iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34538

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 V May 2021- Available at www.ijraset.com

Real Time Object Detection Using YOLO v3 Tiny with Voice Feedback for Visually Impaired

Sangeeta Mahapatra[1], Neha Chugwani[2], Pooja Chuathmal[3], Arti Tekade[4]
[1, 2, 3, 4]*Department of Electronics and Telecommunication Engineering, Pimpri Chinchwad College of Engineering and Research, Ravet-412101*

Abstract: Object Detection is a technique of Computer Vision in which a software system can detect, situate and trace the object from a given image or video. The object's location is identified by drawing a bounding box around the object which is to be identified. In the proposed project we will use this object detection methodology to assist the visually impaired to locate and identify things in their surroundings. This is achieved by using the You Only Look Once (YOLO v3 Tiny) algorithm which is trained on COCO Dataset, TensorFlow, and Open CV. Then the labeled text of the identified object is converted to audio with the help of Google Text TO Speech. Multiple frames are captured by the camera on raspberry pi continuously and then they are translated into voice output. The results thus obtained will help the visually impaired to perceive their environment.
Keywords: Object detection, Open CV, YOLO, Raspberry Pi, Google Text To Speech

I. INTRODUCTION

There are a number of individuals who live in this world with the insufficiency of understanding nature due to impaired vision Because of which they face difficulties in leading their day-to-day life and social awkwardness. For instance, they find difficult to locate things in new surroundings. In our planet with the human population of 7.4 billion, 258 million are visually inefficient. Out of them 39 million are blind completely i.e., they are totally visionless and 246 million suffering from mild or severe impairment.

In recent years there have been considerable breakthroughs in the field of Object detection due to which it has attracted an increasing amount of attention [1]. Identification of objects is associated with computer vision and image processing. This deals with the detection of occurrences of objects of explicit class existing in digital videos and images. This technology thus will give an opportunity to the visually impaired to see this world. There are many computer vision technologies developed that are precise and give accurate results [2]. Primarily, there are three object detectors namely: Shot Detector (SSD), R-CNN, Fast R- CNN, and Faster R-CNN, Single YOLO.

This project uses the sense of hearing to visualize the objects in the surrounding using the "You Only Look Once: Unified, Real-Time Object Detection" algorithm [3] trained on the COCO [4] dataset to identify the object present before the person thereafter the label of the detected object is translated to audio by the aid of Google Text To Speech (gTTS) which will be the expected output. This output will assist the user to perceive the environment.

II. PROJECT REQUIREMENT

A. Raspberry Pi

Raspberry pi is a lightweight, low cost and low-energy computer. It is the size of a credit card. It has 900-Megahertz quad processor and one gigabyte of RAM. People use Raspberry pi for home automation tasks like internet-enabled garage doors or doorbells. It provides greater networking as it comes with gigabit ethernet and onboard wireless networking along with better Bluetooth connectivity. The official operating system of raspberry pi is Raspbian. It is a version of Linux built especially for the Raspberry pi.

B. COCO Dataset

Common Objects in context are segmentation, large-scale object detection, and captioning dataset. COCO dataset contains a large number of object images captured through different angles to identify the image.

Object Detection has

1) 121,408 images are present in the COCO dataset.
2) 883,331 object annotations are available in COCO Dataset.
3) 80 classes are present in COCO Dataset.
4) The median image ratio in COCO Dataset is 640x 480.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

C. OpenCV

It is an open-source Library that contains various real-time programming functions in the field of Computer vision. This makes OpenCV [6] very important for real-time operations. It is easy to use and cross-platform. It has interfaces of Python, MATLAB, C++, and Java.

D. YOLO v3

YOLOv3(You Only Look Once) is the third version of the algorithm used for object detection. This version is faster than the previous version of yolo available. Yolov3 [5] helps us to nearly detect 80 different objects in videos and images. Its accuracy is higher than its previous object detection algorithms. It is a combination object locator (knows the location of the object) as well as an object detector. YOLOv3 runs 22ms at 28.2 maps at 320x320. It makes 3 unique scales of detection. The Deep Learning algorithm is used by yolov3 for object detection.

E. Google-Text-to-Speech (gTTs)

It is a library in Python and a Command Line interface. It writes the spoken data which is in mp3 format into a file or any similar kind of object. This result can be further processed for various audio manipulation or simply outputted as text. gTTs can read an unrestricted length of text.

III. WORKING

Python 3 is being used in this project. Using the Open CV library, the pi camera is initialized. The camera captures the frame with the speed of 30 frames per second. The system makes the use of Yolo v3 Tiny which is trained using the COCO dataset. The object so identified by the system is then converted to audio with the assistance of Google Text To Speech (gTTs) which is also a python library. This audio output gives the identification of the object to the user. With the aid of this the user is able to identify the objects in the surroundings. This will in turn help the person to avoid injuries and be able to perform his day-to-day activities without much trouble.

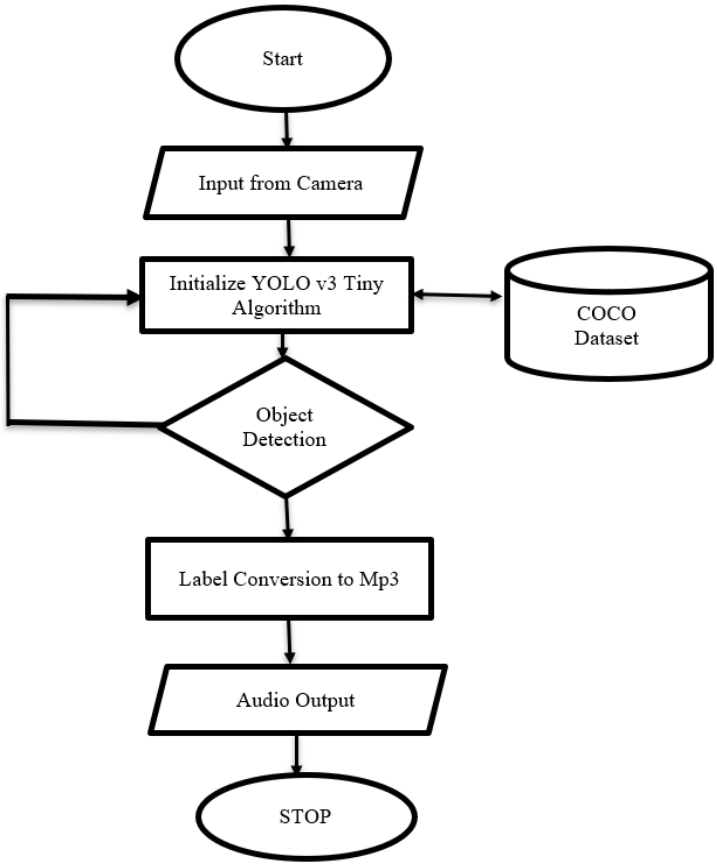


Fig. 1 Flow Chart

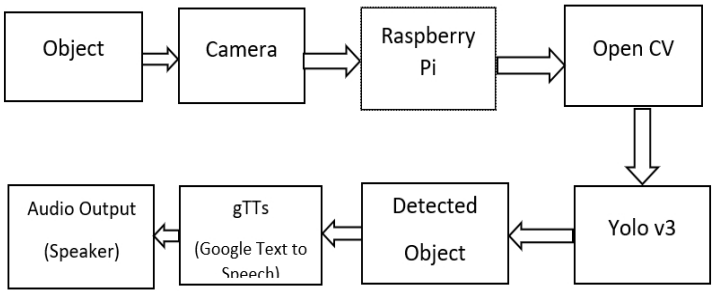


Fig. 1 Block Diagram

IV. EXPERIMENTAL RESULTS

Below are the results that were obtained when the software part of the project was being tested. Fig1 shows the results when we used an image as an input to the software and Fig2 shows the results obtained through live web cam. In both the tests, the objects present in the image and the objects in front of the web cam were detected successfully.

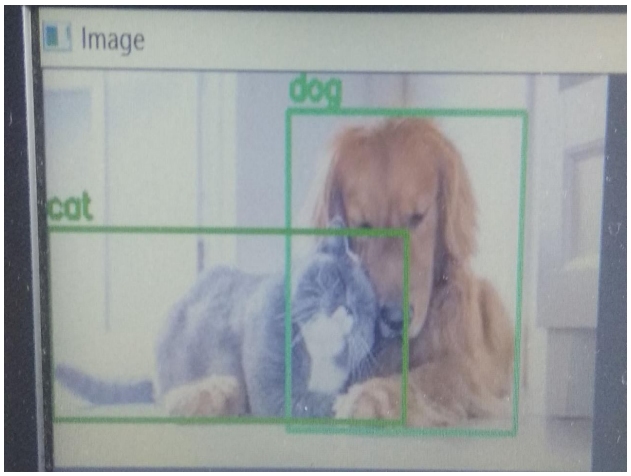


Fig. 3 Object Detection using image as an input

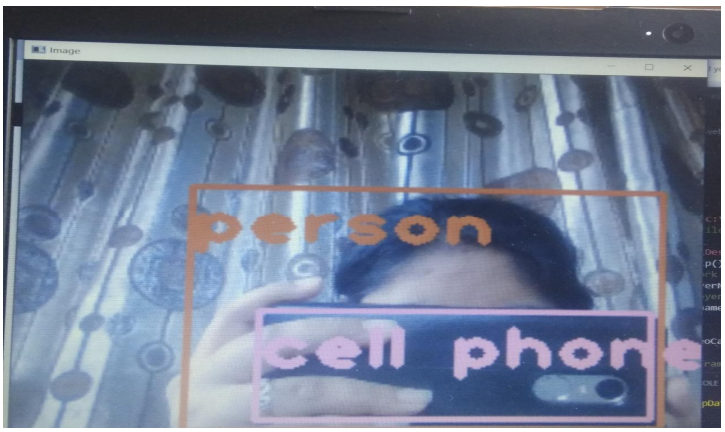


Fig. 4 Object Detection using Web cam

V. CONCLUSION

The project will accurately detect the surrounding objects using YOLO Algorithm and then convert into voice feedback using gTTs. This project made with the help of Deep Leaning and Raspberry pi will greatly help visually impaired individuals to the great extent by acting as a tool connecting them to the world and surpassing their disability of vision.

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.

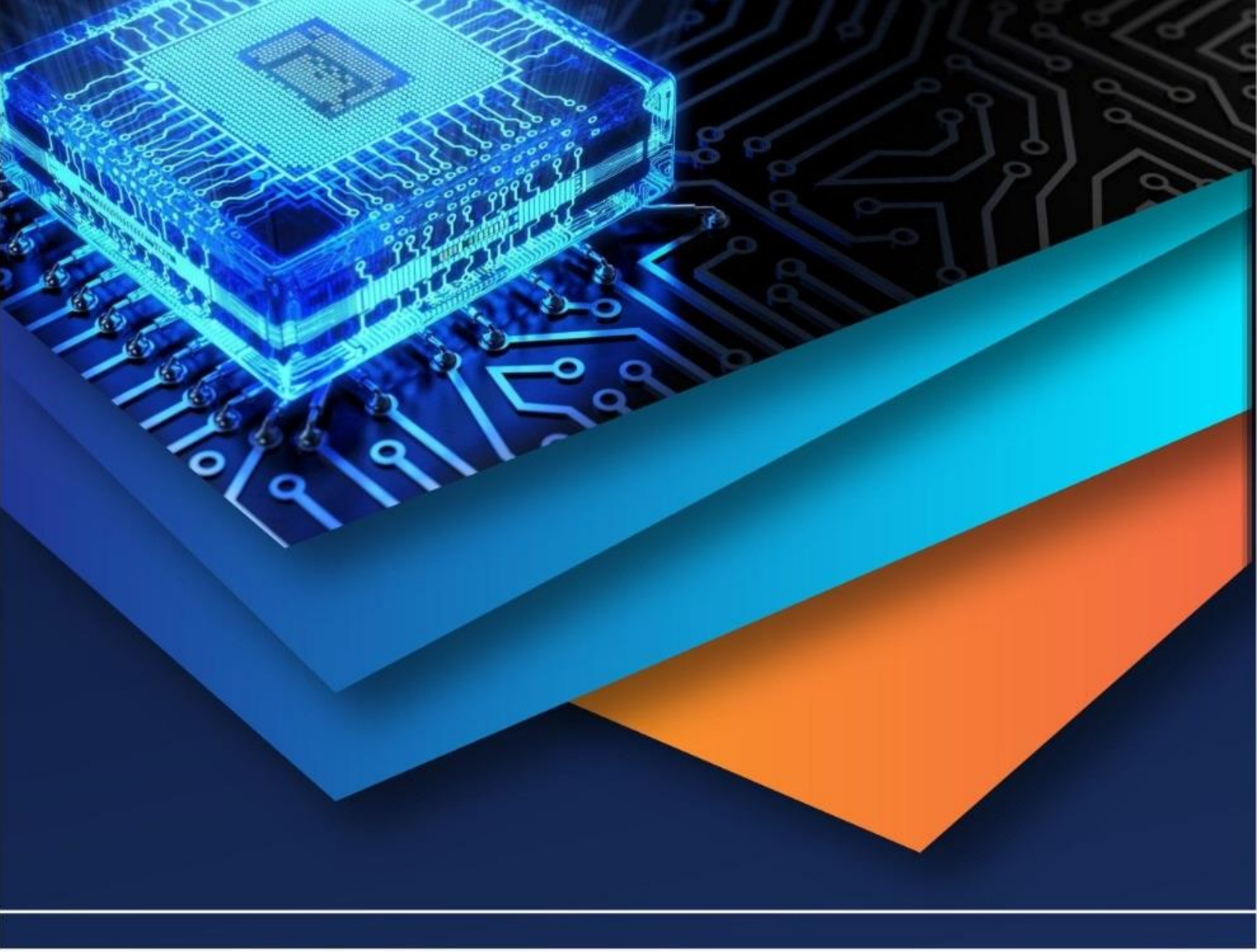
[3] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, University of Washington, Allen Institute for AI , Facebook AI Research, 2016.

[4] 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, Odisha769 008, India, May 2015.

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

[6] 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.

[7] 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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY