A PROJECT REPORT ON

NAVIGATION SYSTEM FOR VISUALLY IMPAIRED PERSONS

Submitted in the partial fulfillment For

the award of degree in

BACHELOR OF TECHNOLOGY

TO

Government College of Engineering, Karad

(An Autonomous Institute of Government of Maharashtra) Submitted

by

Anjali Kevale [17141134]

Simran Shinde [17141135]

Sneha Koli [17141140]

Under the Guidance of

Prof.A.V.Sathe

Department of Information Technology

Government College of Engineering,Karad

Academic Year 2019-2020

Government College of Engineering, Karad

(An Autonomous Institute of Government of Maharashtra)

Department of Information Technology

CERTIFICATE

This is to certify that the project entitled **“ NAVIGATION SYSTEM FOR VISUALLY IMPAIRED PERSONS ”** has been carried out by team:

Name of Student :

1. Anjali Kevale (17141134)
2. Simran Shinde (17141135)
3. Sneha Koli (17141140)

of Third Year B.Tech IT class under the guidance of **Prof.A.V.Sathe** during the academic year 2019 -20 (Sem-VI).

Prof.A.V.Sathe
Project Guide

Dr.S.J.Wagh
Head
Information Technology Department

External Examiner

ACKNOWLEDGEMENT

Apart from individual efforts, the success of any project depends largely on the encouragement & guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental throughout the project work

It is our privilege to express our gratitude towards our project guide**, Prof.A.V.Sathe**, for their valuable guidance, encouragement, inspiration and wholehearted cooperation throughout the project work. We thank her for being a motivation through all our highs and importantly, our lows.

We deeply express our sincere thanks to our Head of Department **Dr. S. J. Wagh** for encouraging and allowing us to present the project on the topic " **Navigation system for visually impaired persons** " and providing us with the necessary facilities to enable us to fulfill our project requirements as best as possible. We take this opportunity to thank all faculty members and staff of Department of Information Technology, who have directly or indirectly helped our project.

We pay our respects to honorable Principal **Dr. A. T. Pise** for their encouragement. Our thanks and appreciations also go to our family and friends, who have been a source of encouragement and inspiration throughout the duration of the project.

ABSTRACT

Visual impairment and blindness caused by various diseases has been hugely reduced, but there are many people who are at risk of age-related visual impairment. This project is focused mainly in the development of the computer vision module of the Smart Vision system. Visual information is the basis for most navigational tasks, so visually impaired people are at disadvantage because necessary information about the surrounding environment is not available. With the recent advances in inclusive technology it is possible to extend the support given to people with visual impairment during their mobility. In this context we propose a system, whose objective is to give blind users the ability to move around in unfamiliar environment, whether indoor or outdoor, through a user friendly interface.[4]

LIST OF FIGURES

ABBREVATIONS

Acronym	Definition
ML	Machine Learning
IXP	Image Processing
CNN	Convolutional neural network
RCNN	Region- Convolutional neural network
YOLO	You only look once
COCO	Common object in context
API	Application program interface

TABLE OF CONTENTS

1.Introduction

1.1 Background

Visually challenged persons face constraints in independent mobility and navigation. Mobility means the possibility of liberally moving, without support of any supplementary person, at home and unfamiliar scenarios. People with visual impairment tackle enormous limitations in terms of mobility. A system which guide or assist people with vision loss, ranging from partially sight to totally blind, by means of sound commands is referred as Navigation assistance for visually impaired (NAVI). Many researches are being conducted to build navigation system for blind people. Most of these technologies have limitations as its challenge involves accuracy, usability, interoperability, coverage which is not easy to overcome with current technology for both indoor and outdoor navigation.

Many problems in computer vision were saturating on their accuracy before a decade. How-ever, with the rise of deep learning techniques, the accuracy of these problems drastically improved. One of the major problem was that of image classification, which is defined as predicting the class of the image. A slightly complicated problem is that of image localization, where the image contains a single object and the system should predict the class of the location of the object in the image (a bounding box around the object). The more complicated problem (this project), of object detection involves both classification and localization.

Object recognition is to describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs. Image classification involves activities such as predicting the class of one object in an image. Object localization is refers to identifying the location of one or more objects in an image and drawing an abounding box around their extent. Object detection does the work of combines these two tasks and localizes and classifies one or more objects in an image. When a user or practitioner refers to the term "object recognition", they often mean "object detection". It may be challenging for beginners to distinguish between different related computer vision tasks.[1]

1.2 Importance of Project

The ability to navigate from place to place is a significant part of daily life. Human beings process the world around them mostly via the sense of sound and vision. It is general belief that vision plays a critical role, but many would have great difficulty in identifying the visual information they use, or when they use it. We find it easy to navigate in extremely familiar places without the sense of vision. This is possible mostly due to muscle memory. This can be experienced in examples such as going to the bathroom from your bedroom in the middle of the night. But only a small minority of people have experienced navigating largescale, unfamiliar environments without the aid of their eyes.[3]

2. Literature study

In various fields, there is a necessity to detect the target object and also track them effectively while handling occlusions and other included complexities.The nature of the techniques largely depends on the application domain. Some of the research works which made the evolution to proposed work in the field of object tracking are depicted as follows.

3.1 Object Detection

Object detection is an important task, yet challenging vision task. It is a critical part of many applications such as image search, image auto-annotation and scene understanding, object tracking. Moving object tracking of video image sequences was one of the most important subjects in computer vision. It had already been applied in many computer vision fields, such as smart video surveillance artificial intelligence, military guidance, safety detection and robot navigation, medical and biological application. In recent years, a number of successful single-object tracking system appeared, but in the presence of several objects, object detection becomes difficult and when objects are fully or partially occluded, they are obtruded from the human vision which further increases the problem of detection.[3]

1)S. TOSUN and E. KARAARSLAN, "Real-Time Object Detection Application for Visually Impaired People: Third Eye," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-6.Ss

By using this thesis and based on experimental results we are able to detect obeject more precisely and identify the objects individually with exact location of an obeject in the picture in x,y axis.This paper also provide experimental results on different methods for object detection and identification and compares each method for their efficiencies.

The system proposed here is a novel method for obstacle detection and identification. It can be easily commercialized and be made to benefit the visually impaired community. Unlike other existing models, it does not require a large database because of the pre-trained

Cognitive Neural Network model. The Single Shot Detection MobileNet model has been trained using the COCO[9] dataset which contains almost 300 thousand images. Hence, it can recognize any object without needing a database. From the experiments it was concluded that the system works extremely accurate in identifying people. It has 21 mAP (Mean Average Precision). Common place objects are also identified with satisfactory accuracy, provided they are sharply defined.

2) Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490. doi:10.1109/TPAMI.2004.108 Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach.Intell.26,1475–1490.

The aim of object detection is to detect all instances of objects from a known class, such as people,cars or faces in an image. Generally, only a small number of instances of the object are present in the image, but there is a very large number of possible locations and scales at which they can occur and that need to somehow be explored. Each detection of the image is reported with some form of pose information. This is as simple as the location of the object, a location and scale, or the extent of the object defined in terms of a bounding box. In some other situations, the pose information is more detailed and contains the parameters of a linear or non-linear transformation. For example for face detection in a face detector may compute the locations of the eyes, nose and mouth, in addition to the bounding box of the face

3) "Smart Navigation System for the Visually Impaired Using Tensorflow"

Mrs.R.Rajalakshmi1 , Ms.K.Vishnupriya2 , Ms.M.S.Sathyapriya3 ,Ms.G.R.Vishvaardhini4

The use of the suggested system is very simple and easy. The system consists of a raspberry pi, a camera and an ultrasonic sensor. The total weight of the system, when made into prototype form will weigh less than a kilogram. The camera will be mounted on a wearable glass, the ultrasonic sensor could be attached to the belt and the Raspberry Pi can be placed in any convenient location on the user. Once the system is switched on it would continuously provide information about the surrounding.

As mentioned earlier there are a lot of research and work going towards making navigation easier for the visually challenged. The use of bionic eyes are revolutionary. Despite so much effort being put into it, many in the blind community still rely on their memory or the kindness of others for their day to day locomotion.

3.Related theory and Problem definition

3.1.Problem definition

Visually impaired individuals will face many difficulties and one of the common difficulties is when they involve in self-navigating at an environment which is strange for them. In fact, physical movement is one of the biggest challenges for them. Besides that, while they travel around or walking at a crowded corridor, it may pose great difficulty. One of the existing problems for visually impaired individuals to travel in a corridor is that they cannot detect either they need to turn left or turn right when reached to the end of the corridor by using only the walking stick.

To walk at the corridor, the visually impaired individuals must find the border of the sidewalk at the corridor and then use their walking stick to define their current location. The reason why the visually impaired individuals do that is because they cannot forecast the obstacle which is far from them while they only can use the walking stick to detect the area around them.[4]

3.2.Related theory

A number of devices have already been developed to address some of the difficulties faced by visually impaired people with regard to travel. For instance, simple Electronic Travel Aids have been in development since 1897 (Brabyn, 1985). Real and more complex developments occurred after the Second World War and through the 1950s and 60s (Heyes, 1983). With the advent of the possibilities of remote sensing in the form of ultrasound and radar more research effort was directed at the problems of remote sensing of the environment for visually impaired people.[5]

3.2.1.TensorFlow

The TensorFlow API is widely used in the field of object detection. In the era of facial recognition, it seems imperative to make use of the advanced technology to recognize objects as well. TensorFlow APIs can be used to detect with bounding boxes, objects in images and or videos using either some of the pre-trained models made available or through models which you can train on your own which the API also makes it easier. TensorFlow, an open source machine library with a branch of machine learning called deep learning. It has led to proficient improvements in many zones mainly image classification and recognition. Deep learning has major advantage when working in the field of any texture of images. It is also done by classifiers. Classifiers are nothing but a set of codes or modules (Functions). The classifiers we use are preferably a high type of classifier namely neural network which could learn more complex functions. In order to make image classification and recognition at a comparatively easier pace, miscellaneous datasets have been created by the TensorFlow open source environment. COCO[9] (Common Objects in Context) has several features such as object segmentation, recognition in context, super pixel stuff segmentation, 1.5 million object instances, 330 K images, 80 object categories, 91 stuff categories, 5 captions per image, and 250,000 people with key points. COCO 2017 train/val browser has over 123,287 images and 886,284 instances.[1]

Fig:3.2.1 COCO Dataset

Deep learning is the basics of the object detection model used here. The concept of deep learning is a new area in machine learning. We are basically training the system to learn on its own. Usually there are three types of machine learning, supervised, semi-supervised and unsupervised learning. In this suggested system we will be using the pretrained model which will be working on the 90 classes of COCO's dataset. Some of the models developed which uses COCO are:

Model name	Speed	COCO mAP	Outputs
ssd_mobilenet_v1_coco	fast	21	Boxes
ssd_inception_v2_coco	fast	24	Boxes
rfcn_resnet101_coco	medium	30	Boxes
faster_rcnn_resnet101_coco	medium	32	Boxes
faster_rcnn_inception_resnet_v2_atrous_coco	slow	37	Boxes

Fig:3.2.2 Object detection model

One of the major competencies in the field of computer vision is the object detection. The RCNNSs could be classified into three advancements namely R-CNN, Fast R-CNN, and Faster R-CNN. The discrimination among these three could be done using parameters such as test time per image, speed up and map. The COCO dataset is used in Faster RCNN. The Regional Convolutional Neural Network plays a significant role in image segmentation.[4]

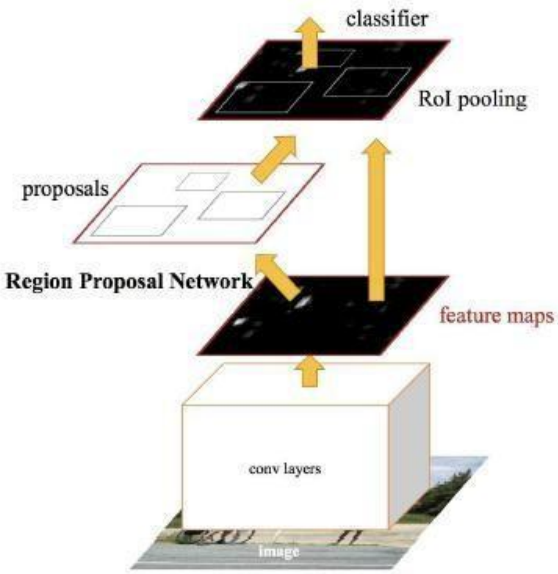

Fig:3.2.3 Faster RCNN model

In the convolutional layer, the neurons apply convolutional operation to input. In this layer, dot product between the image chunk and our filter(w) is done. This results in a single number output to which bias(b) is added. If out input is of size N*N, filter size is F, stride is S and input is added with a 0 pad of size P, then the output size will be: **(N-F+2P)/S+1**

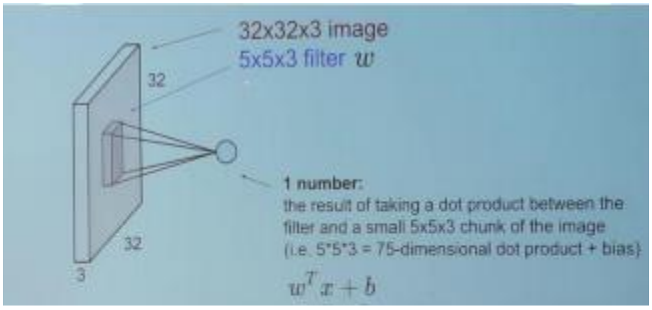

Fig:3.2.4 Convolutional Layer

If we have width w, height h and depth d, then after pooling the image having w1*h1*d1 will be changed as:

w2=(w1-f)/S+1

h2=(h1-f)/S+1

d2=d1

OpenCV :

The OpenCV software is required for handling the operation of the camera used in the system. The function CaptureVideo() is important for beginning the use of the camera. The ffmpeg file is necessary for the separation of each frame. This is important as we need to process each frame separately. The other functions which come into use are:

• read() waitKey() imshow() The functions are used to display a window upon which the detected objects boxed up with their names coming up on the top along with the percentage up to which the match is made.[3]

4.Tools and Technologies

4.1.Faster region convolutional neural network

The faster region convolutional neural network is another state-of-the-art CNN-based deep learning object detection approach. In this architecture, the network takes the provided input image into a convolutional network which provides a convolutional feature map. Instead of using the selective search algorithm to identify the region proposals, a separate network is used to learn and predict these regions. The predicted region proposals are then reshaped using a region of interest (ROI) pooling layer, which is then used to classify the image within the proposed region and predict the offset values for the bounding boxes.

some of the drawbacks of R-CNN to build a faster object detection algorithm and it was called Fast R-CNN. The approach is similar to the R-CNN algorithm. But, instead of feeding the region proposals to the CNN, we feed the input image to the CNN to generate a convolutional feature map. From the convolutional feature map, we identify the region of proposals and warp them into squares and by using a RoI pooling layer we reshape them into a fixed size so that it can be fed into a fully connected layer. From the RoI feature vector, we use a softmax layer to predict the class of the proposed region and also the offset values.[6]

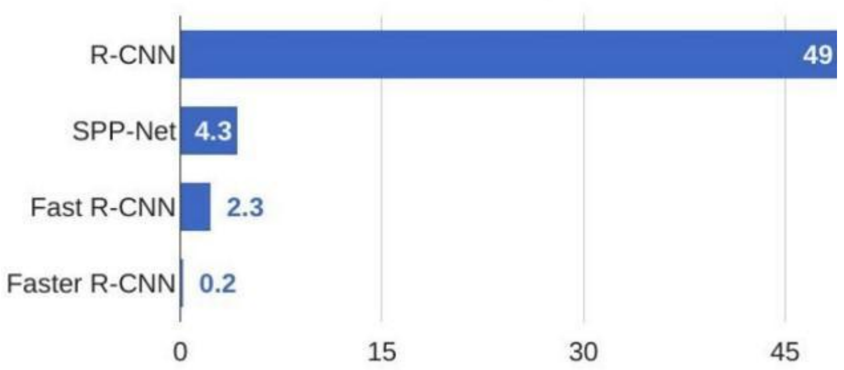

Fig:4.1 Comparison of test-time speed of object detection algorithm

4.2.You only look once

You only look once is a state-of-the-art object detection algorithm which targets real-time applications, and unlike some of the competitors, it is not a traditional classifier purposed as an object detector.

YOLO works by dividing the input image into a grid of S×SS×S cells, where each of these cells is responsible for five bounding boxes predictions that describe the rectangle around the object. It also outputs a confidence score, which is a measure of the certainty that an object was enclosed.Therefore the score does not have any relation with the kind of object present in the box, only with the box's shape.[1]

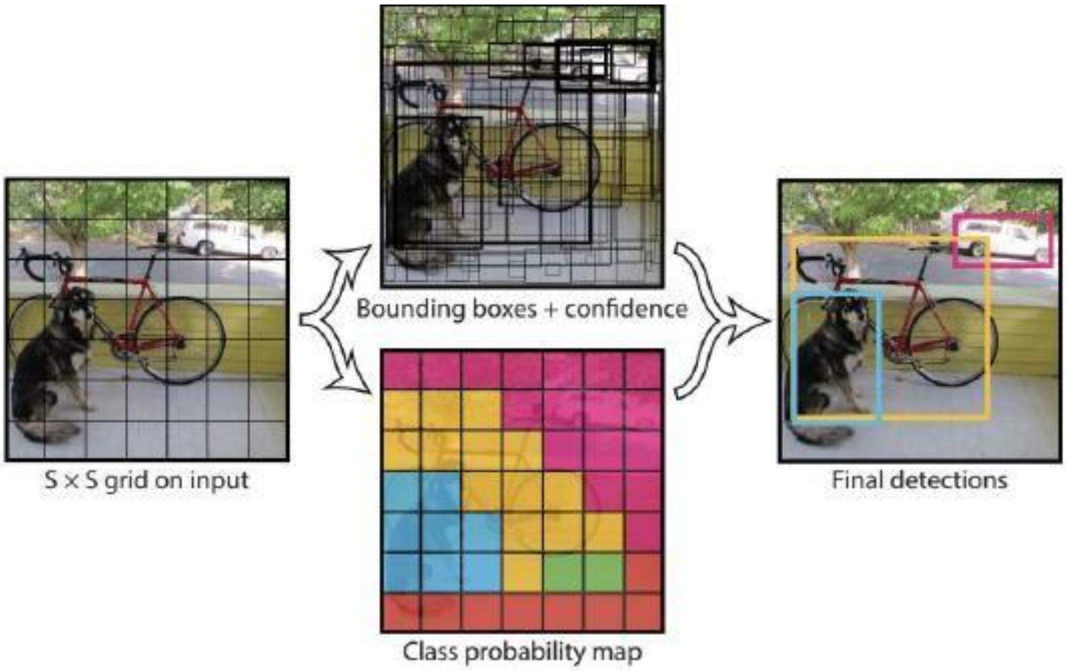

Fig.4.2.YOLO

5. System Design and Analysis

5.1.1 System design and modeling

How-ever, with the rise of deep learning techniques, the accuracy of these problems drastically improved. One of the major problem was that of image classification, which is defined as predicting the class of the image. A slightly complicated problem is that of image localization, where the image contains a single object and the system should predict the class of the location of the object in the image (a bounding box around the object). The more complicated problem (this project), of object detection involves both classification and localization.In this case, the input to the system will be a image, and the output will be a bounding box corresponding to all the objects in the image, along with the class of object in each box.[2]

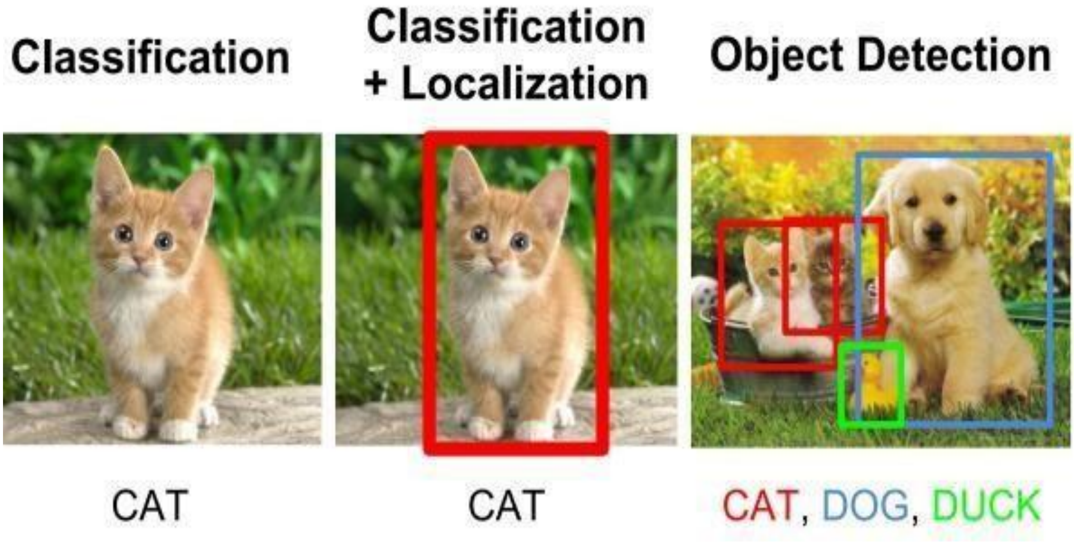

Fig: 5.1: Image classification,Localization and Detection

5.1.2.Bounding Box

The bounding box is a rectangle drawn on the image which tightly fits the object in the image. A bounding box exists for every instance of every object in the image. For the box,4 numbers (center x, center y, width, height) are predicted. This can be trained using a distance measure between predicted and ground truth bounding box. The distance measure is a jaccard distance which computes intersection over union between the predicted and ground truth boxes as shown.[2]

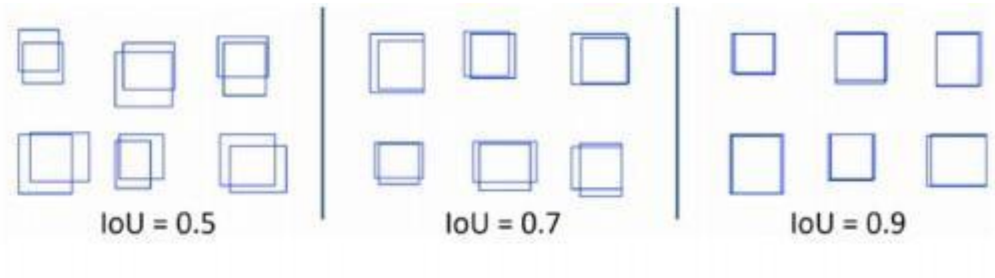

Fig:5.2 Jaccard distance

5.1.3.Classification + Regression

The bounding box is predicted using regression and the class within the bounding box is predicted using classification. The overview of the architecture is shown.

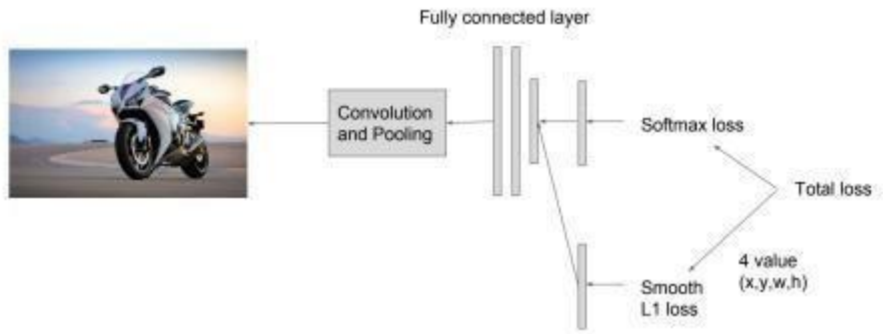

Fig:5.3 Architecture Overview

5.1.4.Two-stage Method

In this case, the proposals are extracted using some other computer vision technique and then resized to fixed input for the classification network, which acts as a feature extractor. Then an SVM is trained to classify between object and background (one SVM for each class). Also a bounding box regressor is trained that outputs some some correction (offsets) for proposal boxes. The overall idea is shown in Fig. These methods are very accurate but are computationally intensive (low fps).

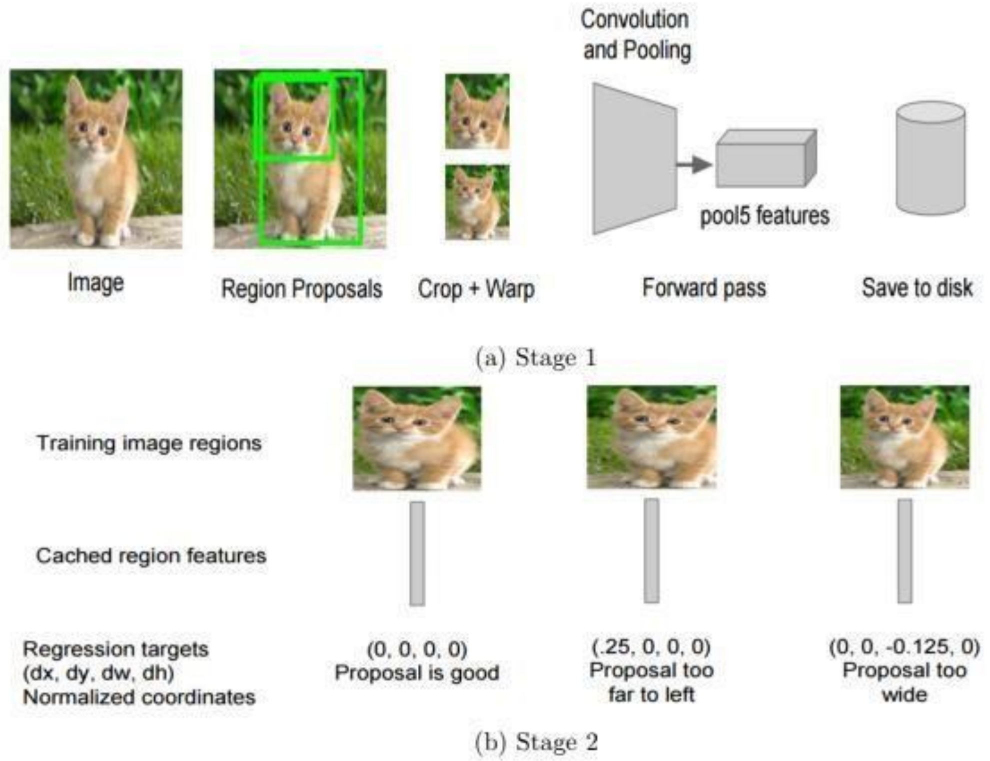

Fig:5.4 Two stage method

5.2 Web Application

Web sites most likely to be referred to as "web applications" are those which have similar functionality to a desktop software application, or to a mobile app. HTML5 introduced explicit language support for making applications that are loaded as web pages, but can store data locally and continue to function while offline.

Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source. Django was designed to help developers take applications from concept to completion as quickly as possible. Django takes security seriously and helps developers avoid many common security mistakes. Some of the busiest sites on the Web leverage Django's ability to quickly and flexibly scale.

For our project, we have created a web application using Django using Python. It has a page for guidance to visually impaired persons that it will give the instructions to person like there is table or there is car likewise. So that the person get aware of that things. Also, it provides a web page which take quires of person. Or one can give suggestion for that application.

6.Result

6.1 Implementation of proposed system

The project is implemented in python 3. Tensorflow was used for training the deep network and OpenCV was used for image pre-processing.

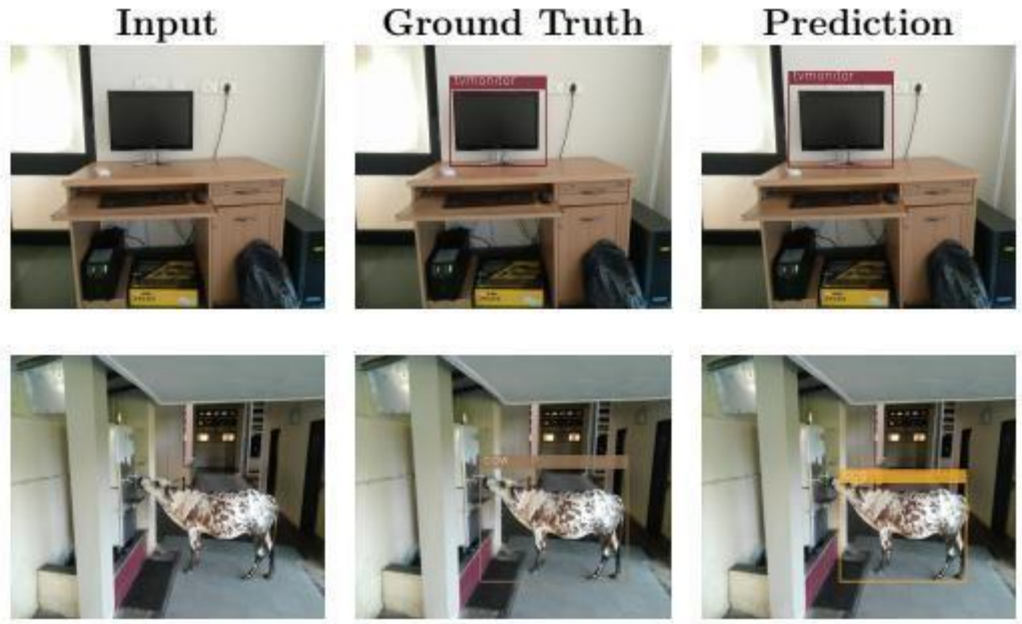

Fig:6.1.1 Detection result on custom dataset

6.2 Experimental Results

6.2.1 Dataset

COCO stands for Common Objects in Context. As hinted by the name, images in COCO dataset are taken from everyday scenes thus attaching “context” to the objects captured in the scenes. We can put an analogy to explain this further. Let’s say we want to detect a person object in an image. A non-contextual, isolated image will be a close-up photograph of a person. Looking at the photograph, we can only tell that it is an image of a person. However, it will be challenging to describe the environment where the photograph was taken without having other supplementary images that capture not only the person but also the studio or surrounding scene.

COCO was an initiative to collect natural images, the images that reflect everyday scene and provides contextual information. In everyday scene, multiple objects can be found in the same image and each should be labeled as a different object and segmented properly. COCO dataset provides the labeling and segmentation of the objects in the images. A machine learning practitioner can take advantage of the labeled and segmented images to create a better performing object detection model.[5]

Fig:6.2.1 Image classification

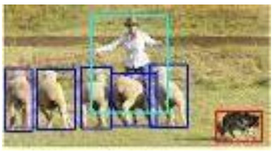

Fig:6.2.2 Image localization

Fig:6.2.3 Semantic segmentation

Fig:6.2.4 This work

6.2.2 Real time implementation

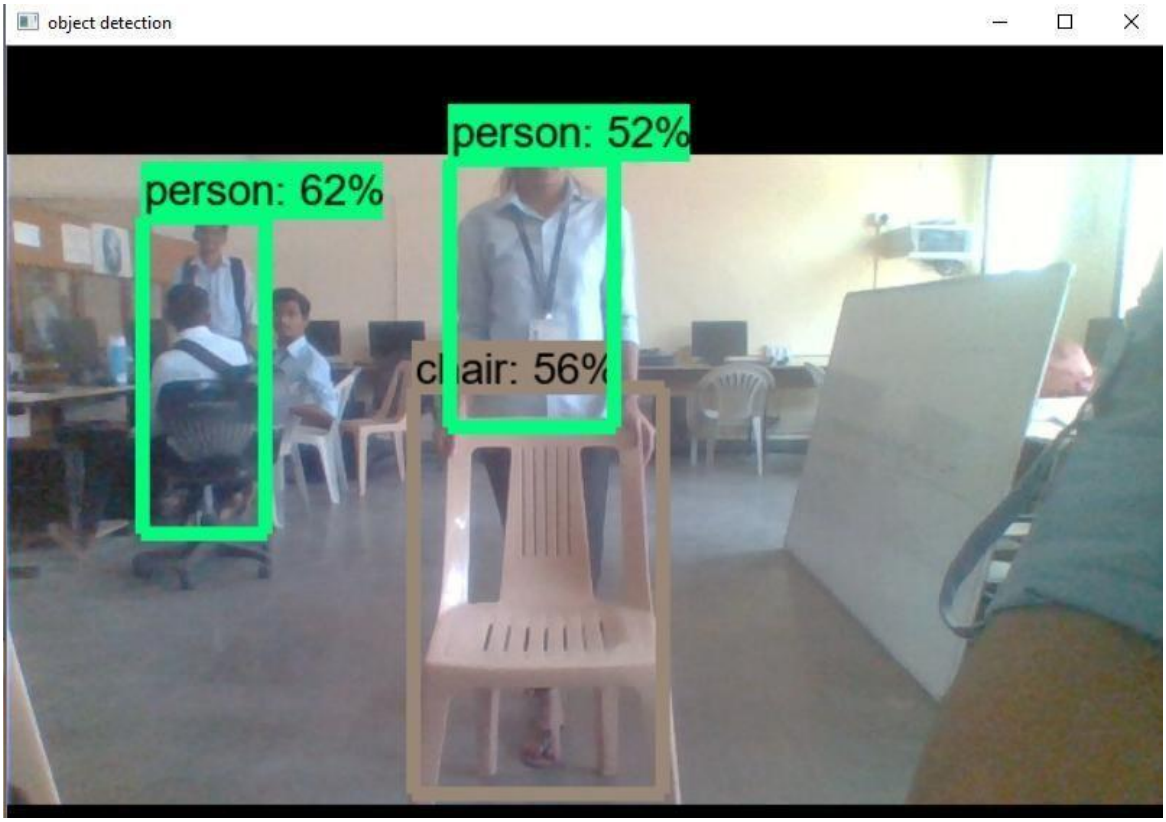

Fig 6.2.5 : Real time implementation

7. Conclusion and Future Scope

7.1 Conclusion

The system proposed here is a novel method for obstacle detection and identification. It can be easily commercialized and be made to benefit the visually impaired community. Unlike other existing models, it does not require a large database because of the pre-trained Cognitive Neural Network model. The Single Shot Detection MobileNet model has been trained using the COCO[9] dataset which contains almost 300 thousand images. Hence, it can recognize any object without needing a database. From the experiments it was concluded that the system works extremely accurate in identifying people. It has 21 mAP (Mean Average Precision). Common place objects are also identified with satisfactory accuracy, provided they are sharply defined.

7.2 Future Scope

The future scope of the project determines to recognize any kind of object irrespective of its nature and scope. With the aid of more complex processors and improvement in profound technology, the extension of this module could identify any kind of entity with even more faster frame rate. The text to speech part could also be developed according to the futuristic pace. Instead of using the pre-trained models we can train the model by ourselves. The model can be trained to recognize objects which are frequently encountered by the user. Thus, it can be customized for the specific needs of the user and ensure safer navigation. The faster RCNN model which is already available, though very accurate is still slow. Improvements in the model will be a huge boost towards accurate and quick object detection.

References

[1] S. TOSUN and E. KARAARSLAN, "Real-Time Object Detection Application for Visually Impaired People: Third Eye," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-6.

[2] Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490.doi:10.1109/TPAMI.2004.108 Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490.

[3] Smart Navigation System for the Visually Impaired Using Tensorflow

Mrs.R.Rajalakshmi1 , Ms.K.Vishnupriya2 , Ms.M.S.Sathyapriya3 ,Ms.G.R.Vishvaardhini4

[4] S. S. Nijhawan, A. Kumar, S. Bhardwaj and G. Nijhawan, "Real-time Object Detection for Visually Impaired with Optimal Combination of Scores," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 307-311.4

[5] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[6] International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 1, January 2015