International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

Visually Impaired Aid using Deep Learning

Aditi Adwalpalkar[1], Pratiksha Sharma[2], Nitesh Redkar[3], Arpitha Naik[4], Swarali Shirodkar[5], Kedar Sawant[6], Nora Naik[7]

[1,2,3,4,5]*Department of Computer Engineering Agnel Institute of Technology and Design Assagao, Goa*
[6,7]*Assistant Professor, Department of Computer Engineering Agnel Institute of Technology and Design Assagao, Goa*

Abstract— Several projects have been implemented to help the visually impaired, however, they are most of the time restricted to the outdoor environment only, like detecting any obstacle or detecting the elevation of the ground etc. It is necessary to know that visually impaired people are dependent for the tasks that are done in the indoor environment also. In order to make them self-dependent as much as possible, the idea to create an aid in the form of a Smart Glass seemed appropriate. The Smart Glass is considered to be a potential aid for people who are visually impaired that will help them to deal with their day-to-day activities quite easily than they could actually think of. The idea is to create a novel artificial vision device, composed of a mini camera, which is strapped to the eyeglasses of the user. When a user gives an audio command to a specific module, the device translates the visual information into audio, conveying the information to the user. The device will also be able to recognize colors, currencies and objects and read texts and time as well.

Keywords: Smart glasses, Object detection, Currency recognition, Color recognition, Image to text

1. INTRODUCTION

Artificial intelligence is a field where machines can perform tasks on their own which generally require human intelligence. In Deep Learning artificial neural network algorithms, influenced by the human brain learn from large amounts of data. Similarly how humans learn from their experience Deep Learning performs a task repeatedly, each time to give more accurate results compared to the previous. Using Deep Learning, machines can solve complex problems while using a data set that is unstructured, interconnected and diverse. An organization's data mostly exists in different formats such as images, texts, pdf files etc, hence it is unstructured. Analyzing unstructured data is difficult, this is where deep learning can help. Deep Learning algorithms can be trained using distinct data formats, and still obtain insights that are relevant to the purpose of its training.

We know that life of a visually impaired person is filled with several challenges. Even though several methods have been figured out by latest technologies to ease their lives, yet many have failed to meet maximum requirements which, include day to day tasks like knowing what item

you're holding or knowing what you are reading etc. The main aim of this project is to make the user do its daily activities without depending on others. This paper will include details on how with the help of deep learning and CNNs, the smart spectacles, which is wearable technology, put on the eyes similar to our normal spectacles, will include five primaries and essential task performed by the device 1) Real-time object detection using the SSD-MobileNet object detection algorithm. 2) Text detection using OCR (Optical Character Recognition). 3) Currency detection using trained dataset. 4) Color recognition 5) Date and Time from the system. So the Visually challenged human can get all these benefits in one product itself. This paper also includes how we integrate the different hardware components used in the Smart spectacle including Raspberry pi, a NoIR camera, USB microphone, a headset and a SD card.

2. LITERATURE SURVEY

2.1 Convolutional Neural Network Based Object Detection

According to the trio Asim Suhail, Manoj Jayabalan and Vinesh Thiruchelvam who are a part of School of Computing, Asia Pacific University of Innovation & Technology, Computer vision is excelling in the field of segmentation, feature extraction, and object detection from image data. They were of the view that object detection is gaining immense interest in the fields of healthcare, traffic monitoring, surveillance, robotics etc. Their main focus was on the ability to detect the object more precisely than usual. Over the past few years, researchers have strived to cope up with this challenge. This study presents a review of the object detection approach considered using Convolutional Neural Network (CNN). CNN is applicable in all categories of object detection.[1][2][3]

2.2 Object Detection Applied To Convolutional Neural Network

In the opinion of Kai Kang Wanli Ouyang Hongsheng Li Xiaogang Wang Department of Electronic Engineering, the latest ImageNet task on object detection from video (VID) brings the object detection task into the video domain, in which objects locations at each frame are required to be annotated with bounding boxes. The evaluation protocol for the VID task is similar to that of the DET task and so they have used the conventional mean average precision on all classes as the evaluation metric.They decided to introduce a complete framework for the VID task based on still-image object detection and general object tracking together. It has the discriminative ability from the object detectors.[1][2]

2.3 Deep Learning Based Indian Currency Detection for Visually Challenged using VGG16

According to Nijil Raj N, Anandu S Ram, AneetaBinoo Joseph andShabna S of International Journal of Recent

Technology and Engineering, banknote recognition is one of the major problems faced by visually Challenged people. So they have come up with a system with almost 99.07% accuracy level. According to their project, bank notes with different positions are directly fed into VGG 16 which is supposed to be a pretrained model of convolution neural network. It is a DL algorithm which accepts an input image and assigns importance to number of objects in the image and is in a position to differentiate 5 classes of notes from one another that includes the Rs20, Rs50, Rs100, Rs200 and Rs500 notes. Then random snaps are taken out of these classes of notes Furthermore 15 images of 11 positions are considered. [4][5]

2.4 Banknote Portrait Detection Using Convolutional Neural Network

Ryutaro Kitagawa∗, Yoshihiko Mochizuki∗, Satoshi Iizuka∗, Edgar Simo-Serra∗, Hiroshi Matsuki†, Naotake Natori†, Hiroshi Ishikawa from Department of Computer Science and Engineering, Nagoya University were of the view that banknotes generally exist in different designs according to their denominations.They decided to come up with a sorting system for banknotes which could recognize portraits in each banknote and sort it accordingly. Their main aim is automating the configuration of such a sorting system by automatically detecting portraits in sample banknotes, for the purpose of their deployment in the country. They have used CNN to detect portraits in a completely new set of banknotes [20]. Here, a moving window generates candidate regions so that all possible candidates are systematically enumerated. Using CNN, the probability that the candidate region is a portrait is obtained. .[7][8][9]

2.5 Recognition of Fake Currency Note using Covolutional Neural Networks

In the opinion of Navya Krishna G, Sai Pooja G, Naga Sri Ram B, Yamini Radha V and Raja Rajeswari P of the International Journal of Innovative Technology and Exploring Engineering (IJITEE) ,designed the Automatic Fake Currency Recognition System to detect whether the currency it is fake or original. This method is conceptually based on Deep Learning. This technique can guide people and machines in identifying and determining if the currency is fake. We need to perform perspective transformations on the image to make the model data in a standard format so that the training is improved, accurate, and faster. The final result is achieved by training an artificial neural network on an image data set of currency in order to predict which class the image belongs to, and determines whether note is fake or original.

3. EXISTING SYSTEM

Several products as an aid to the visually impaired are already existing in the market but are not widely accepted due to limited features or functionality and the costly price; a price cannot be afforded by the small and middle class visually challenged people. A very well-known developed product is the OrCam. OrCam comes in two version: OrCam MyEye and OrCam MyReader. The difference between these two products is that OrCam MyEye device has three main features: Reading, Identifying Products and recognizing faces whereas OrCam MyReader is focused on reading and does not include the Product Identification or Face Recognition features. Although the former has comparatively more features than the latter, the market value for both these products are very high. Currently, OrCam MyEye is priced at $4500 and OrCam MyReader is priced at $3500. Both these products are well above the budget of any standard middleclass visually impaired person. Another system is the smart cane. The smart cane can help people avoid obstacles by providing vibrations only at ground level. Additionally, it does not provide currency, color object and text recognition. Hence provides very limited features. Another issue is that the users have to carry it everywhere they want to go. The products that have already been developed are restricted to text recognition and obstacle detection itself. The Braille Reader is only restricted to reading and writing texts. Also, these devices are too costly to buy as the price range from $3000-$15000. Some devices also include detecting objects using ultrasonic waves which get reflected on colliding with the obstacles, but with the advancement in learning, we can use deep neural networks to obtain more accurate and efficient results. The proposed design of the smart glasses in this paper facilitates multiple features all combined under a single platform including real-time obstacle avoidance by object detection, text detection in the vicinity, color detection and currency detection for all types of visually challenged people. The main advantage of such smart glasses will be that the user will be able to do its routine activities without being dependent on others and they can avail this device at a very cost friendly price.[5][6][8]

4. PROPOSED SYSTEM

Amidst the entire population, we find ourselves independently doing our day-to-day tasks, which implies we are self-sufficient. But if we decide to conduct a survey, we encounter that approximately 45% of the world's population comprises visually challenged people.

So, to eradicate this problem to a certain extent our team has decided to come up with smart glasses for the visually impaired using the power of Deep Learning, which are much more flexible in terms of the features it inculcates and much affordable than the pre-existing ones devices. The device will also be able to recognize colors, currencies, text and detect objects.

4.1 Overview

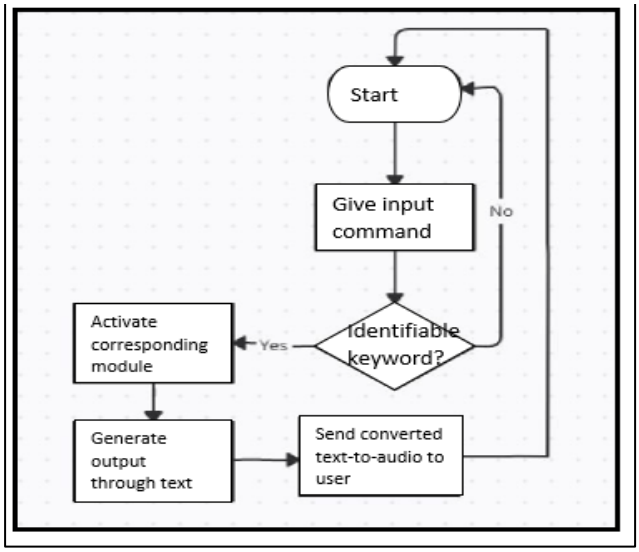

Fig. 1. Entire Process

The glasses will function as follows:

1. The camera attached to the glasses along with its system core will be in the 'ACTIVE' mode at the time of use.
2. The input command given by the user, will be recorded by the microphone. If the particular keyword falls within the predefined words fed to the system go-to step 3, else go-to step 8.
3. Upon activation of the corresponding module, the raw image captured from the camera is passed to the blur detection frame in order to obtain a clean image.
4. The clean image is then passed as an input to the algorithm.
   - If the module called upon is 'READING_TEXT' or 'COLOR_RECOGNITION', go-to step 5, else,
   - If the module called upon is 'CURRENCY_RECOGNITION' or 'OBJECT_RECOGNITION', go-to step 6, else,
   - Go-to step 7, for the 'TIME_MODULE'.
5. The clean image is passed on to the triggered algorithm where output in the form of a string is produced. This string output is then passed to the audio module for text-to-audio conversion.
6. The clean image is either passed to the MobileSSD model for 'OBJECT_RECOGNITION' or to the custom trained model for 'CURRENCY_RECOGNITION'. The image will then be tested with its train set to return the validated output.
7. The time from the system will be directly passed to the audio module.
8. Since the input command doesn't contain any of the predefined keywords, an audio message will be passed to indicate the same.

4.2 Overall Look

Fig. 2. Overall Look

In the proposed design of the smart glasses, there will be np buttons or switches. To perform any function the user gives commands in the form of audio speech via Speech-to-Text module. While receiving the output the sentence in the form of text gets converted into speech by Text-to-Speech module. Below is the description of Speech-to-Text and Text-to-Speech module.

1. Speech-to-Text

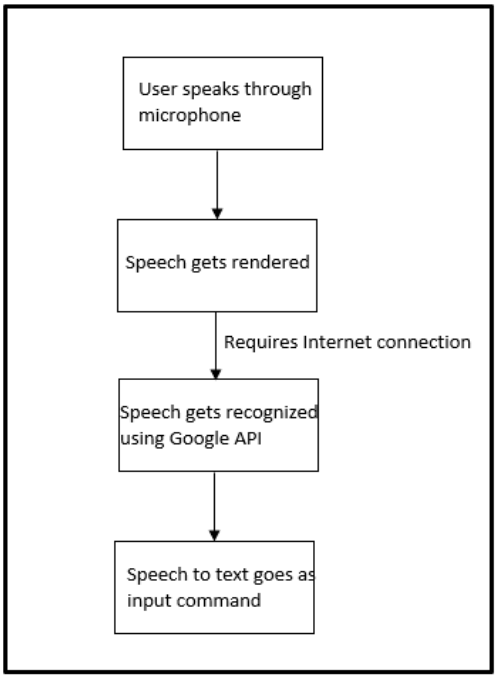

Fig. 3. Speech-to-Text

2. Text-to-Speech

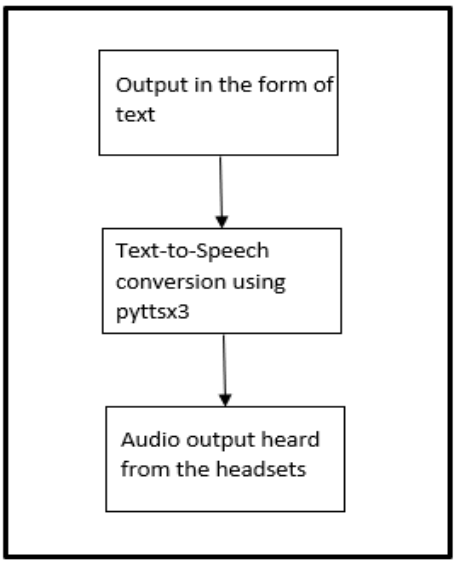

Fig. 4. Text-to--Speech

A. Text detection

If a visually impaired person wants to read a book, they have to be dependent on others to do the task. It would be great if they can do it on their own. One way they can read is using Braille but it is neither widely found in public places nor are they largely available in form of books, Braille books even cost more than normal books. So to overcome with this obstacle the user just has to keep what they want to read in front of the Smart glasses, the glasses will scan it and inform the user what is written on it.

Here the use of python library is more reliable and it also cuts down the development by a big margin. The python library in use is TesseractOCR. TesseractOCR is an Optical Character Recognition tool for python which can be used to read text in images. pytesseract is a wrapper for Google's Tesseract-OCR Engine.

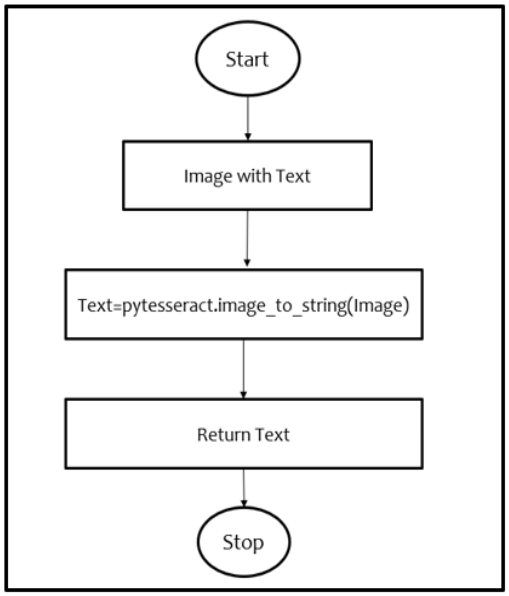

Fig. 5. Text detection

B. Object recognition

Using Deep learning object detection algorithm, we have Faster R-CNN‟s, SSD (Single Shot Detectors), YOLO (You Only Look Once. MobileNets network architecture can be used inside the object detection pipeline but these heavy architectures are unsuitable for devices like Raspberry Pi. Hence we combine MobileNet architecture and SSD framework to arrive at a very fast and efficient real-time object detection algorithm. The SSD predicts different scale feature maps to gain high accuracy on predictions made by it and this also improves its capabilities of detecting different size objects. The object recognizer that we will be using is the large Coco 2020 model that has been trained to recognize 91 different objects. The recognizer can recognize multiple objects at same time with high accuracy. Large coco 2020 is pretrained and can be deployed.

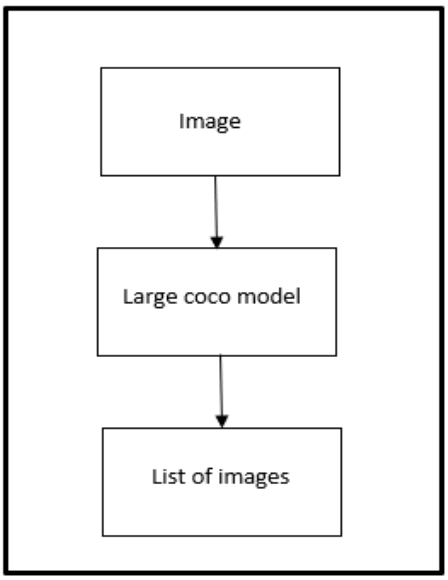

Fig. 6. Object detection

C. Color Recognition

Color Recognition features in our smart glasses will help the user to identify the color. There are times where the user have lost his/her eyesight due to an accident. In such cases, unlike the people who are blind by birth, the user actually knows how colors look like. This module is an advantage to such type of users. Too execute this module, we will use the K-Means algorithm. K-means is a clustering algorithm used in Machine Learning where a set of data points are to be categorized to 'k' groups. It works on simple distance calculation.

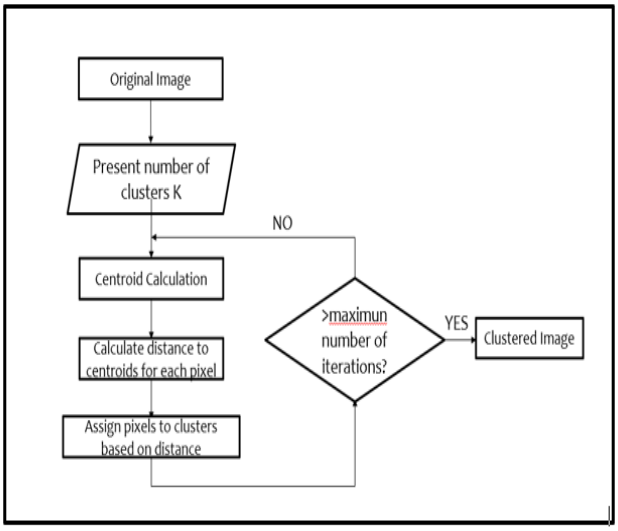

Fig. 7. Color recognition

D. Currency Recognition

Here in the currency recognition model, the Smart glasses will scan the currency and tell the user which note one is holding. The user just needs to hold it in front of the camera of the Smart glass and he will be informed about the amount.

Model selection is one of the starting steps of developing any image recognition model which will recognize the image you want. There are multiple neural network models such as LSTM, CNN, GAN, RNN and ANN etc. CNN is the most ideal when it comes to dealing with image recognition problems.

CNN has multiple implementations, each implementation differs based on accuracy of the model. High accuracy requires higher processing power. Preferred implementation for the project is Mobile Net as Raspberry pi is the hardware

Mobile Net is a lightweight architecture of CNN. It uses depth-wise separable convolutions which basically means it performs a single convolution on each color channel rather than combining all three and flattening it. This has the effect of filtering the input channels.

Building a neural network for image classification is not always easy. Transfer learning is one of the practices used where we take a pre-trained model trained on the ImageNet dataset and use those weights as a starting point for a different data set. This is done by replacing the last fully-connected layer and training the model while only updating the weights of the linear layers and letting the convolutional layers keep their weights. This cuts down the time taken building the model from scratch.

By analyzing the training dataset, more accuracy can be obtained and improved according to our expectations. In this feature, there would be an adequate number of images in the dataset which can be used for training purpose. Initially, multiple images are captured until the appropriate image is found for further processing. The software interface that we are proposing could be used only for Indian currency.

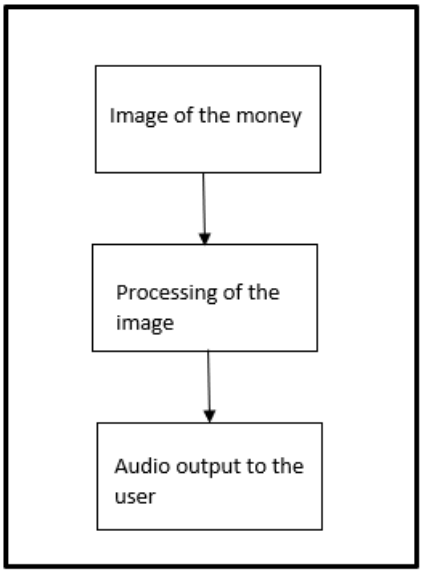

Fig. 8. Currency Recognition

E. Date and Time

Using the python datetime module comes in handy as we just need to import the library. When the time library module is called it extracts the time from the system and returns a string value.

```
datetime.datetime.now()
```

This returns a string value which can later be converted to voice output.

5. IMPLEMENTTION AND RESULTS

The images below are the screenshots from the output we receive.

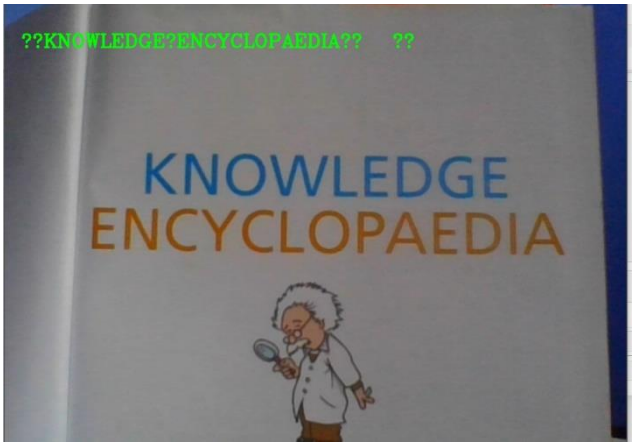

Fig. 9. Text Detection

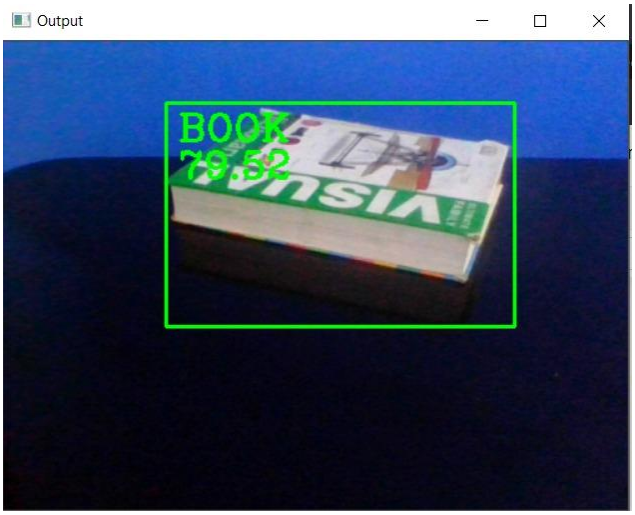

Fig. 10. a single object on display

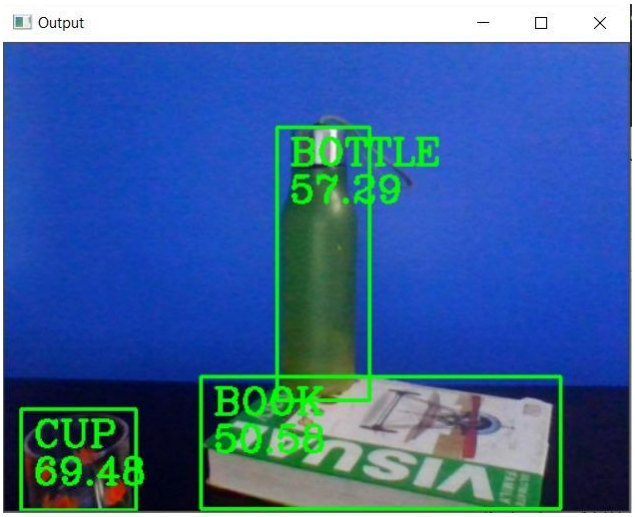

Fig. 11. multiple objects on display

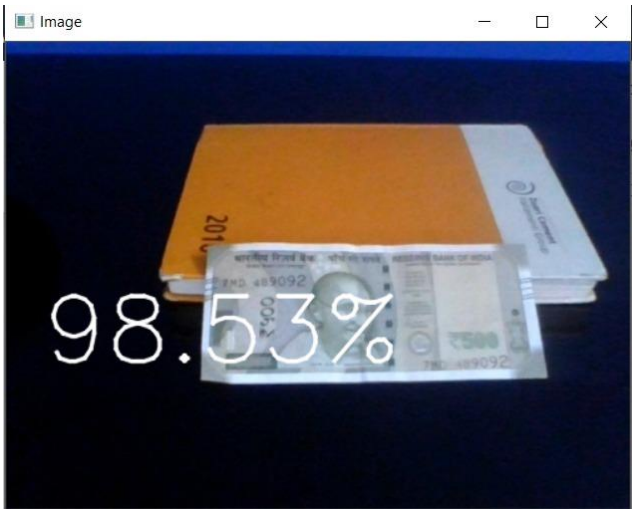

Fig. 12. Currency Recognition(Rs 500 Note)

Fig. 13. Currency Recognition(Rs 200 Note)

Fig. 14. Currency Recognition(Rs 200 note Folded)

6. Conclusions

The project proposed the design of Smart Glasses that would assist the visually impaired with their daily routines. Although several projects have been implemented as an aid to the visually impaired, they are usually restricted to the outdoor environment. The Smart Glasses helps performing indoor as well as outdoor tasks such as reading text, color recognition, object detection, currency recognition and informing the user about the current time.

It can further be improved to carry out tasks such as recognizing humans, sign boards and act as a tool. The main aim of this project revolves around solving the difficulties faced by the blind people in their daily life.

7. References

1) Smart Guiding Glasses for Visually Impaired People in Indoor Environment Jinqiang Bai, Shiguo Lian, Member, IEEE, Zhaoxiang Liu, Kai Wang, Dijun Liu.

2) Kai Kang Wanli Ouyang Hongsheng Li Xiaogang Wang Department of Electronic Engineering, The Chinese University of Hong Kong

3) Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. NIPS, pages 1097–1105, 2012

4) Deep Learning Based Indian Currency Detection for Visually Challenged using VGG16 Nijil Raj N, Anandu S Ram, AneetaBinoo Joseph, Shabna S International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-2, July 2020

5) Gouri Sanjay Tele, Akshay Prakash Kathalkar, Sneha Mahakalkar, Bharat Sahoo, and Vaishnavi Dhamane. Detection of fake indian currency. International Journal of Advance Research, Ideas and Innovations in Technology, 4

6) Qian Zhang, Wei Qi Yan, and Mohan Kankanhalli. Overview of currency recognition using deep learning. Journal of Banking and Financial Technology, 3(1):59–69, 2019.

7) Snigdha Kamal, Simarpreet Singh Chawla, Nidhi Goel, and Balasubra- manian Raman. Feature extraction and identification of indian currency notes. In 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pages 1–4. IEEE,

8) CONVOLUTIONAL NEURAL NETWORK BASED OBJECT DETECTION: A REVIEW Asim Suhail1, Manoj Jayabalan2, Vinesh

9) Thiruchelvam3 1 School of Computing, Asia Pacific University of Innovation & Technology, Kuala Lumpur, Malaysia

10) https://www.pyimagesearch.com/2017/07/10/using-tesseract-ocr-python/

11) https://buildmedia.readthedocs.org/media/pdf/pyttsx3/latest/pyttsx3.pdf