and Multi-Scale Feature Fusion in Video”.IEEE Access 8 (2020):25811-25822.

[10] Kiruthika, U., & Somasundaram, T. S. (2018). Efficient agent-based negotiation by predicting opponent preferences using AHP. Journal of applied research and technology, 16(1), 22-34.