Object Detection with Voice Feedback

Rajat Lilhare[1], Jitendra Meena[2], Nikhil More[3], Shubhangi Joshi[4]

[1,2,3]*Student Final Year ECE, MIT School Of Engineering, Pune, Maharashtra, India*
[4]*Assistant Professor, Dept. Of Electronics & Communication Engineering MIT SOE, Pune, Maharashtra, India*

---***---

Abstract - *Object Detection is a field of Computer Vision that detects instances of semantic objects in images or videos (by creating bounding boxes around them in our case). In this project, we will convert image to text and then text to speech for the visually impaired person who deserve to live independently by using You Only Look Once V3 (YOLO v3) algorithm that runs through a variation of an extremely complex Convolutional Neural Network architecture called the Darknet with OpenCV and Google Text to Speech, We can then convert the annotated text into audio responses and give the location of the objects in the camera's view. The system will continuously capture multiple frames using a camera on raspberry pi and the frames then converted to audio segment, the obtained results manage to achieve the success of the proposed prototype in giving visually impaired users the capability to understand unfamiliar surroundings, through a user friendly device with this profound object identification model.*

Key Words: *Object detection, YOLO, Deep neural network, Tensorflow, OpenCV, Python, Raspberry Pi3b+, Google Text To Speech.*

1.INTRODUCTION

With the recent rapid development of information technology (IT), a lot of research has been carried out to solve inconveniences ineveryday life, and as a result, various conveniences for people have been provided. Nevertheless, there are still many inconveniences for the visually impaired. The greatest inconveniences that a blind person feels in everyday life include finding information about objects and indoor mobility problems. They have difficulty recognizing simple objects, and it is not easy to distinguish objects thathave similar forms. Previous studies included object analysis using ultrasonic sensors. However, with these methods, it is difficult to know exactly where an object is located, especially in the presence of obstacles. In this paper, we analyze accurate object information and obtain a location using a deep learning object recognition technique. In addition, voice recognition and voice guidance technologies are synthesized so the visually impaired can know the location of the objects they want to find by speaking to the system. Object recognition algorithms are designed based on theSingle Shot MultiBox Detector (SSD) structure, an object recognition deep learning model, to detect objects using a camera. In addition, voice recognition technology designed to use speech-to-text (STT) technology converts a user's vocal commands into text, from which only specific words are extracted and retrieved by the system. In the voice guidance technology, the technique of synthesizing the position of the article so it can be output, and synthesizing the name of the article, is done by using text-to-speech (TTS). In this paper, we propose an efficient object detection system to help find objects in a certain space without help from others, with special consideration for the blind.

1.1 Object Detection

Before Object recognition has developed rapidly, starting with the deep learning–based convolutional neural network (CNN) technique [5] that drew attention at the ImageNet 2012 competition. The CNN, however, was accurate with object classification, but it was difficult to determine where inside the image the object was located. Subsequently, the model for solving this problem was the region-based consolidated neural network (R-CNN), which uses a linear regression method. However, due to the slow speed of theR-CNN, Fast R-CNN was developed. It utilizes a deep learning technique to not only classify the object but also to find the area the object is located in. Nonetheless, there was a limit in that the above model's object recognition processing speed was insufficientfor real-time object recognition. Since then, You Only Look Once (YOLO), which comprises all the processes of object recognitionas a deep learning network, has emerged, and technologies with fast detection speeds, such as Single Shot MultiBox Detector (SSD),have been developed. YOLO estimates the type and location of objects using regression inference on the problem of area selection and classification. On the other hand, SSD does not create candidate areas separately, but recognizes objects using a feature map of various sizes. Since it does not generate candidate areas, it is faster to train than the Faster R-CNN and is more accurate than YOLO because it uses different sizes of feature map.

1.2 Image Processing

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Image processing basically includes the following three steps:

- Importing the image to the system.

- Analyzing and manipulating the image.
- Output in which result can be altered image or report that is based on image analysis.

1.3 Yolo v3

In this project we have used YOLO v3 which is faster than the prior version. It works three times faster, at 320 ×320 YOLOv3 runs 22ms at 28.2 map. It has a similar performance but 3.8× faster. The most notable characteristic of v3 is that it makes 3 distinct scales of detections. YOLO v3 is a fully convolutional neural network and it generates its resultant output by applying a 1 x 1 kernel to a feature map.In YOLO v3, the recognition is obtained by implementing 1 x1 detection kernels to three-size feature maps at three different regions in the network.

Within each boundary the network predicts 4 coordinates tx,ty,tw,th. Whereas if a cell is offset in the upper left corner of the image by (cx,cy) and prior bounding boxes has pw, ph width and height respectively then the prediction is done .

1.4 OpenCV

Techniques for Object Recognition in Images and Multi-Object Detection and segmentation is the most significant and testing central undertaking of Computer vision. It is a basic part in numerous applications, for example, image search, scene understanding, and so far. However it is as yet an open issue because of the assortment and multifaceted nature of item classes and foundations.

The most effortless approach to identify and fragment an item from a picture is the shading based techniques. The tem and the foundation ought to have a critical shading distinction so as to effectively portion objects utilizing shading based strategies.

OpenCV usually captures images and videos in 8-bit, unsigned integer, BGR format. In other words, captured images can be considered as 3 matrices; BLUE, GREEN RED (hence the name BGR) with integer values ranging from 0 to 255.In genuine pictures, these pixels are little to such an extent that the natural eye can't separate.

1.5 Hardware

We are using Raspberry Pi 3 Model B+ which has Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz and 1GB LPDDR2 SDRAM with class 10 micro SD card where the OS and project is stored, the reason behind we are using the class 10 memory card is that it helps to retrieve data at higher speed so that the project can take lesser time to execute The camera used in our system is Raspberry Pi Camera Module v2 which has Sony IMX219 8-megapixel sensor to feed images at 30 frame per second to this trained model,Ultrasonic Distance Sensor HC-SR04.

2. Working

We are using Python3 for this project, the camera is initialized by using OpenCV library and the camera starts capturing frames with the rate of 30 frames per second to the algorithm. Then the system uses YOLO v3 which is trained on the COCO dataset and Dark Neural Network (DNN) to identify the object kept before the user.

The object identified is later converted to an audio segment using gTTs which is a python library. The audio segment is the output of our system that gives the spatial location and name of the object to the person. Now by using this information the person can have a visualization of the objects around him. The proposed system will even protectthe person from colliding to the objects around will secure him from injuries & Ultrasonic sensor will detect object & give the distance between came.

2.1 Block Diagram & Proposed Platform

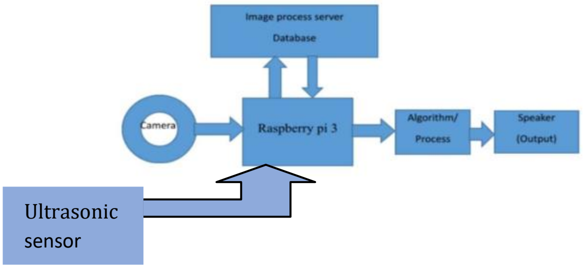


Figure 1: Block Diagram

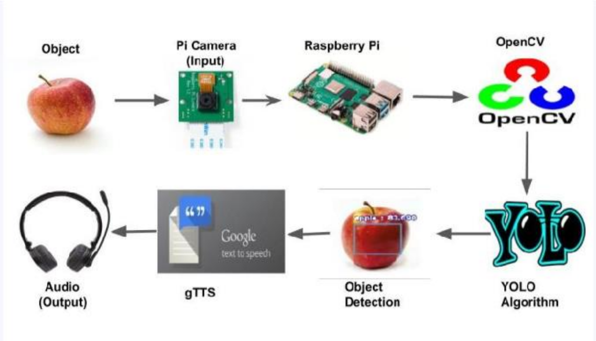


Figure 2: Proposed Platform

3. Result and Experiment

The proposed system will be able to identify the object in front of the camera and will later on convert it into mp3 using gTTS. The proposed system is very low cost, FIG 3 shows the whole system which is a Raspberry pi 3b+, Bluetooth headphones and a power bank in order to provide power to the raspberry pi. The hardware implementation is shown in the given figure.

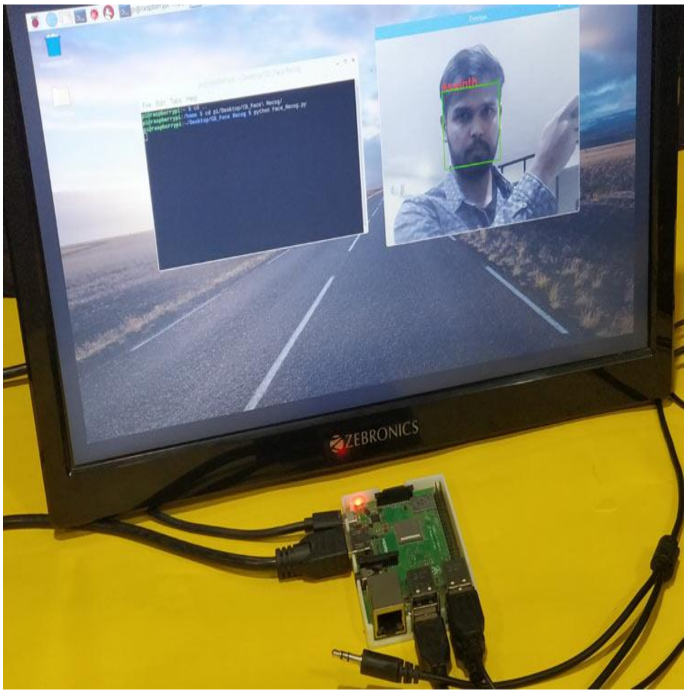
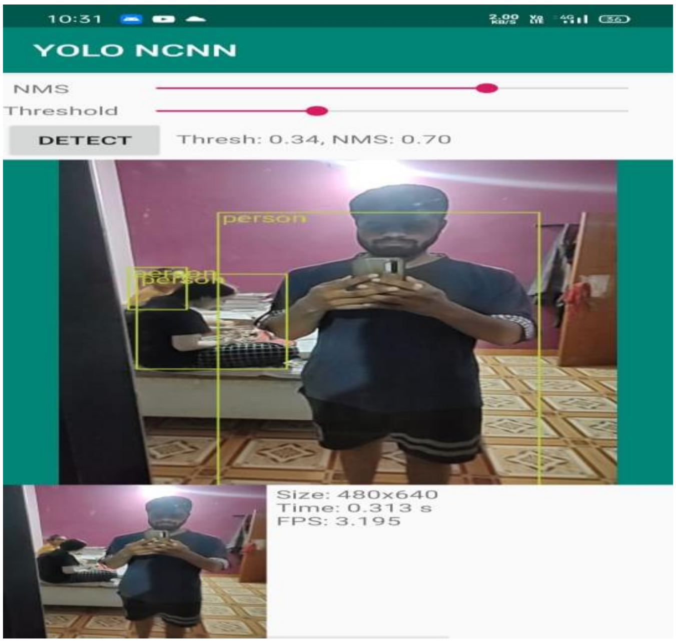


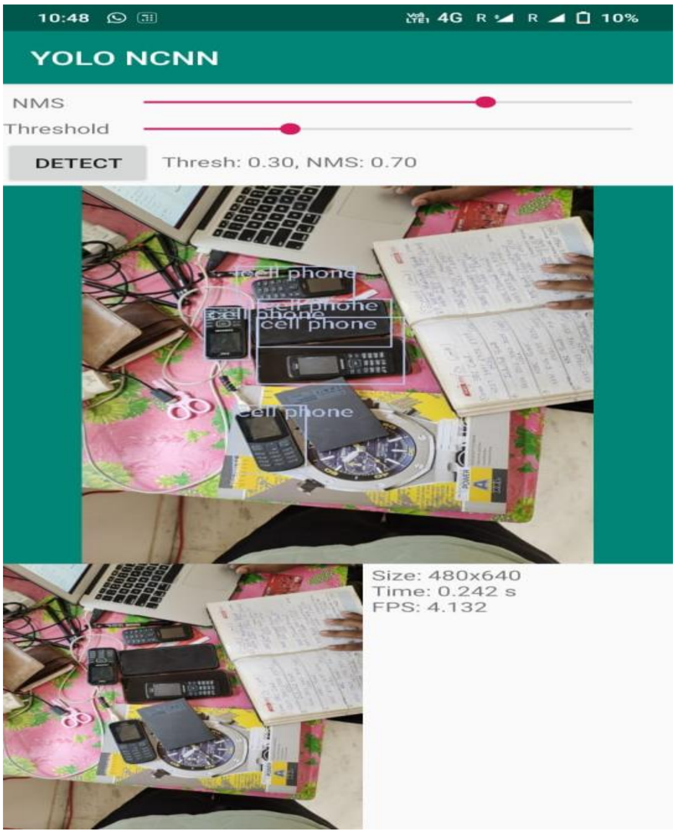
Figure 3.1: Interfacing of Hardware.



Thresh: 0.34, NMS: 0.70

Size: 480x640
Time: 0.313 s
FPS: 3.195

Figure 3.2: Object detection with Python.



Thresh: 0.30, NMS: 0.70

Size: 480x640
Time: 0.242 s
FPS: 4.132

Figure 3.3: Multiple Object detected.

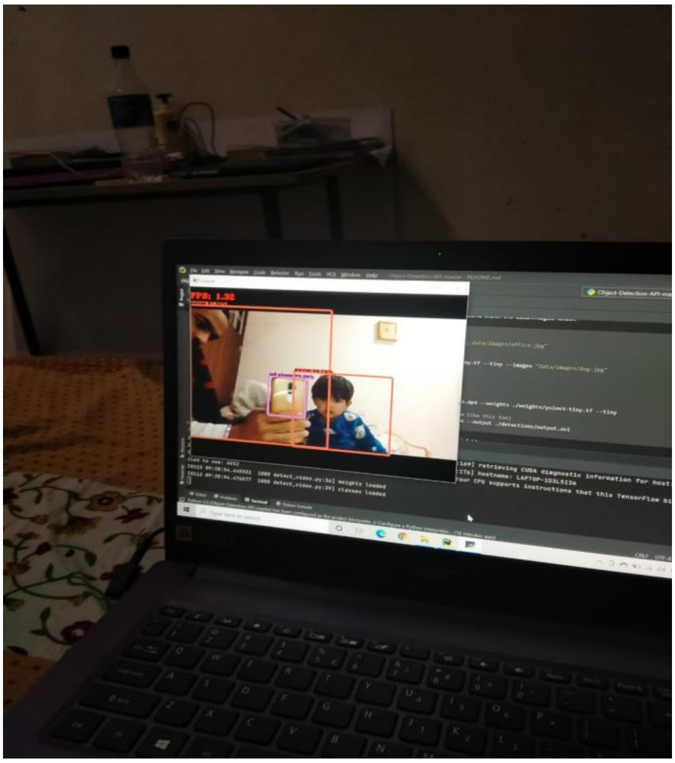


Figure 3.4: Working of the system which is detecting 2 person & 1 cellphone.

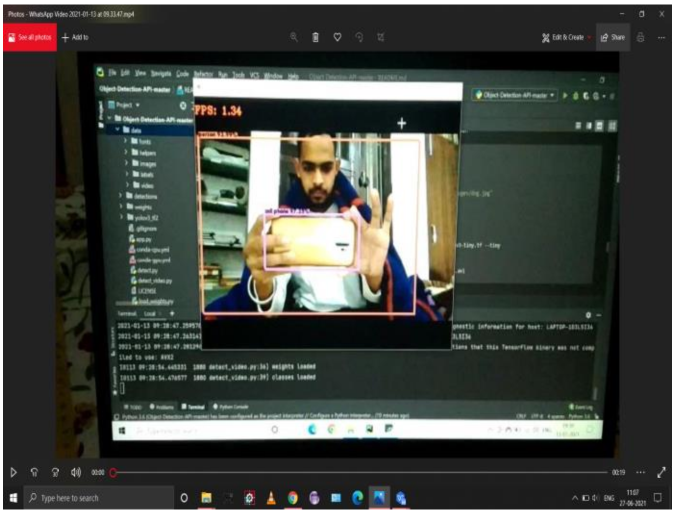


Figure 3.5: Working of the system which is detecting person & cellphone.

FUTURE SCOPE

This project is for the blind people who are incapable to see this colorful and beautiful world, our initiative will support them to have a better life. By this project one will be able to understand what object is present in front of him and by continuous research and development our team will be able improve this product by feeding more data to the Deep Learning algorithm by which the accuracy of the model will increase as well as the power of the algorithm to recognize more objects will increase. The object recognition system can be applied in the area of surveillance system, face recognition, fault detection, character recognition etc. The objective of this thesis is to develop an object recognition system to recognize the 2D and 3D objects in the image. The performance of the object recognition system depends on the features used and the classifier employed for recognition. This research work attempts to propose a novel feature extraction method for extracting global features and obtaining local features from the region of interest.

CONCLUSIONS

This study can be used widely to provide the blind with privacy and convenience in everyday life. Also, it is expected to be applied to industrial areas where diminished visibility occurs, such as coal mines and sea beds, to greatly help production and industrial development in extreme environments. The study aims to enable people with visual impairment to live more independently. Peoplewith visual impairment will be able to overcome some threats that they may come across in their day to day life that may be either while reading a book or traveling through the city by making efficient use of the application and its associative voice feedback. Therefore, it will help to prevent possible accidents. The mobile devices can be carried easily and the camera of the device can be used to detect object from the surroundings and give output in audio format. Thus, helping visually impaired people to 'See throughthe Ears'.

REFERENCES

[1] Kedar Potdar, Chinmay Pai and Sukrut Akolkar, ”A Convolutional Neural Network based Live Object Recognition System as Blind Aid”, arXiv:1811.10399v1 [cs.CV] 26 Nov 2018 https://arxiv.org/pdf/1811.10399.pdf

[2] Liam Betsworth, Nitendra Rajput, Saurabh Srivastava,and Matt Jones. Audvert: Using spatial audio to gain a sense of place. InHuman-Computer Interaction–INTERACT 2013, pages 455–462. Springer, 2013

[3] Evanitsky, Eugene. ”Portable blind aid device.” U.S. Patent No. 8,606,316, 10 Dec. 2013.

[4] A.Culjak, D.Abram, T. Pribanic, H. Dzapo and M. Cifrek, A brief introduction to OpenCV,"2012 Proceedings of the 35th International Convention MIPRO, Opatija, 2012, pp. 1725-1730.

[5] Rahul Kumar and Sukadev Meher, "Assistive System for Visually Impaired using Object Recognition, M.Sc. Thesis at Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, Rourkela, Odisha-769 008, India, May 2015.

[6] Khushboo Khurana, Reetu Awasthi, 2013, "Techniques for Object Recognition in Images and Multi-Object Detection," International Journal of Advanced Research in Computer Engineering & Technology.

[7] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", University of Washington, Allen Institute for AI , Facebook AI Research, 2016.

[8] J. Redmon and A. Farha, Yolov3: An incremental improvement. arXiv, 2018.

[9] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pages 6517–6525. IEEE, 2017.

BIOGRAPHIES

Rajat Lilhare
B.Tech Student at MIT School of Enigeering, Pune, Maharashtra, India

Jitendra Meena
B.Tech Student at MIT School of Enigeering, Pune, Maharashtra, India

Nikhil More
B.Tech Student at MIT School of Enigeering, Pune, Maharashtra, India

Shubhangi Joshi
Assistant Professor, Dept.of Electronics & Communication Engineering at MIT School of Enigeering, Pune, Maharashtra, India