Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Blind Assistance System using Image Processing

By P. Rama Devi, K. Sahaja, S. Santrupth, M. P. Tony Harsha
& K. Balasubramanyam Reddy

Gitam Institute of Technology

Abstract- Eye diseases usually cause blindness and visual impairment. As per the statistics, there are over 285 million visually impaired people living worldwide. They come across many troubles in their daily life, especially while navigating from one place to another on their own. They often depend on others for help to satisfy their day-to-day needs. So, it is quite a challenging task to implement a technological solution to assist them. Several technologies were developed for the assistance of visually impaired people. One such attempt is that we would wish to make an Integrated Machine Learning System that allows the blind victims to identify and classify real-time objects generating voice feedback and distance. Which also produces warnings whether they are very close or far away from the thing.

Keywords: *blindness, visual impairment, machine learning, real-time objects.*

GJCST-D Classification: *DDC Code: 617.7 LCC Code: RE52*

Strictly as per the compliance and regulations of:

Blind Assistance System using Image Processing

P. Rama Devi [α], K. Sahaja [σ], S. Santrupth [ρ], M. P. Tony Harsha [ω] & K. Balasubramanyam Reddy [¥]

Abstract- **Eye diseases usually cause blindness and visual impairment. As per the statistics, there are over 285 million visually impaired people living worldwide. They come across many troubles in their daily life, especially while navigating from one place to another on their own. They often depend on others for help to satisfy their day-to-day needs. So, it is quite a challenging task to implement a technological solution to assist them. Several technologies were developed for the assistance of visually impaired people. One such attempt is that we would wish to make an Integrated Machine Learning System that allows the blind victims to identify and classify real-time objects generating voice feedback and distance. Which also produces warnings whether they are very close or far away from the thing.**

Keywords: blindness, visual impairment, machine learning, real-time objects.

I. Introduction

The fast progress of data and organized technology has advanced from the Internet to applying innovations in life. One of the technologies to consider is objected acknowledgment innovation, later known as object detection. This term denotes a capacity to identify the shape size of diverse objects, and the device's camera catches their position. The practice of detecting real-world object instances in still photos or videos, such as a car, bike, TeleVision, flowers, and humans, is known as object detection. It lets us recognize, localize, and detect many things inside an image, giving us a better overall knowledge of the picture. Image retrieval, security, surveillance, and sophisticated driver assistance systems are all examples of areas where it's applied (ADAS).

Developing accurate Machine Learning Models capable of identifying and localizing multiple objects in a single image has long been a significant challenge in computer vision. However, thanks to recent advances in Deep Learning, developing Object Detection applications is now easier than ever. TensorFlow's Object Detection API is an open-source framework built on top of TensorFlow that makes building, training, and deploying object detection models simple.

Detection of objects can be accomplished in a variety of ways. It is a known fact that the statistical number of visually impaired individuals in the world is nearly 285 million. They face a lot of trouble and constant challenges in Navigation, especially when they are on their own. They need to often depend on someone to get their fundamental daily needs met. So, it is a very challenging task to make a mechanical arrangement for them which is most significant. One such attempt from our project is that we would like to develop an Integrated Machine Learning Framework that permits visually challenged people to distinguish and classify everyday day-to-day objects with voice assistance calculating distance and producing warnings whether the person is close or distant from the thing. The same framework can be used for obstacle detection instruments.

We'll concentrate on Deep Learning Object Detection in this Object Detection project because TensorFlow is based on Deep Learning. Each Object Detection Algorithm works somewhat differently, but they all follow the same basic principles.

Feature Extraction: They use their hands to extract features from input images and utilize these features to identify the image's class. MATLAB, OpenCV, Viola-Jones, and Deep Learning are just a few examples. Tensors are multidimensional arrays that extend the functionality of two-dimensional tables to data with a higher dimension. TensorFlow has numerous properties that make it suitable for Deep Learning. So, without spending any time, let's look at how we can use TensorFlow to develop Object Detection.

COCO dataset comprises around 330K annotated images for Common Objects in Context. Now you must choose a model because you must make a crucial trade-off between speed and accuracy. The main motto for object detection is to find things, drawing rectangular bounding box-like structures around them with distance. Object detection applications are emerging in numerous diverse areas counting, recognizing people, checking agricultural crops, and real-time applications in sports.

Many methods and techniques are introduced to solve the problems of visually impaired people.

This paper gives a compelling presentation on object detection and analyzing the gesture of an object using computer vision and machine learning.

This paper proposed a well-known computer technology part of image processing and computer

Author α: Assistant Professor.
Author σ ρ ω ¥: Students, Department of Computer Science and Engineering, Gitam Institute of Technology, Visakhapatnam, Andhra Pradesh, India.
e-mail: rponnaga@gitam.edu

vision that focuses on detecting objects in computerized pictures or videos. There are various object detection applications with high requirements for face detection, vehicle calculator, and character recognition. Object detection can be used for different applications, including recovery and surveillance. Other essential concepts used in object detection, like using the OpenCV library of python 2.7 progressing in the exactness and effectiveness of object detection, are displayed.

This paper described that everyone wants to live independently, especially the disabled ones. Over the past few decades, technology has helped disabled ones control their livelihood. In this study, an assisting system is propped for the blind using YOLO for the object detection within images and video streams based on deep neural networks to make precise detection, and OpenCV under Python using Raspberry Pi3.The result obtained indicates the proposed approach in providing blind users the capability to travel in unfamiliar indoor and outdoor environments through an object identification model and user-friendly device.

With the rise of more up-to-date and current developments, the world of innovation has prospered at a rapid rate over the last decade. Our lives have become faster due to the use of more recent advances. The rapid advancement of information and arranged innovation has progressed from the internet and mechanization frameworks, which were initially used for regulatory workplaces and mechanical and commercial applications, to the apparatus of those advances all over life. The Internet has also grown in popularity over time. Each family has devised a strategy. Individuals began to seek a more beneficial and superior living environment. They began to consider the use of portable gadgets, apps, and versatile systems in natural checking, machine automation, smart home, and so on. Proficient and precise object recognition is a critical point in advancing computer vision frameworks. The introduction of machine learning and deep learning methods has dramatically increased the precision for object location. The project aims to integrate an Android application for object recognition and localization to achieve high accuracy and real-time performance.

The proposed system aims to create a visual aid image processing system for visually impaired people in which the user accepts speech commands. Its functionality addresses the identification of objects and signs. This will help the visually impaired person manage day-to-day activities and navigate their surroundings.

The paper intends to incorporate cutting-edge object detection techniques in order to achieve high accuracy and real-time performance. In this paper, we use Python in conjunction with a TensorFlow-based approach to solving the problem of object detection from start to finish. The resulting system is quick and precise. A TensorFlow-based application for an Android mobile device is built to detect objects using the device's built-in camera, specifically:

The framework is set up so that an android application (a'suming you're executing it on an Android gadget) will capture real-time outlines and send them to the background of the application, where all the computations take place.

- The video stream is sent and received as an input in the application's background, where it is tested and detected using accurate metrics by the COCO DATASETS object detection model.
- After testing with voice modules, the object's path will be converted into default voice notes, which can be sent to blind victims for assistance.
- In addition to object discovery, we have used an alarm framework to calculate an estimate. If the Blind Person is exceptionally close to the diagram or is far away in a more secure location, it will produce voice-based results in addition to distance units.

The main objective is to identify objects and signboards to help visually impaired persons manage everyday activities. This study will assist blind people by taking speech commands to detect objects using the image processing technique and will provide audio output to the person to track their way around the obstacles. This study will recognize some prominent signboards such as assign for "Washroom" and inform the blind person as soon as the sign is recognized.

II. Review of Literature

1. The current approaches for detecting models were explained in this work, as well as the standard datasets. This work discussed several detectors, such as one-stage and two-stage detectors, which aided in the analysis of various object detection methods and gathered some classic as well as innovative applications. There were also some branches relating to object detection identified. In addition, several development tendencies were identified in order better to follow the set of art algorithms and subsequent processes.
2. A fully convolutional network based on regions was given in this paper. For precise and efficient object detection, R-FCNN is used. As a result, this work can readily use ResNets as fully convolutional image classifier backbones for object detection. For object detection, this research offered a simple but effective R-FCNN architecture. When compared to the quicker R-FCNN, this approach obtains the same accuracy. As a result, it was easier to incorporate state-of-the-art picture classification backbones.
3. This Challenge serves as a reference point for object classification and detection. More than 100 item types and 1 million photos were categorized

and detected in this work. The method for collecting enormous amounts of data is described in this publication. Also, the most efficient algorithm for this data was explained, as well as the successes and failures of other algorithms.

4. The findings of this study revealed that oriented gradient grids outperform the present feature set for human recognition.
5. As object identification technology has advanced, many technologies have been used to autonomous vehicles, robots, and industrial facilities, according to this article. The benefits of these technology, however, are not reaching the visually handicapped, who are in desperate need of them. Using deep learning technologies, this paper suggested an object detection system for the blind. A voice guidance technique is also used to advise visually impaired people about the position of objects. The You Only Look Once (YOLO) technique is used in the object identification deep learning model, and a voice announcement is synthesized using text-to-speech (TTS) to make it easier for the blind to acquire information about items. As a result, it employs an effective object-detection system that aids the blind in locating objects within a given space.

III. Methodology

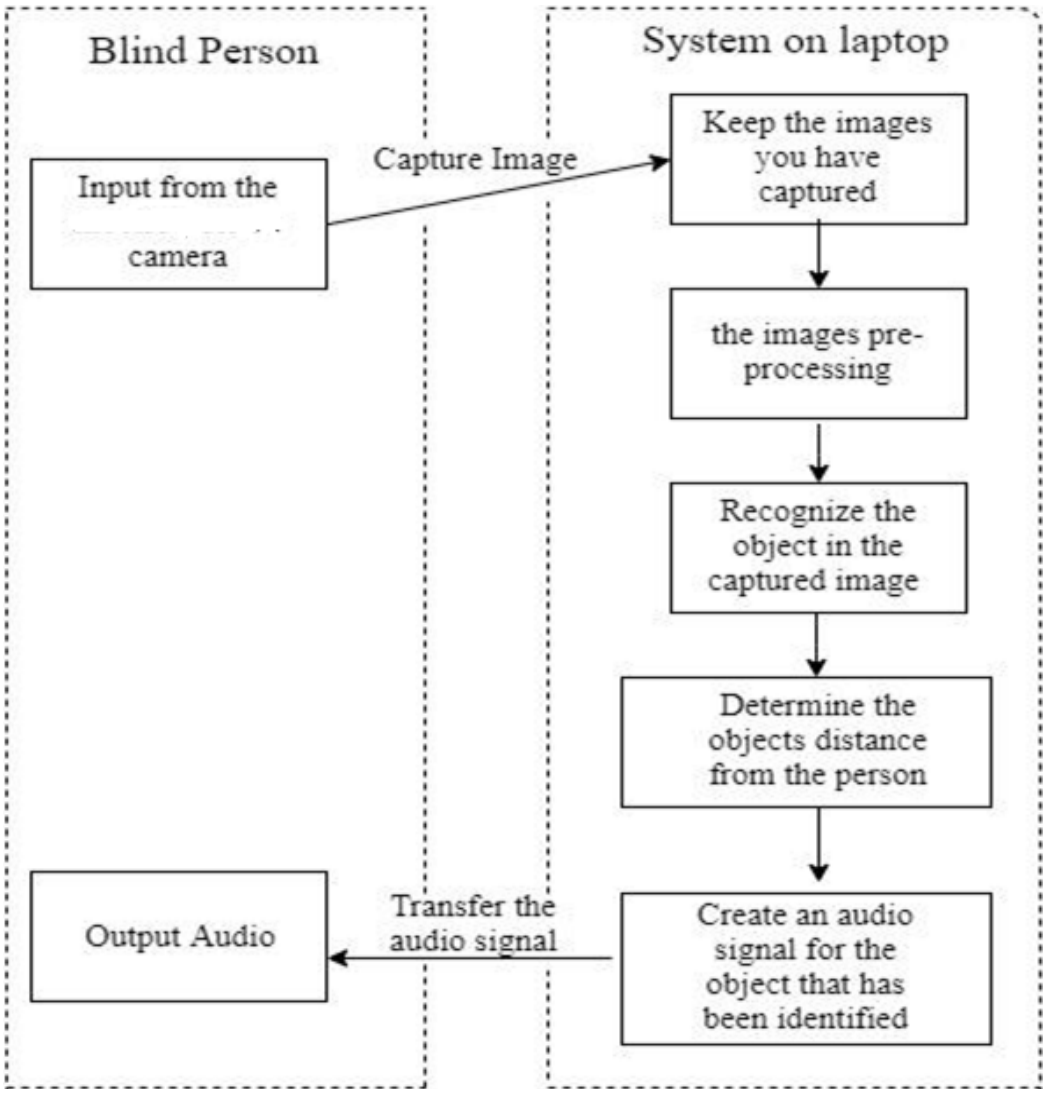

Fig. 1: System Methodology

IV. Overview of Technologies

TensorFlow API:

Fig. 2: Logo of Tensorflow

We have implemented by using TensorFlow APIs. The advantage we have by using APIs is that it provides us with a set of standard operations. So, we need not write the code for a program from scratch. APIs offer us convenience, and hence they are time savers, beneficial, and efficient. The TensorFlow object detection API is generally a structure built for creating deep learning networks that solve the problem of object detection. There are so many trained models in their framework, and they refer to it as 'Model Zoo.' This includes a collection of the COCO dataset, the KITTI dataset, and the Open Images Dataset.

TensorFlow Object Detection API depends on the libraries mentioned:

- Protobuf 3.0.0
- Python-tk
- Pillow 1.0
- Lxml
- Tf-slim
- Slim
- Jupyter notebook
- Matplotlib
- Tensorflow (1.15.0)
- Cython
- Contextlib2
- Cocoapi

V. Models

Now, a bunch of pre-trained models is with Tensorflow. You can use any one of them. They are pretty good and depending upon your system specifications you can choose one. For a faster accuracy, you can go with SSD DETECTION, and for better accuracy, you can go with MASK RCNN, but most of the system shows smooth performance with SSD Mobile Net DETECTION. So, I'll elaborate on SSD ALGORITHM.

VI. SSD Architecture

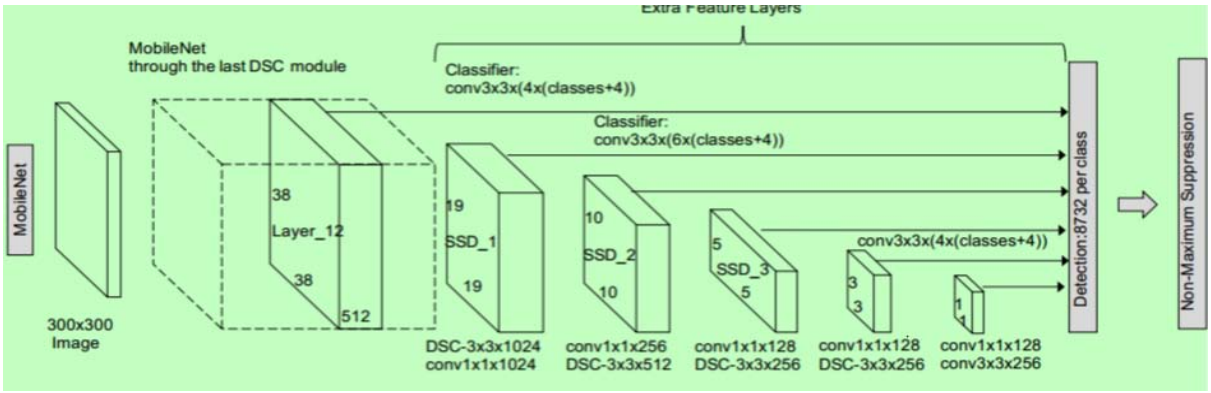

Fig. 3: SSD Architecture

SSD has two components: an SSD head and a backbone model.

The backbone model is basically a trained image classification network as a feature extractor. Like ResNet, this is typically a network trained on ImageNet from which the final fully connected classification layer has been removed.

The SSD head is nothing but one or more convolutional layers added to the backbone, and the outputs are explained as the bounding boxes and classes of objects in the spatial location of the final layer's activations. We are hence left with a deep neural network that is able to extract the meaning of the input image while preserving the spatial structure of the image at a lower resolution.

For an input image, the backbone results in 256 7x7 feature maps in ResNet34. SSD classifies the image using a grid and grid cell responsible for detecting

objects in the region of the picture. Detecting objects basically means predicting the class and location of an object within that region.

VII. Anchor Box

Multiple anchor boxes can be assigned to each grid cell in SSD. These designated anchor boxes are predefined, and each one is responsible for size and shape within a grid cell. The matching phase is used by SSD while training so that there's an appropriate match to anchor box with bounding boxes of each ground truth object within an image. The anchor box with the highest degree of overlap with an object is responsible for predicting that object's class and location. Once the network has been trained, this property is used to prepare the web and predict the detected objects and their places. Practically, each anchor box is specified with an aspect ratio and a zoom level. Well, we know that all things are not square. Some are shorter, some are very long, and some are wider by varying degrees. The SSD architecture allows predefined aspect ratios of the anchor boxes to account for this. The different aspect ratios can be specified using the ratios parameter of the anchor boxes associated with each grid cell at each zoom/scale level.

VIII. Zoom Level

The anchor boxes don't need to have the same size as the grid cell. The user might find both smaller or larger objects within a grid cell. To specify how much the anchor boxes need to be scaled up or down concerning each grid cell, the zooms parameter is used.

IX. Mobilenet

This model is based on the ideology of the MobileNet model based on depth-wise separable convolutions, and it forms a factorized Convolutions. This converts basic standard convolutions into depth-wise convolutions. These one × one convolutions are also called pointwise convolutions. For MobileNets to work, these depth-wise convolutions apply a general single filter-based concept to each input channel. These pointwise convolutions use a one × one convolution to merge with the outputs of the depthwise convolutions. As a standard convolution, both filters combine the inputs into a new set of outcomes in one single step. The depth-wise identifiable convolutions split this into two layers — a separate layer for the filtering purpose and the other separate layer for the combining purpose. This factorization methodology has the effect of drastically reducing the computation and that of the model size.

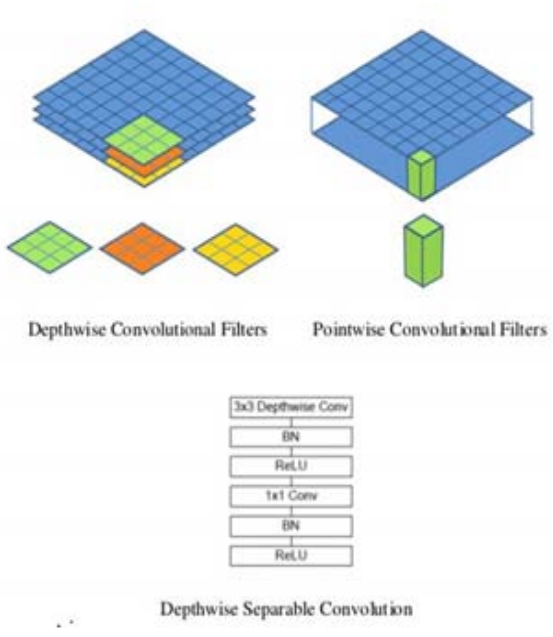

Fig. 4: Mobilenet Architecture

X. Depth Estimation

Depth estimation or extraction feature is nothing but the techniques and algorithms which aim to obtain a representation of the spatial building of a scene. In simpler words, it is used to calculate the distance between two objects. Our prototype is used to assist blind people, which aims to issue warnings to blind people about the hurdles coming on their way. To do this, we need to find how much distance the obstacle and person are located in any real-time situation. After the object is detected, a rectangular box is generated around that object.

Fig. 5: Depth Estimation

If that object occupies most of the frame, then concerning some constraints, the approximate distance of the thing from the particular person is calculated. Following code is used to recognize objects and return the information of the space and location.

```
(boxes, scores, classes, num_detections) = sess.run([boxes,
scores, classes, num_detections],feed_dict={image_tensor:
image_np_expanded})
```

Here, we have established a Tensorflow session comprised of Crucial Features for Detection. So, for further analysis, iteration is done through the boxes. Boxes are an array inside of a collection. So, for iteration, we need to define the following conditions.

```
for i,b in enumerate(boxes[0]):
boxes[0][i][0] – y axis upper start coordinates
boxes[0][i][1] – x axis left start coordinates
boxes[0][i][2] – y axis down start coordinates
boxes[0][i][3] – x axis right start coordinates
```

Index of the box in boxes array is represented by i. Analysis of the box's score is done by index. It is also used to access class. Now the width of the detected object is measured. This is done by asking the width of an object in terms of pixels.

```
apx_distance = round(((1 - (boxes[0][i][3]boxes[0][i][1]))**4),1)
```

We got the center of two by subtracting the same axis start coordinates and dividing them by two. In this way, the center of our detected rectangle is calculated. And at the end, a dot is drawn in the center. The default parameter for drawing boxes is a score of 0.5. if scores[0][i] >= 0.5 (i.e., equal or more than 50 percent) then we assume that the object is detected. if scores[0][i] >= 0.5:

```
            mid_x = (boxes[0][i][1]+boxes[0][i][3])/2
            mid_y = (boxes[0][i][0]+boxes[0][i][2])/2
            apx_distance = round(((1 - (boxes[0][i][3] -
boxes[0][i][1]))**4),1)
```

In the above formula, mid_x is the center of the X-axis, and mid_y is the center of the y axis. If the distance apx_distance< 0.5 and if mid_x> 0.3 and mid_x< 0.7 then it can be concluded that the object is too close to the particular person. With this code, the object's relative distance from a particular person can be calculated. After detecting an object, the code is used to determine the relative distance of the object from the person. If the object is too close, then a signal or a warning is issued to the person through the voice generation module.

XI. Voice Generation Module

After the detection of an object, it is of utmost importance to acknowledge the person about the presence of that object on their way. For the voice generation module, PYTTSX3 plays an important role. Pyttsx3 is a conversion library in Python which converts text into speech. This library works well with both Python 2 and 3. To get a reference to a pyttsx. Engine instance, a factory function called pyttsx.init() is invoked by an application. Pyttsx3 is a tool that converts text to speech easily.

This algorithm works whenever an object is being detected, and the approximate distance is being calculated. With the help of the cv2 library and cv2.putText() function, the texts are getting displayed on the screen. To identify the hidden text in an image, we use Python-tesseract for character recognition. OCR detects the text content on images and encodes it in a form that is easily understood by the computer. This text detection is done by scanning and analysis of the picture. Thus, the text embedded in images is recognized and "read" using Python-tesseract. Further, these texts are pointed to a pyttsx. Engine instance, a factory function called pyttsx.init() is invoked by an application. During construction, a yttsx.Driver. DriverProxy object is initialized by an engine that is responsible for loading a speech engine driver from the pyttsx.driver's module. After construction, an entity created by a machine is used by the application to register and unregister event call-backs; produce and stop speech; get and set speech engine properties; and start and stop event loops.

Pytorch is primarily a machine learning library. Pytorch is mainly applied to the audio domain. Pytorch helps in loading the voice file in standard mp3 format. It also regulates the rate of audio dimension. Thus, it is used to manipulate the properties of sound like frequency, wavelength, and waveform. The numerous availabilities of options for audio synthesis can also be verified by looking at the functions of Pytorch.

XII. Results and Discussion

The suggested system is focused on object detection. The technology has been designed to be wearable and portable. The system is attached to the person's chest. The video of the scene is captured by the Raspberry Pi camera, which is then translated into frames by the processor. The auditory output from the system directs the user to the object. Figure 6 depicts the detection of an object (blue cell phone) and a person. A chair is spotted with a person in Figure 7. The system outputs the object's name as well as the object's likelihood as a percentage. As a result, the system will only detect items with a probability larger than the set threshold. Because You Only Look Once (YOLO) is employed to implement the system on the Android platform, the accuracy of object recognition is reduced. It shows the object's name as well as its probability. Through the device's speakers, the programme also informs the user of the class designation and the distance between the object and the camera.

Fig. 6: Image Recognition of person and cell phone

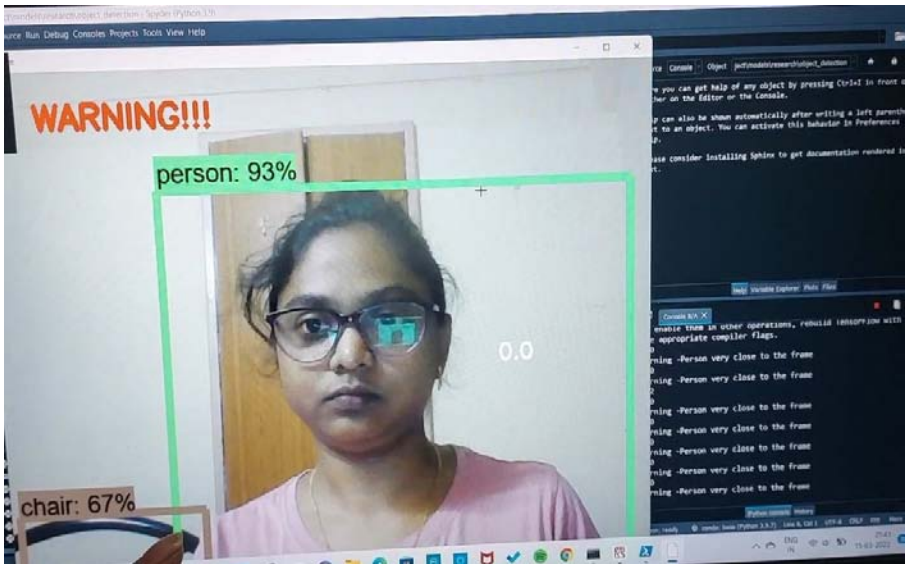

Fig. 7: Image Recognition of person and chair

XIII. Conclusion

Several technologies were developed for the assistance of visually impaired people. One such attempt is that we would wish to make an Integrated Machine Learning System that allows the blind victims to identify and classify real-time objects generating voice feedback and distance. Which also produces warnings whether they are very close or far away from the thing. For visually blind folks, this technology gives voice direction. This technology was created specifically to assist blind individuals. The precision, on the other hand, can be improved. Furthermore, the current system is based on the Android operating system, which may be modified to make it compatible with any convenient device.

Acknowledgments

We also express our thanks to the project reviewers Department of CSE, GITAM University, for their valuable suggestions and guidance for doing our project. We consider it a privilege to express our deepest gratitude to the Head of the Department, Computer Science Engineering for her valuable suggestions and constant motivation that immensely helped us complete this study. Finally, we deem it a great pleasure to thank everyone who helped us directly and indirectly throughout this project.

References Références Referencias

1. Choi D., and Kim M. (2018). Trends on Object Detection Techniques Based on Deep Learning, *Electronics and Telecommunications Trends*, 33(4): 23-32.
2. Dai J*et al*., (2016). R-FCN: Object Detection via Region-based Fully Convolutional Networks. *Conf. NeuralInform. Process. Syst*., Barcelona, Spain, Dec. 4-6, p. 379-387.
3. Dalal N. and Triggs B., Histograms of Oriented Gradients for Human Detection (2015). *IEEE Comput. Soc.Conf.Comput. Vision Pattern Recogn*., San Diego, CA, USA, June 20-25, p. 886-893.
4. Russakovsky O *et al*., (2015). ImageNet Large Scale Visual Recognition Challenge, *Int..J.Comput. Vision*, 115(3): 211-252.
5. Rajeshvaree Ravindra Karmarkar (2021). Object Detection System for the Blind with Voice Guidance, *International Journal of Engineering Applied Sciences and Technology*, 6(2): 67-70.