C.Santhosh Kumar, et.al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 5, (Series-II) May 2021, pp. 16-23

RESEARCH ARTICLE OPEN ACCESS

DLWAVIP - Deep Learning Based Web Assistance for Visually Impaired People

[1]C.Santhosh Kumar, [2]Yasaswini.P, [3]Dr.R.Vijayabhasker, [4]Geetha.B, [5]Roja.S
[1]Assistant Professor/CSE, Er.Perumal Manimekalai College of Engineering, Hosur, Tamilnadu
[3]Assistant Professor/ECE, Anna university Regional Campus, Coimbatore
[2,4,5] Final Year, Department of CSE, Er.Perumal Manimekalai College of Engineering, Hosur ,Tamilnadu

ABSTRACT
The visually challenged and blind people face more difficulties in their day to day living. Many people suffer from the blindness, these leads them the need of assistance or guide to move around in a surrounding and also to do their day to day activities .Our project is going to be the assistant for blind people. Human vision has the capability of capturing and storing billions of images in brain and realizing the images by comparing with pre images .We in our project going to do the same thing wetrain the image to AI using CNN algorithm by giving thousands of images as input.So,our AI can detect the images easily .We also classify the images between indoor and outdoor .Our application is easy to use and output will be in the form of audio through google voice.

Date of Submission: 21-04-2021 Date of Acceptance: 06-05-2021

I. INTRODUCTION

Beyond the limits,the blind people can travel in a known environment without the sense of vision. But,it is very difficult for them to move in an unknown environment. It is possible due to muscle memory. The blind people will have the sense of smell and their hearing more sharply, as they rely lot on these sensors. The application we developed detects the objects in user surrounding it alerts the user if any things available in this front. For example: At home,while the user is moving ,if staircase, chairs, tables, walls, persons etc..come as obstacle the AI detects and alerts the person by saying through google voice. for example, "Electronic display item detected move aside" .While moving in outdoors ,it detects the path holes, traffic lanes, animals ,vechicles etc. For example ,If path holes are detected it says the person "path hole detected walk carfully". We use CNN algorithm for classifying and d detecting the item. our main motive is to help the blind people and make their lives easy.

II. LITERATURE SURVEY

[1] Daniyal Rajput et al discusses about Smart Obstacle Detector for Blind Person .They proposed different types of obstacles in front of the user, their size and their distance from the user. MATLAB Software is used for signal processing. The camcorder is used for recording videos. Video processing methods are used after that. The output of this system not only gives output in audio format but also vibration. A vibrating motor has been connected with an ultrasonic sensor. The ultrasonic sensor detects objects coming in its range and this makes the vibrating motor vibrate. [2] Khushboo Khurana, et al system tries to detect multiple objects in an image. That is the core specialty of the system. It is a system where N object detectors are trained for N different objects. When an image is sent to the system, all object detectors do their work. If an object is found by a detector, it will mark its boundary and label the object name. After the process completes for all N detectors, the image is displayed with all the tags. Moving a cursor over an object in the image shows the complete boundary of the object with its label beside. This system is a little slower than other systems because a lot of object detectors are working on a single image. The performance can increase by allowing more than one object detectors to run in parallel. [3] Payal Panchal et al reviewed system which first subtracts the current frame from the previous one and obtains a maximum value of the difference between two pixel values. Maximum value > given pixel = Foreground. Maximum value < given pixel = Background. The brightness distortion and Chromaticity distortion are also taken care of in this project by using shadow detection technique theory. From a video, objects are detected by taking templates out of the video. Of course this is not the best way. It works if the object is present in the whole video. Compared to other features like SIFT, Shape features are better used to detect objects in images. Hence, in this project, the local features are

replaced by Shape features. [4] Prof. SeemaUdgirkar et al proposes wearable device and consists of a blind stick and sensor based detection circuit. It uses an infrared sensor which uses infrared waves to scan the surroundings of a person. It uses object detection and gives them audio information about it. The system must be trained about object information. Feature extraction is also a part of the process.[5]Bing jiang et al explains evaluation to select the captured images through vision sensors, which can ensure the input quality of scenes for the final identification system. They used binocular vision sensors to capture images in a fixed frequency and choose the informative ones based on stereo image quality assessment. Then they captured images will be sent to cloud for further computing. [6] Dakopoulos.D et al in his paper deals with the design, simulation and implementation of a 2D vibration array used as a major component of an assistive wearable navigation device for visual impaired. [7] Hseuh-cheng-wang et al uses techniques from computer vision and motion planning to (1) identify walkable space (2) plan step-by-step a safe motion trajectory in the space and (3) recognize and locate certain types of objects, for example the location of an empty chair.[8] In this model, a down swept FM ultrasound signal is emitted from a transmitting array with broad directional characteristics in order to detect obstacles. The ultrasound reflections from the obstacles are picked up by a two-channel receiver. With it a blind person can recognize a 1-mm-diameter wire. It was also proved that the blind could discriminate between several obstacles at the same time without any virtual images. This mobility aid, modeled after the bat's echolocation system, is very effective at detecting small obstacles placed in front of the head.[9] Jinqiang Bai et al illustrates the prototype consisting of a pair of display glasses and several low-cost sensors is developed and its efficiency and accuracy were tested by a number of users. The experimental results shows the smart guiding glasses can effectively improve the user's travelling experience in complicated indoor environment. Thus it serves as a consumer device for helping the visually impaired people to travel safely.[10] P B Meijer et al in his paper presents an experimental system for the conversion of images into sound patterns. The system was designed to provide auditory image representations within some of the known limitations of the human hearing system, possibly as a step towards the development of a vision substitution device for the blind. [11] satyanarayana et al proposed processing methodology is found to be effective for object identification and for producing stereo sound patterns in the NAVI system.

PROPOSED APPROACH

Current wearable device consists of a blind stick and sensor based detection circuit. It uses an infrared sensor which uses infrared waves to scan the surroundings of a person. It uses buzzer sound notification. The issue in the existing system is If the object is detected it gives the buzzer sound continuously. It will not detect whether it is moving or idle object. It is applicable for outdoor only.It will not recognize the object's identity.It needs some wearable device and hardware cost is high. The proposed system captures the image through the continuous video stream rather than taking the picture of each object every time. The object detection and recognition have to be accurate and the detecting speed of the object should be high so that the navigation for the blind people would be easier. For such high-speed detection, CNN algorithm has been implemented. The first step is to capture the objects using the camera of this the application should get the camera permission of the device. Then the image is sub-segmented so that multiple regions can be formed from a single image. The algorithm then combines similar regions to form a larger region and finally produce the region of interest. The pre-trained convolutional neural network is trained again based on the number of classes that has to be detected then the region of interest is identified. Based on the region of interest, the objects and the backgrounds are classified. For each identified object in the image, tighter bounding boxes are generated based on the linear regression model. The input image to the CNN, the object is detected in indoor environment like person, chair, bench and outdoor environments like car, bus, animal, motorcycle etc. The recognized information is in the image format but to the blind people, the information has to be provided in the vocal format based on speech synthesizer. The speech synthesizer used is Pyttsx3i.e Text to speech. It is a text to speech conversion library in python. The pyttsx3 module supports two voices: the first is female and the second is male which is provided by "sapi5" for windows. It is a very easy use tool which converts the entered text into speech. An application invokes the pyttsx3.init() factory function to get a reference to a pyttsx3.The advantages is this project is applicable for both outdoor and indoor. It identifies whether the object is moving or idle. It also recognizes the identity of the object. Instead of buzzer sound and vibration it tells what object is present through google voice.

ARCHITECTURE DESIGN

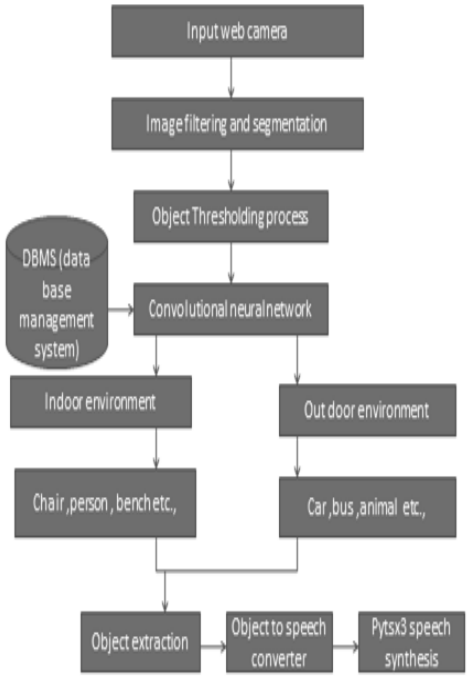


Fig.No.1.Architecture Design for Proposed System

SYSTEM MODULES

There are 5 modules used in this project

- Camera Module
- Image Pre-processing and Segmentation Module
- Database Module
- Recognition Module
- Audio Module

Camera Module:

In this module the input video streaming can be captured by using camera. The capture video frames can be converted as image by extracting the image from video.

Image Pre-Processing and Segmentation Module:

In image pre-processing module, the input extracted images can be sub-segmented so that multiple regions can be formed from a single image.

Database Module:

The model of trained images can be stored in data base management system for detection of objects in indoor and outdoor environments.

Recognition Module:

The conventional neural network algorithm is used for recognition of objects in indoor and outdoor environments by comparing with data base system.

Audio Module:

The recognized object images format can be converted as audio format based on speech synthesizer and the audio output is easy for understanding blind and self-navigate without any persons help.

Convolution Neural Network

When programming a CNN, the input is a tensor with shape (number of images) x (image width) x (image height) x (image depth). Then after passing through a convolution layer, the image becomes abstracted to a feature map, with shape (number of images) x (feature map width) x (feature map height) x (feature map channels). A convolution layer within a neural network should have the following attributes: Convolution kernels defined by a width and height (hyper-parameters).The number of input channels and output channels (hyper-parameter).The depth of the Convolution filter (the input channels) must be equal to the number channels (depth) of the input feature mapping.put feature map.

Hidden Layer: The input from Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layers can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by addition of learnable biases followed by activation function which makes the network nonlinear.

Output Layer: The output from the hidden layer is then fed into a logistic function like sigmoid or soft max which converts the output of each class into probability score of each class.

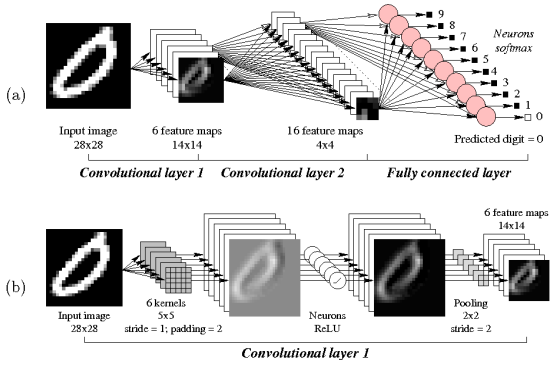


Fig No.2. Convolutional Neural Network

The data is then fed into the model and output from each layer is obtained this step is called feed forward, we then calculate the error using an error function, some common error functions are cross entropy, square loss error etc.

After that, we back propagate into the model by calculating the derivatives. This step is called Back propagation which basically is used to minimize the loss.Here's the basic python code for a neural network with random inputs and two hidden layers.

```
activation = lambda x: 1.0/(1.0 + np.exp(-x)) # sigmoid function
input = np.random.randn(3, 1)
hidden_1 = activation(np.dot(W1, input) + b1)
hidden_2 = activation(np.dot(W2, hidden_1) + b2)
output = np.dot(W3, hidden_2) + b3
```

W1, W2, W3, b1, b2, b3 are learnable parameter of the model.

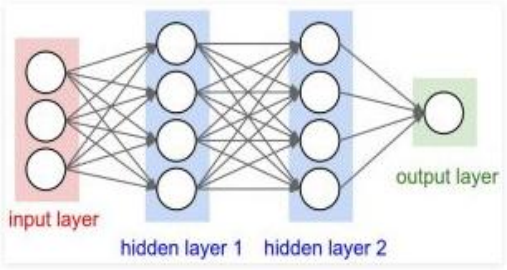


Fig.No.3.convolution layers

Convolution Neural Network

Convolution Neural Networks or covnets are neural networks that share their parameters. Imagine you have an image. It can be represented as a cuboid having its length, width (dimension of the image) and height (as image generally have red, green, and blue channels).

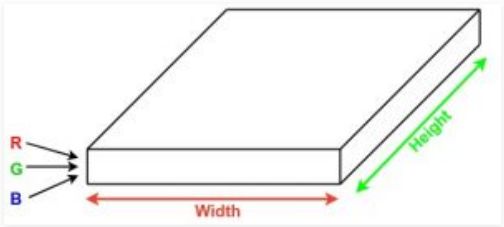


Fig.No.4.Small Neural Network

Instead of just R, G and B channels now we have more channels but lesser width and height. his operation is called Convolution. If patch size is same as that of the image it will be a regular neural network. Because of this small patch, we have fewer weights.

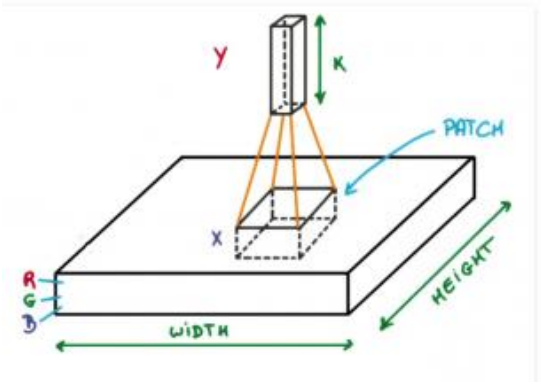


Fig.No.5.Deep learning

Convolution layers consist of a set of learnable filters . Every filter has small width and height and the same depth as that of input volume (3 if the input layer is image input).For example, if we have to run convolution on an image with dimension 34x34x3. Possible size of filters can be axax3, where 'a' can be 3, 5, 7, etc but small as compared to image dimension.

During forward pass, we slide each filter across the whole input volume step by step where each step is called stride (which can have value 2 or 3 or even 4 for high dimensional images) and compute the dot product between the weights of filters and patch from input volume. As we slide our filters we'll get a 2-D output for each filter and we'll stack them together and as a result, we'll get output volume having a depth equal to the number of filters. The network will learn all the filters. Let's take an example by running a covnets on of image of dimension 32 x 32 x 3.

Input Layer: This layer holds the raw input of image with width 32, height 32 and depth 3.

Convolution Layer: This layer computes the output volume by computing dot product between all filters and image patch. Suppose we use total 12 filters for this layer we'll get output volume of dimension 32 x 32 x 12.

Activation Function Layer: This layer will apply element wise activation function to the output of convolution layer. Some common activation functions are RELU: max(0, x), Sigmoid: 1/(1+e^-x), Tanh, Leaky RELU, etc. The volume remains unchanged hence output volume will have dimension 32 x 32 x 12.

Pool Layer: This layer is periodically inserted in the covnets and its main function is to reduce the size of volume which makes the computation fast reduces memory and also prevents from overfitting. Two common types of pooling layers are max pooling and average pooling. If we use a max pool with 2 x 2 filters and stride 2, the resultant volume will be of dimension 16x16x12.

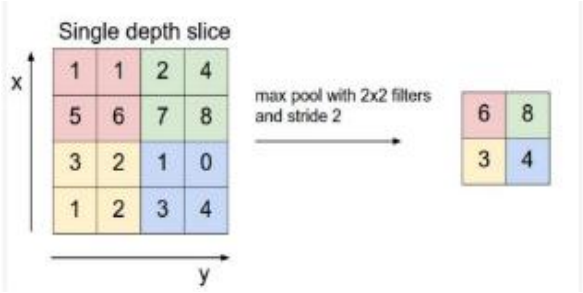


Fig No.6. Pool layer

Fully-Connected Layer: This layer is regular neural network layer which takes input from the previous layer and computes the class scores and outputs the 1-D array of size equal to the number of classes.

SEGMENTATION PROCESS

Segmentation partitions an image into distinct regions containing each pixel with similar attributes. It is useful for image analysis and interpretation, the regions should strongly relate to depicted objects or features of interest. Segmentation is the first step from low-level image processing transforming a grey scale or color image into one or more other images to high-level image description in terms of features, objects, and scenes. The success of image analysis depends on reliability of segmentation, but an accurate partitioning of an image is generally a very challenging problem. Segmentation techniques are either contextual or non-contextual. Non Contextual take no account of spatial relationships between features in an image and group pixels together on the basis of some global attribute. Image segmentation is a technique to determine the shape and size of the border. It separates the object from its background based on different features extracted from the image. After removing the noise and hair from the lesion area, the lesion needs to be separated from the skin, and therefore the analysis for diagnosis is conducted purely using the necessary area.

Thresholding

This Method determines the threshold and then the pixels are divided into groups based on that criterion. It includes bi-level and multi thresholding. Thresholding method includes Histogram and Adaptive thresholding.

Color-based segmentation

Algorithms Segmentation based on color discrimination. Include principle component transform/ spherical coordinate transform.

Discontinuity-based segmentation

Detection of lesion edges using active contours / radial search techniques / zero crossing of Laplacian of Gaussian (LoG). It covers Active contours, Radial search & LoG .Region-based segmentation It is a method of splitting the image into smaller components then merging sub images which are adjacent and similar in some sense. It includes Statistical region merging, multi scale region growing, and morphological flooding. It is based on the techniques such as Split and merges Statistical Region Merging Multi-Scale Morphological flooding

Edge Detection Technique

Edge detection is an image processing technique for finding the boundaries of objects within images. It works by detecting discontinuities in brightness. Filter is used to blur image and remove noise. It is used for image segmentation and data extraction. Ringworm is detected best using edge detection.

Gray Scale

Grayscale is a range of monochromatic shades from black to white. It contains only shades of gray and no color. Grayscale values are represented as binary values as 0's and 1's.Grayscale images are composed of pixels represented by multiple bits of information, typically range from 2 to 8 bits or more. Grayscale measures the intensity of the light reflected from an area(dot) of plane surface and defines each pixel(picture element) as a byte.

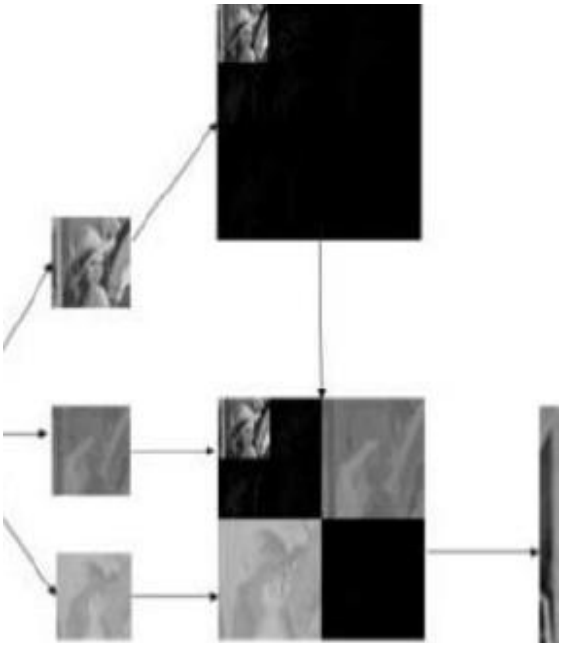


Fig.No.7. Gray Scale Conversion

RGB COLOR MODEL

The RGB color model is use a color coordinate system with three primary colors:
R(red), G(green), B(blue)

Each primary color can take an intensity value ranging from 0(lowest) to 1(highest). Mixing these three primary colors at different intensity levels produces a variety of colors. The collection of all the colors obtained by such a linear combination of red, green and blue forms the cube shaped RGB color space.

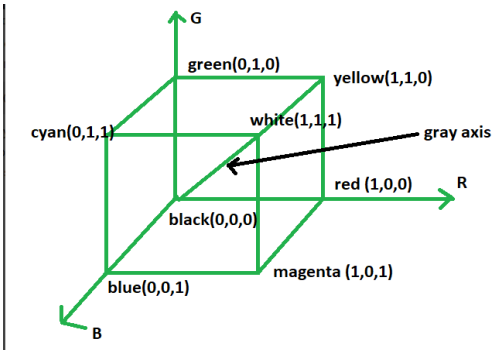


Fig No.8. RGB color model

The corner of RGB color cube that is at the origin of the coordinate system corresponds to black, whereas the corner of the cube that is diagonally opposite to the origin represents white. The diagonal line connecting black and white corresponds to all the gray colors between black and white, which is also known as gray axis.In the RGB color model, an arbitrary color within the cubic color space can be specified by its color coordinates: (r, g.b).

Example:
(0, 0, 0) for black, (1, 1, 1) for white,
(1, 1, 0) for yellow, (0.7, 0.7, 0.7) for gray

Color specification using the RGB model is an additive process. We begin with black and add on the appropriate primary components to yield a desired color. The concept RGB color model is used in Display monitor. On the other hand, there is a complementary color model known as CMY color model. The CMY color model use a subtraction process and this concept is used in the printer. In CMY model, we begin with white and take away the appropriate primary components to yield a desired color

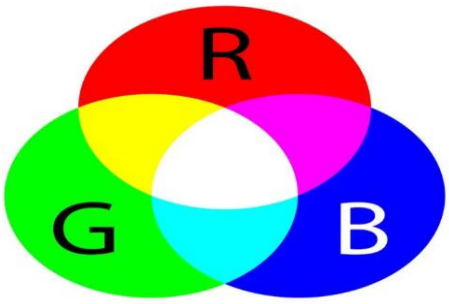


Fig.No.9.RGB COLOR MODEL

Example:

If we subtract red from white, what remains consists of green and blue which is cyan. The coordinate system of CMY model use the three primaries' complementary colors:

The corner of the CMY color cube that is at (0, 0, 0) corresponds to white, whereas the corner of the cube.

.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix} \qquad \begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

that is at (1, 1, 1) represents black. The following formulas summarize the conversion between the two color models:

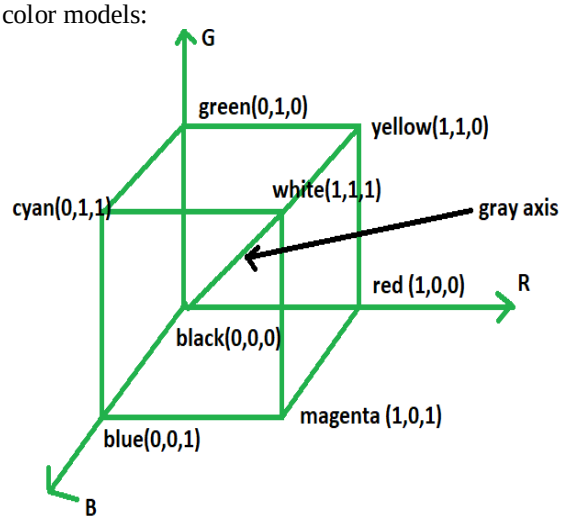


Fig.No.10. COLOUR MODEL

The corner of the CMY color cube that is at (0, 0, 0) corresponds to white, whereas the corner of the cube that is at (1, 1, 1) represents black.

OPENCV-PYTHON

Numpy is a highly optimized library for numerical operations. It gives a MATLAB-style syntax. All the OpenCV array structures are converted to-and-from Numpy arrays. So whatever

operations you can do in Numpy, you can combine it with OpenCV, which increases number of weapons in your arsenal. Besides that, several other libraries like SciPy, Matplotlib which supports Numpy can be used with this.So OpenCV-Python is an appropriate tool for fast prototyping of computer vision problems.

QT DESIGNER

Qt Designer is the Qt tool for designing and building graphical user interfaces (GUIs) with Qt Widgets. We can compose and customize our windows or dialogs in a what-you-see-is-what-you-get (WYSIWYG) manner, and test them using different styles and resolutions.

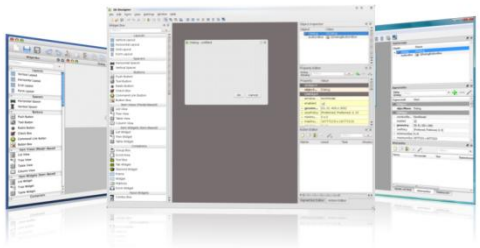


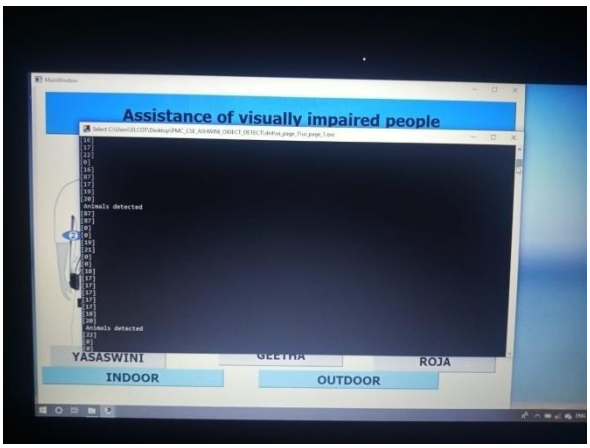
Fig.No.11. Qt Designer

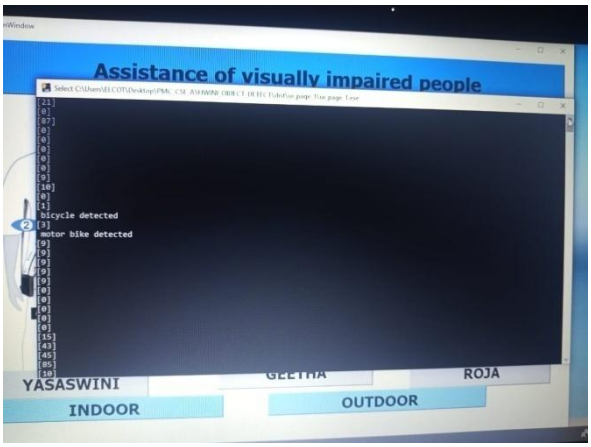
Widgets and forms created with Qt Designer integrate seamlessly with programmed code, using Qt's signals and slots mechanism, so that we can easily assign behavior to graphical elements. All properties set in Qt Designer can be changed dynamically within the code. Furthermore, features like widget promotion and custom plugins allow us to use our own components with Qt Designer.

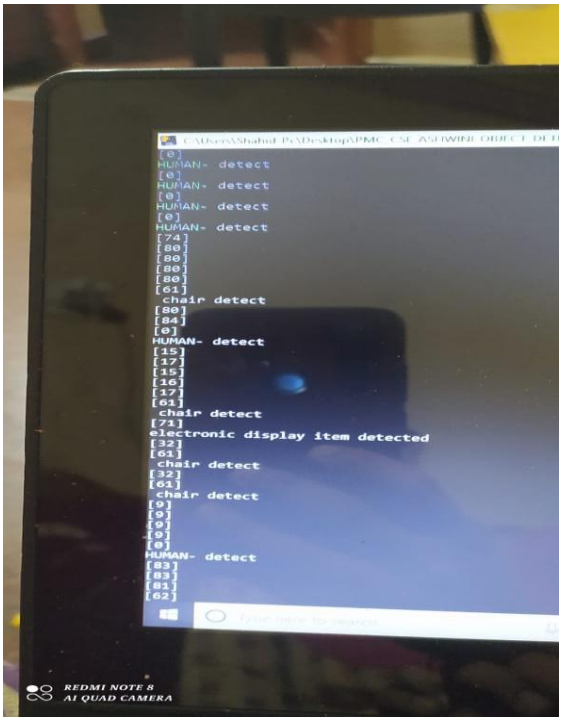
III. RESULTS

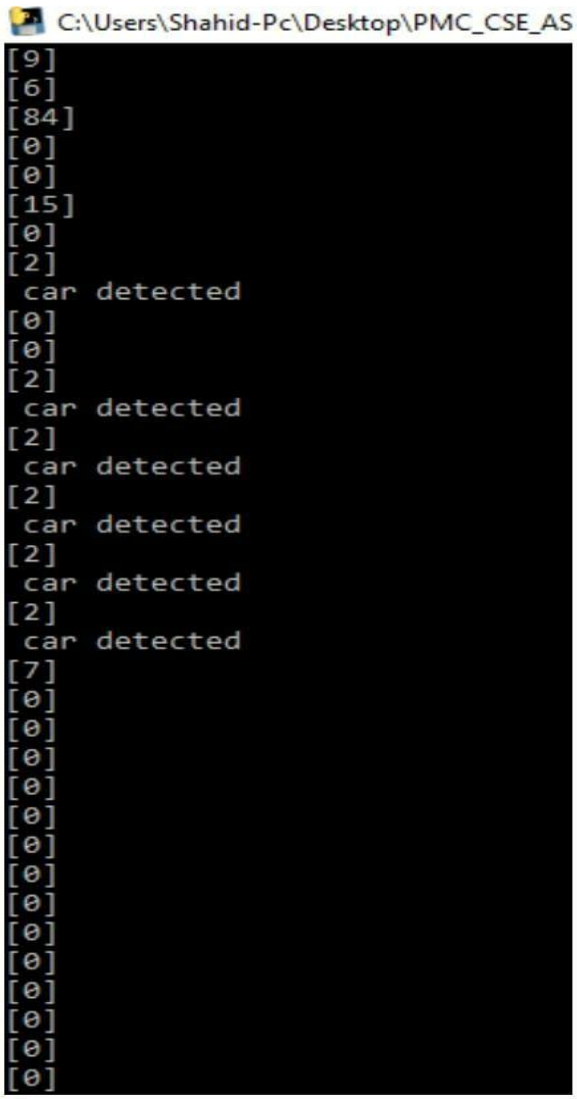
INPUT:

Input is in the form of any objects like chair, pen, any vehicles etc.









The output will also be in the form of voice through google voice.

IV. CONCLUSION & FUTURE ENHANCEMENT

The system has a novel method for object detection and identification. It can be easily commercialized and be made to benefit the visually impaired community. In our study, we used large database because of the pre-trained Convolution Neural Network model. The Single Shot Detection Mobile Net model has been trained using the COCO dataset which contains almost more 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 objects in both indoor and outdoor environment. Thus the proposed approach will help Visually impaired people in an effective way. The future enhancement is to use face recognition system to identify the person's identity and animals identity.

REFERENCES

[1]. Daniyal Rajput, Faheem Ahmed, Habib Ahmed, Engrzakir Ahmed Shaikh, Aamirshamshad, "Smart Obstacle Detector For Blind Person,"2017.

[2]. KhushbooKhurana, ReetuAwasthi, "Techniques for Object Recognition in Images and Multi-Object Detection," 2018.

[3]. Payal Panchal, Gaurav Prajapati, Savan Patel, Hinal Shah and JitendraNasriwala, "A Review on Object Detection and Tracking Methods,"2018.

[4]. Prof. SeemaUdgirkar, ShivajiSarokar, Sujit Gore, Dinesh Kakuste, SurajChaskar, 2017, "Object Detection System for Blind People," International Journal of Innovative Research in Computer and Communication Engineering.

[5]. Bin Jiang, Jiachen Yang, HoubingSong (2018)"Wearable Vision Assistance System Based on Binocular Sensors for Visually Impaired Users," DOI:10.1109/JIOT.2018.2842229, IEEE INTERNET OF THINGS JOURNAL.

[6]. Dakopoulos.D, S. K. Boddhu, and N. G. Bourbakis. A(2017)2d vibration array as an assistive device for visually impaired. pages 930–937. IEEE Computer Society.

[7]. Hsueh-Cheng Wang, Robert K. Katzschmann,SantaniTeng, Brandon Araki, Laura Giarre(2017)"Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System", IEEE International Conference on Robotics and Automation (ICRA).

[8]. T. S. T. Ifukube and C. Peng. (2016) A blind mobility aid modeled after echolocation of bats. IEEE Transactions on Biomedical Engineering, 38(5):461–465.

[9]. JinqiangBai, ShiguoLian, Zhaoxiang Liu, Kai Wang, DijunLiu(2018) "Smart Guiding Glasses for Visually ImpairedPeople I "Indoor Environment," IEEE Transactions on Consumer Electronics, Vol. 63, No. 3.

[10]. Meijer.P.B(2018) An experimental system for auditory image representations. IEEE Trans Biomed Eng, 39(2):112– 121.

[11]. Sainarayanan.G , R. Nagarajan, and S. Yaacob(2018) Fuzzy image processing scheme for autonomous navigation of human blind. Appl. Soft Comput., 7(1):257–264.