iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35881

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Image to Speech Conversion using CV to Help Visually Challenged People

Prof. S.G. Latake[1], Sourabh Shinde[2], Sneha Shinde[3], Abhijeet Pawar[4], Ajinkya Thombare[5], Pratik Sonawane[6]
[1]*Professor, Dept. of Information Technology Engineering, Jayawantrao Sawant College of Engineering,Maharashtra, India.*
[2, 3, 4, 5, 6]*Student, Dept. of Information Technology Engineering, Jayawantrao Sawant College of Engineering, Maharashtra, India*

Abstract: ***This work aims to assist the visually impaired people for reading a text material and detect objects in their surroundings. The input is taken in the form of an image captured from the web camera. This image is then processed either for the purpose of text reading or for object detection based on user choice. The main aim of this project is to build a system that detects objects from the image or a stream of images given to the system in the form of previously recorded video or the real time input from the camera. Bounding boxes will be drawn around the objects that are being detected by the System. The system will also classify the object to the classes the object belongs. Python programming and a machine Learning technique named yolo (you only look once) algorithm using convolutional neural network is used for the object detection. The smart blind navigation is fill gap, providing accurate and contextually rich information about the environment around the user current location, and simplifying the navigation and increasing the overall accuracy of the System. Preventing the user from dangerous locations. They have very little information on self-velocity objects, direction which is essential for travel. The navigation systems is costly which is not affordable by the common blind people. The navigation system are heavy complicated to operate.***
Keywords: Machine Learning, Yolo, Convolutional Neural Network, Bounding Box, Object Detection

I. INTRODUCTION

The technology for navigation of the blind is not sufficiently accessible devices rely heavily on infrastructure requirements. Without vision it can be challenging for visually im-paired persons to navigate through rooms or different road paths .The main aim to develop the project is to help the visually impaired people and to detect the obstacles to detect the road traffic signs. The blind persons life become easier and they can go anywhere where they wants without anyone helps .They can walk alone through street they does not need anyone to assist them they can handle their self correctly. The project comes under the domain Machine Learning which is the part of Artificial Neural Network. Machine Learning concepts makes the system learn on its own from the experiences it gains, without the interference of the external factors. The YOLO (You Only Look Once) algorithm using Convolutional Neural Network is used for the detection purpose. It is a Deep Neural Network concept from Artificial Neural Network. Artificial Neural Network is inspired by the biological concept of Nervous System where the neurons are the nodes that form the network. Similarly, in Artificial Neural Network perceptions act like the nodes in the network. Artificial Neural Network has three layers that are, Input Layer, Hidden Layer and the output Layer. Deep

Learning is the part of the Artificial Neural Network that has multiple Hidden Layer that can be used for the Feature Extraction and Classification purposes. Convolutional Neural Network (CNN) is the part of Deep Learning that is used in analysis of visual imagery. It has four different kinds of layers, they are, Convolutional Layer,Pooling Layer, Activation Layer and Fully Connected Layer. Convolution Layer uses filter and strides to obtain the Feature Maps. These Feature Maps are the matrix that is obtained after the Convolution Layer. It can be simplified using ReLU (Rectified Linear Unit) that maps negative values

to 0. The resulted Feature Map is reduced by sending it into the Pooling Layer where it is reduced to the smaller sized matrix. This is how the features are extracted. At the end of the convolutional neural network is the Fully Connected Layer where the actual Classification occurs.

II. LITERATURE REVIEW

Zibo Gong1, Tianchang He1, and Ziyi Yang1 They work on object detection using neural network and other computer vision features. We use Faster Region based Convolutional Neural Network method (Faster R-CNN) for detection and then match the object with features from both neural network and features like histograms of gradients (HoG). We are able to achieve real-time performance and satisfactory matching results.. We firstly used a region proposal network (RPN) to generate detection proposals. Then we employed the same network structure to Fast R-CNN to classify the object and modified the bounding box. Furthermore we extracted feature from object detected via algorithm developed by us. At last we matched object detected with ones stored in database.

Roshan Rajwani [1], Dinesh Purswani [2], Paresh Kalinani [3] Deesha Ramchandani [4], Indu Dokare [5]
They experimentally and numerically solve the problem of a visually impaired person needs absolution to help him overcome problems in navigation due to his disability. The project is mainly focused on providing a type of visual aid to the visually impaired people. With the current advances in comprehensive innovation, it is conceivable to stretch out the help given to individuals with visual hindrance during their mobility. In this context we propose system in which an Android smartphone is used to help a blind user in obstacle detection and navigation. Today, smartphones are available to anyone. In fact, they have become the most common device available everywhere. Hence, this project uses an Android smartphone that uses its camera to identify objects in surroundings and gives an audio output. The hearing ability of the user tries to fulfil his seeing ability. In this paper a model has been proposed which makes the use of smartphone, a common device available to anyone and used technology to make an application which can help the blind user detect objects in his surroundings and help him in navigating from one place to another. The output of the system is in audio form that can be easily understandable for a blind user.

M. Murali, Shreya Sharma and Neel Nagansure This work aims to assist the visually impaired people for reading a text material and detect objects in their surroundings. The input is taken in the form of an image captured from the web camera. This image is then processed either for the purpose of text reading or for object detection based on user choice. The Raspberry Pi acts as the microcontroller for processing of the entire process. The text reading is supported by software named OCR. The read text is changed into an audio output using the TTS Synthesis. Other dependencies required for the process include Tesseract Library. The Object Detection is another aspect of the project which is implemented using a TensorFlow Object Detection API. It is able to detect various objects in its surroundings and provide an audio feedback about the same. The dataset can be trained on various different situations depending on the user needs, thus making it scalable. With the aim to provide assistance to disabled people we decided to innovate something for the visually disabled or blind people of our society

Piyush Pimplikar Avanti Dorle Pranit Bagmar Atharva Rajurkar This project proposes an android application to help blind people see through handheld device like mobile phone. It integrates various techniques to build a rich android application that will not only recognize objects around visually impaired people in real time but also give an audio output to assist them as quickly as possible. SSD (Single Shot Detector) Algorithm is used for the object recognition as well as detection. Also this algorithm gives nearly accurate results for real time object detection and is proven to be faster than other relative algorithms. The application further uses android tensorflow APIs and android TextToSpeech API to give audio output. This application aims to help the visually impaired people to know their surrounding objects that could be just basic everyday objects or can create obstacle in their activity. The application is build to recognise or detect some household objects like chair, table, bed, refrigerator, laptops etc and some on outdoor objects like cars, motorbikes, potted plant, people etc. The application will use mobile phone camera to scan the surrounding in real time and take the frames from the ongoing video. The frames will be sent to the next module where the SSD algorithm will create bounding boxes around the objects in the frame and classify them into given categories. At last the application will produce an audio output of the
object detected which has the maximum confidence score among all other present in the frame. The frames are selected at a particular time interval to avoid the hindrance in the audio output.

Pooja Maid1, Omkar Thorat2, Sarita Deshpande3 The Smart Blind Navigation is fill gap, providing
accurate and contextually rich information about the environment around the user current location, and simplifying the navigation and increasing the overall accuracy of the system. Preventing the user from dangerous locations. They have very little information on self-velocity objects, direction which is essential for travel. The navigation systems is costly which is not affordable by the common blind people. The navigation system are heavy complicated to operate. In this project we present a visual system for blind people based on object like images and video scene. This system uses Deep Learning for object identification. In order to detect some objects with different conditions. Object detection deals with detecting objects of inside a certain image or video. The TensorFlow Object Detection API easily
create or use an object detection model Blind peoples they have a very little information on self-velocity objects, direction which is essential for travel. The navigation systems is costly which is not affordable by the common blind people. So this project main aim is to the help of blind people.

III. PROPOSED SYSTEM

The Fig-1 shows the Architecture Diagram of the Proposed YOLO Model. Images are given as the input to the system. If Video can also be taken as input as it is nothing but a stream of images. As the name suggests You Only Look Once, the input goes through the network only once and the result of detected object with Bounding Boxes and Labels are obtained. The images are divided into SXS grid cells before sending to the Convolutional Neural Network (CNN).

B Bounding boxes per grid are generated around all the detected objects in the image as the result of the Convolutional Neural Network. On the other hand, the Classes to which the objects belong is also classified by the Convolutional Neural Network, giving C Classes per grid. Then a threshold is set to the Object Detection. In this project we have given a Threshold of 0.3. Lesser the Threshold value, more number of bounding boxes will appear in the output resulting in the clumsy output.

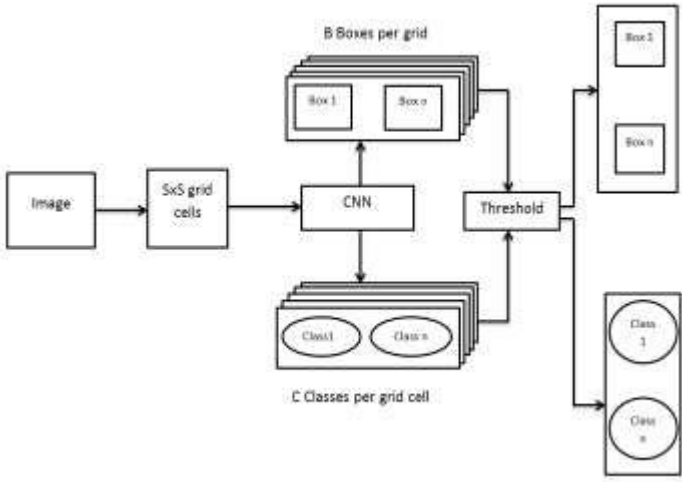


Fig -1: YOLO Architecture

The Fig-2 illustrates the Flow of data in the System. Initially User will be given the options to choose the type of the File to be given to the System as an input. Thus, User can either choose option of File Selection or start the Camera. In the former, User can choose either Image File or a Video File and, in the latter, User can start the Camera module. Once the input is selected Preprocessing is done, where the SXS grids
are formed. The resultant thus formed with the grids is send to the Bounding Box Prediction process where the Bounding Boxes are drawn around the detected objects. Next the result from the previous process is sent to the Class Prediction where the Class of the object to which it belongs is predicted.

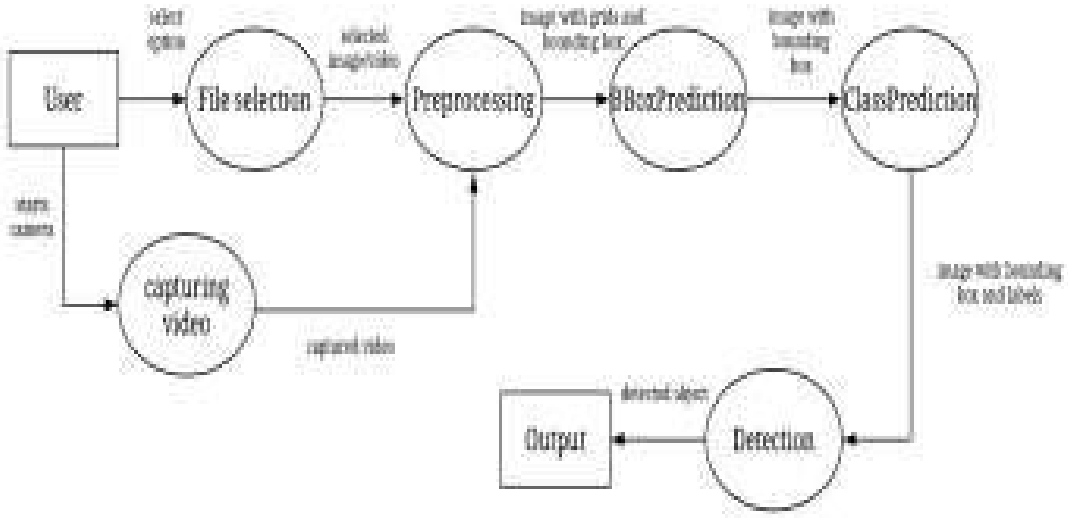


Fig -2: Data Flow Diagram of the System

Then it is sent to the detection process where a Threshold is set in order to reduce clumsiness in the output with many Bounding Boxes and Labels in the final Output. At the end an image or a stream of images are generated for image and video or camera input respectively with Bounding Boxes and Labels are obtained as the Output.

IV. IMPLEMENTATION

This chapter describes the methodology for implementing this project. Following is the algorithm for detecting the object in the Object Detection System.

A. Algorithm for Object Detection System

1) The input image is divided into SxS grid
2) For each cell it predicts B bounding boxes Each bounding box contains five elements: (x, y, w, h) and a box confidence score
3) YOLO detects one object per grid cell only regardless of the number bounding boxes
4) It predicts C conditional class probabilities
5) If no objects exists then confidence score is zero Else confidence score should be greater or equal to threshold value
6) YOLO then draws bounding box around the detected objects and predicts the class to which the object belongs

A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech.

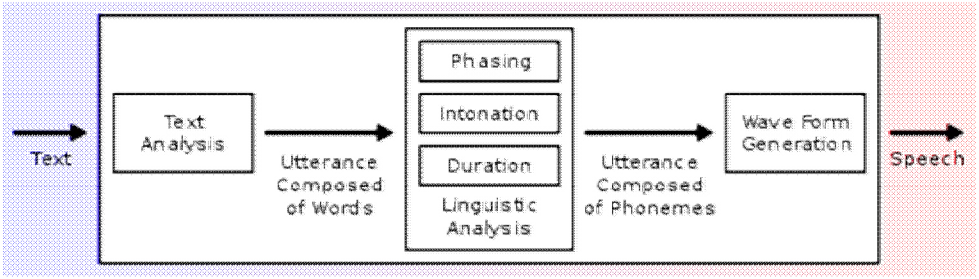


Fig.3Overview of a typical TTS system

A text-to-speech system (or "engine") is composed of two parts: a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end—often referred to as the synthesizer—then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody (pitch contour, phoneme durations), which is then imposed on the output speech.

V. RESULT

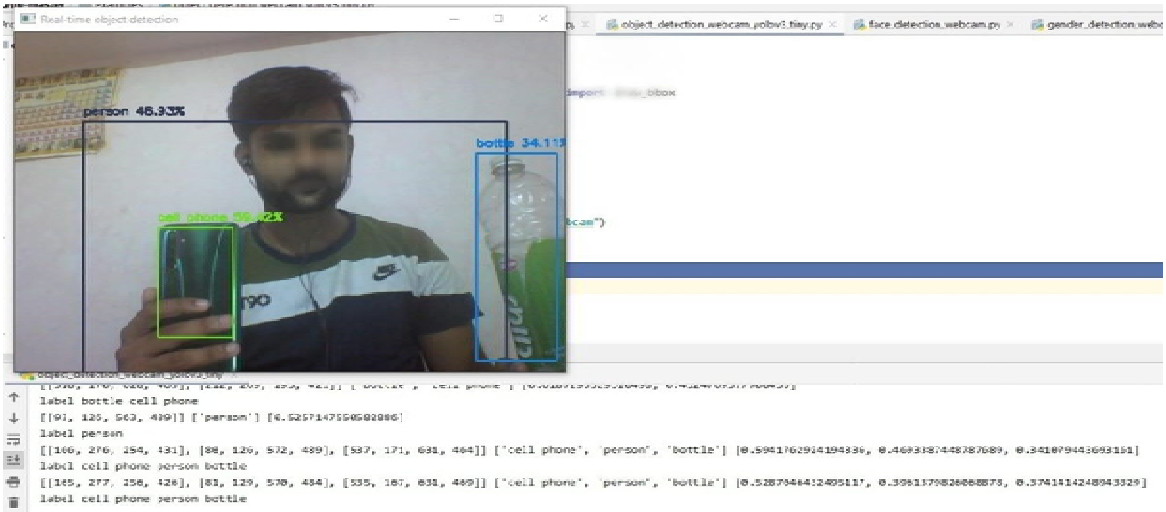


Fig.4 Screenshot of result

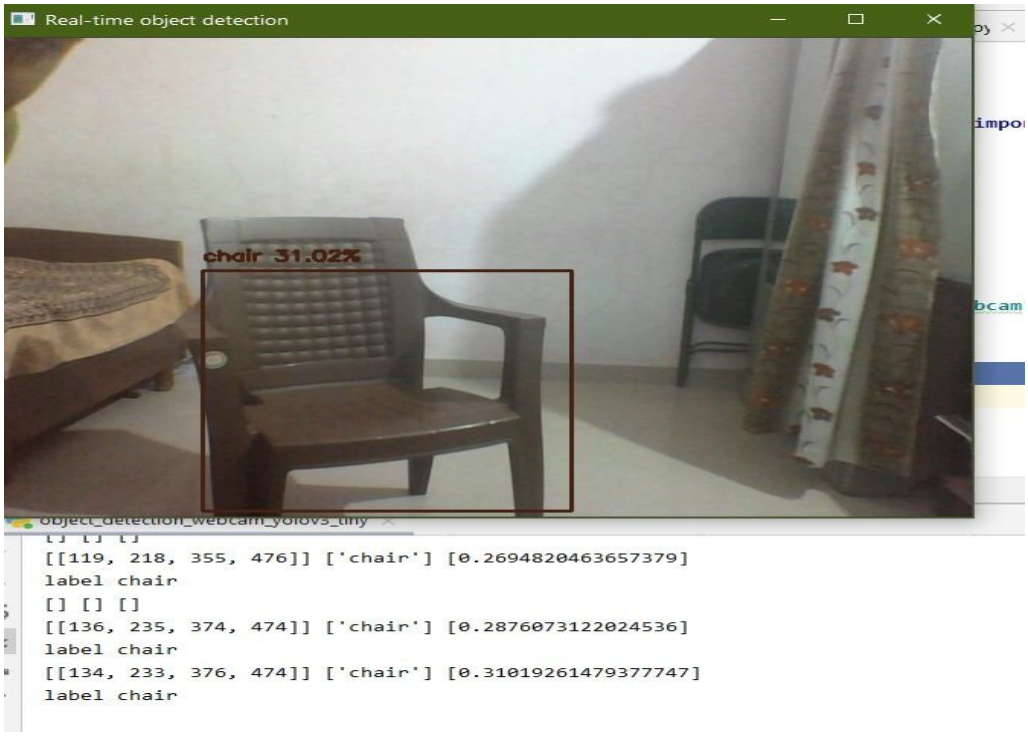


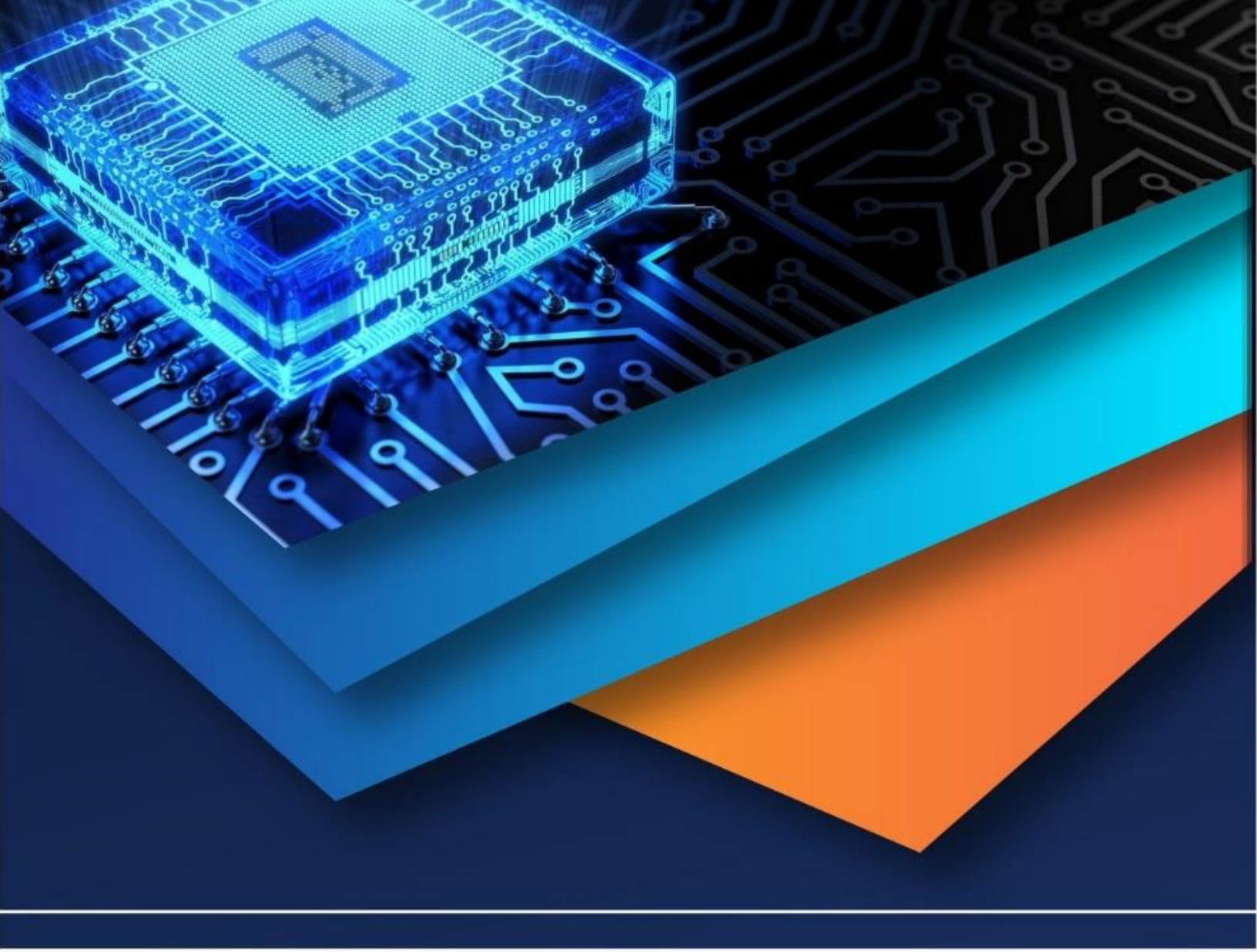
Fig 5. Screenshot of result

VI. CONCLUSIONS

The project is developed with objective of detecting real time objects in image, video and camera. Bounding Boxes are drawn around the detected objects along with the label indicating the class to which the object belongs. We have used CPU for the processing in the project. Future enhancements can be focused by implementing the project on the system having GPU for faster results and better accuracy

REFERENCES

[1] Juan and O. Gwon, aˆA Comparison of SIFT, PCASIFT and SURFˆa. International Journal of Image Processing(IJIP), 3(4):143 aˆ 152, 2009.
[2] Hanen Jabnoun, Faouzi Benzarti ,Hamid Amiri, ˆa Visual substitution system for blind people based on SIFT descriptionˆ, International Conference of Soft Computing and Pattern Recognition 2014 IEEE.
[3] Hanen Jabnoun, Faouzi Benzarti,and Hamid Amiri, ˆa Object recognition for blind people based on features extraction IEEE IPASˆa14: INTERNATIONAL IMAGE PROCESSING APPLICATIONSAND SYSTEMS CONFERENCE 2014.
[4] Raspberry pi user guide;Eben Upton and Gareth Halfacree; 312 pages;2017;ISBN 9781118921661
[5] Esteban Bayro Kaiser, Michael Lawo,aˆWearable Navigation System for the Visually Impaired and Blind Peopleˆ IEEE, 2012.
[6] Widodo Budiharto, Alexander A S Gunawan, Jarot S.Suroso, Andry Chowanda, Aurello Patrik and Gaudi Utama, “Fast Object Detection for Quadcopter Drone using Deep Learning”, 2018 3rd International Conference on Computer and Communication System(ICCCS), Nagoya, Japan, ISBN (e): 978-1-5386-6349-3, April- 2017.
[7] Pushkar Shukla, Beena Rautela and Ankush Mittal, “A Computer Vision Framework for Automatic Description of Indian Monuments”, 2017 13th International Conference on Signal Image Technology and Internet-Based Systems(SITIS), Jaipur, India, ISBN (e): 978-1-5386-4283-2, December-201.
[8] M. Buric, M. Pobar and Ivasic-Kos, “Object Detection in Sports Videos”, 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics(MIPRO), Opatija, Croatia, ISBN (e): 978-953-233-095-3, May-2018.
[9] Yin-Lon Lin, Yu-Min Chiang and Hsiang-Chen Hsu, “Capacitor Detection in PCB Using YOLO Algorithm”, 2018 International Conference on System Science and Engineering(ICSSE), New Taipei City, Taiwan, ISBN (e): 978-1-5386-6285-4, December-2017
[10] Romain Vial, Hongyuan Zhu, Yonghong Tian and ShijianLu “Search Video Action Proposal with Recurrent and Static YOLO”,2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, ISBN (e): 978-1- 5090-2175-8, Sep



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY