TELL ME (AN ASSISTANT FOR THE VISUALLY CHALLENGED)

Moeez Ur Rehman
BS Scholar (Computer Engineering)
Sir Syed University of Engineering & Technology
moeezurrehman0@gmail.com [1]

Muhammad Huzaifa
BS Scholar (Computer Engineering)
Sir Syed University of Engineering & Technology
huzaifamuhammad1994@gmail.com [2]

Qateba Moin Shahzaib
BS Scholar (Computer Engineering)
Sir Syed University of Engineering & Technology
Qatebamoinshahzaib@gmail.com [3]

Syed Babar Kamal
BS Scholar (Computer Engineering)
Sir Syed University of Engineering & Technology
babar_judean1994@hotmail.com [4]

Dur-e-Shawar Agha
Lecturer (Computer Engineering)
Sir Syed University of Engineering & Technology
engr.dureshawaragha@gmail.com [5]

Abstract: It is anticipated that around 258 million people are visually impaired worldwide. Majority of these people are the citizens of developing countries. The difficulties faced by the visually impaired persons are the recognition of objects, colors and shapes. The computational power of current mobiles gives inspiration and motivation to developed applications that can assist the visually impaired persons. The proposed mobile application deals with the difficulties faced by the visually impaired persons by recognize objects, colors, shapes and the Tell Me app will convert the object's name, color or shape into audio speech. For the implementation of the system Ionic and Django Rest framework is used. The inception architecture which uses deep convolutional neural network is used for classifying objects or inputs. The system was tested on dozen of trained objects. Experiment results show the accuracy and promising object detections.

Keywords: Object Recognition, Django Rest framework, Ionic framework, Convolution neural network

I. INTRODUCTION

One hundred years ago, being visually impaired meant being condemned to a life of confinement and institutionalization. It was almost unheard of that visually impaired or challenged people undertake steps to become educated and employed. Often seen as helpless by society and as a burden by the family, they could only dream of having a family and living an independent life.

With the advancement of technologies and better understanding of the human sciences, it is possible to create such a system which would be able to help these challenged groups of people

to become a part of the society as if they had no disabilities at all. Many mobile applications are there which help visually impaired people to recognize objects, color, face or even the path.

II. LITERATURE REVIEW

According to June 2012 approximately [1], 285 million people are visually impaired worldwide: 39 million are blind and 246 million have low vision [2] according to the World Health Organization. About 90% of the worlds visually weaken people live in developing countries. With the advancement of technologies and better understanding of the human sciences, it is possible to create such a system which would be able to help this challenged group of people to become a part of the society as if they had no disabilities at all. Among all the areas of computer sciences, we are particularly interested in the techniques which can be used for supporting blind or visually challenged.

Alexander Dreyer Johnsen et al. [3] anticipated a Touch-based mobile system for the visually weaken people. The researchers introduced the idea of using finger touch mobile devices to help the visually impaired. Their system is a combination of sound and haptics. It depends upon haptic and voice feedback, where the user receives information when tapping or dragging in finger across the system.

K. Matusiak et al. [4] proposed a system which recognizes the objects in mobile phone. In this work, they introduced a system for Android devices for the blind users. It can recognize objects by scanning them and then matching with its database of objects. It can also recognize colors and locate maximum brightness regions in the captured scenes.

Chaudhry, Shonal, and Rohitash Chandra [5] proposed the mobile based system for the face detection using mobile computing for the people who are visually impaired. The mobile system was assisted by a server based support system and used computer vision techniques for face detection. This system allows device to process several video frames in real time. The system also maintains application response while doing frame processing in real time.

Thakoor, Kaveri A., et al. [6] developed a naturally inspired computer vision algorithm AB-SURF (Attention Biased Speeded Up Robust Features) for object recognition. This algorithm consist of two parts, the Attention Biasing select the major task driven prominent region in an image while SURF algorithm recognized the object by the narrowed subsection of the image selected by AB algorithm.

Patil, Mrunmayee, and Ramesh Kagalkar [7] proposed a technique to convert images into text and speech using different image processing techniques like image segmentation, object detection, edge detection and speech synthesis. For object detection from images Canny Edge detection algorithm was used while object recognition is done on the basis of texture, shape and size of the object.

Mascetti, Sergio, et al.[8] developed an mobile application that can detect the color of an object. The user has to bear a marker to point the object for color recognition. Poggi, Matteo, and

Stefano Mattoccia [9] developed a wearable aid for visually impaired persons based on machine learning techniques. Their system used deep learning techniques to detect obstacles. The system allows real time obstacle detection on a computer by processed the cloud provided by RGBD sensor which is based on passive stereo vision. It can also work with smart phone, connected wirelessly to the mobility aid.

III. OVERVIEW OF THE PROJECT

Identification of objects for the visually impaired persons is really a difficult task. With the computational power of present mobile phones many applications are developed to assist these people. The main goal of our proposed system is to identify objects and convert it into audio speech. In order to achieve a system which should be simple but powerful, various combinations can be exercised. Firstly, consider the computational power of the present mobile devices. It is possible to achieve inference in mobile devices but it is not considered prime, although as the devices become more powerful by time this thought will be changed with it. So, for the time being it is better to let the server side compute for inference rather than device itself and as most devices would not have the computational power needed.

A. System Functions

On the user side the user will only have to aim his back-camera at the object they want to be recognized, considering the Tell Me app has been launched using any app for voice search or tap launching it the conventional way. The Tell Me app will:

1. Take a picture of the object being aimed.
2. Convert it into a base64 string.
3. Create a JSON object which contains several entities like image base64 string, authentication token etc.
4. An Http request is issued which sends the JSON object to the API endpoint of the server.

The server side is a REST framework which will use this JSON object for authentication of the user and then classification of the requested image string:

1. The data is extracted from the JSON object.
2. The token, from JSON object, is used for authentication of the user.
3. If the user is authenticated, then the image base64 string is converted back to an image and then passed into the tensorflow script.
4. The tensorflow script calls the pre-trained data along with the labels of the objects and parses both.
5. The image is feed into the pre-trained model, which contains the softmax function that maps the input data into probabilities into an expected output.
6. The tensorflow script returns the label with the greatest probability.
7. The result is sent to the user as a Http Response.

The result is now displayed onto the device's screen and the Tell Me app will convert the object's name into audio speech.

B. Technical Implementation

The system is achieved with the combination of Ionic framework as the user end and the Django Rest Framework as the server, used for managing the server end-point and classification of images via tensorflow scripts.

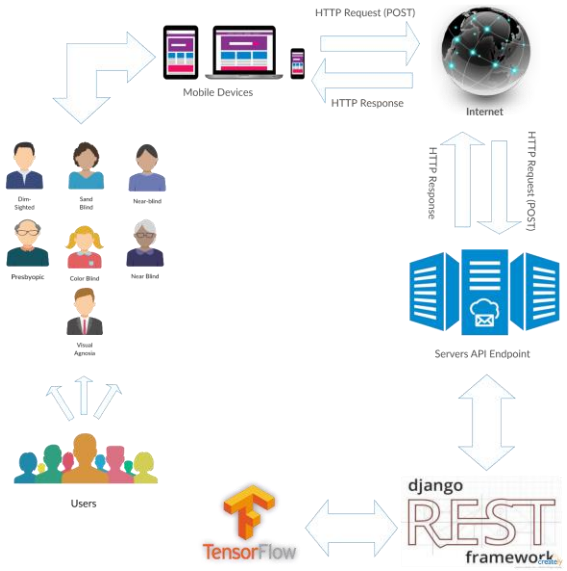

Figure 1: The overall system view of Tell Me.

C. Ionic Framework

The ionic framework [10] is a hybrid app development framework with the motto of "write once, run everywhere". The latest version of ionic is compatible with Angular 4, Typescript 2.1 and Typescript 2.2. This framework enables the developers to create powerful apps rapidly which are able to run on various platforms, such as Android and IOS, with fairly good compatibility. The user end is created with the Ionic framework latest version i.e. ionic3.

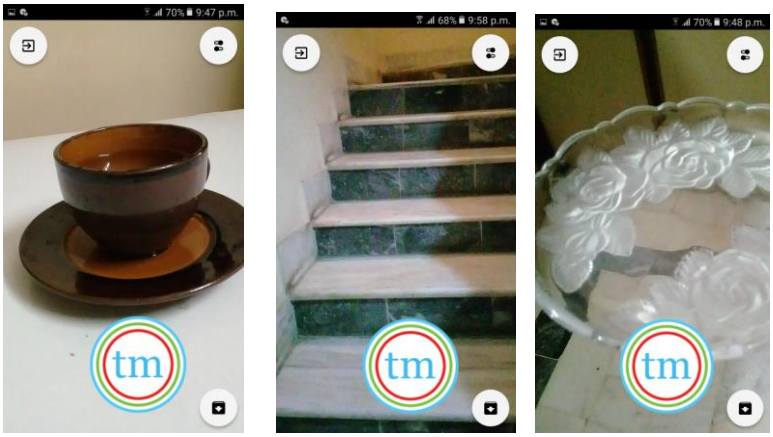

Figure 2: Audio speech generated for various objects.

D. Django Rest Framework

The Django Rest framework or simply DRF [11] is a powerful toolkit for developing web APIs, which is built on top of Django framework. The Django framework [12] is a Python based server side programming framework for developers with the need for rapid development but with powerful design.

E. Tensorflow

Tensorflow [13] is an open source machine learning library based on Python. It can be used to create, train and test neural networks of various kinds such as CNNs, RNNs etc. it was developed to be used for Google by its "Google Brain" team. The version 1.0.0 was released in February 11 2017.

F. Inception Architecture

The inception architecture [14] is a deep convolutional neural network used for classifying objects or inputs. It is important to mention this model because the DCNN used in Tell Me is the Inception architecture. It at least uses 1x1, 3x3 and 5x5 convolutions, max pooling and filter concatenation per each module.

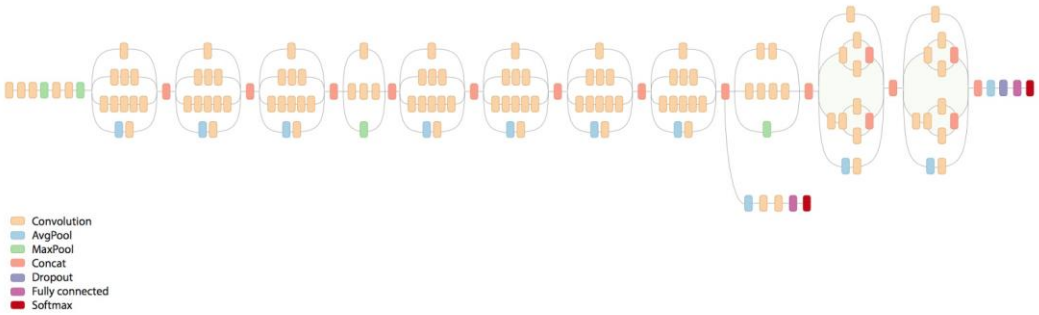

Figure 3: The overview of Inception Architecture

IV. FLOW OF THE PROPOSED SYSTEM

The following diagram represents the flow of the overall system. The app is launched and selects the category for object recognition. The images are automatically taken at regular intervals, create its JSON object and send as HTTP request to API Endpoint. The data is send to TF script which recognizes the object and classifies it and return it as HTTP Response. The output is generated as audio speech of returned label.

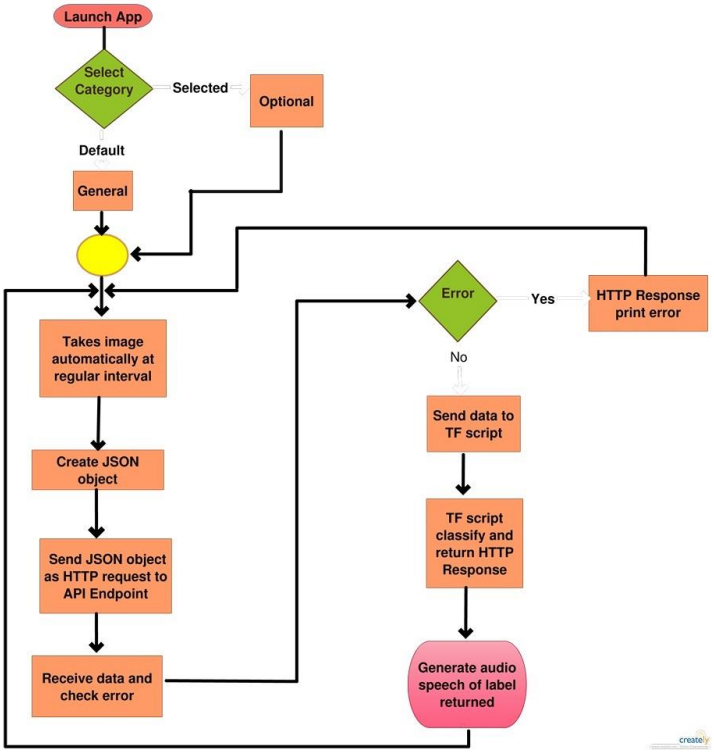

Figure 4: Flow diagram of overall system

V. FEATURES

The app is for android devices, human friendly and will prove to be very useful for the visually challenged. The main objective is to facilitate the low vision users, to provide them the convenience to walk among people with independence and prosperity. Tell Me allows users to understand their surrounding objects.

This goal has become achievable through agile classification by implementing Deep Convolutional Neural Networks. The data trained through DCNN is input for tensorflow scripts being used by the Django Rest Framework, hence a restful API service is established which communicates with the user through Http protocols.

Tell Me is able to recognize a variety of categories of objects. The category section is for the users that are able to see to some extent, such as color blinds. The selection of categories actually results in an even faster inference outputs due to lesser number of objects within a category, comparative to the "General" category containing every object trained within the system.

The categories include:

1. General
2. Basic objects
3. Colors
4. Animals

5. Fruits

Each of these categories includes dozens of trained objects.

VI. COMPONENTS USED FOR SYSEM GENERATION

1. Ionic 3
2. Django Rest Framework (DRF)
3. Tensorflow
4. Google Cloud Platform
5. Windows 10
6. Ubunto
7. Microsoft Visual Studio Code
8. Sublime Text
9. Laptops
10. Mobile Devices

VII. EXPERIMENTS AND RESULTS

The captured image/object are send at time interval (t) and received after some time interval having some delay and the result comes with accuracy.

Local Host Server				
Object	Request sent(t)	Response time	delay	Accuracy
Bag	03:06:00 03:07:15	03:06:06 03:07:21	6sec 6sec	0.65784 0.74321
Chair	03:12:19 03:15:02	03:12:26 03:15:08	7sec 6sec	0.75677 0.78123
Dog	05:04:12 05:10:12	05:04:19 05:10:18	7sec 6sec	0.61344 0.69623
Cat	06:01:20 06:05:23	06:01:26 06:05:29	6sec 6sec	0.65211 0.66875
Black	06:30:34 06:40:23	06:30:40 06:40:28	6sec 5sec	0.64444 0.67832
White	07:02:10 07:15:10	07:02:15 07:15:16	5sec 6sec	0.67892 0.69754
Banana	07:07:00 07:15:03	07:07:06 07:15:08	6sec 5 sec	0.58901 0.59021
Watermelon	07:20:24 07:27:24	07:20:30 07:27:31	6sec 7sec	0.60012 0.59104

Table 1: Local Host experiment values

Amazon Web Service				
Object	Request sent(t)	Response time	delay	Accuracy
Bag	03:06:00 03:07:15	03:06:04 03:07:20	4sec 5sec	0.75784 0.74321
Chair	03:12:19 03:15:02	03:12:23 03:15:08	4sec 6sec	0.79677 0.78123
Dog	05:04:12 05:10:12	05:04:16 05:10:18	4sec 6sec	0.81344 0.79623
Cat	06:01:20 06:05:23	06:01:26 06:05:26	6sec 3sec	0.75211 0.77875
Black	06:30:34 06:40:23	06:30:37 06:40:28	4sec 4sec	076532 0.75632
White	07:02:10 07:15:10	07:02:15 07:15:17	5sec 7sec	0.67892 0.69754
Banana	07:07:00 07:15:03	07:07:03 07:15:07	3sec 4sec	0.58901 0.59021
Watermelon	07:20:24 07:27:24	07:20:28 07:27:29	4sec 5sec	0.60012 0.60104

Table 2: Amazon web service experimental results

VIII. CONCLUSION

Our goal was to develop an app for visually impaired people that can recognize objects in different categories. . For this purpose Ionic and Django Rest framework is used. The inception architecture which uses deep convolutional neural network is used for classifying objects or inputs. The app is able to detect the objects quickly and accurately and developed results in the form of audio.

Acknowledgement

We would like to thank the Computer Engineering Department, Sir Syed University of Engineering and Technology, Karachi Pakistan for all the key guidance, support and constant encouragement towards our project. Also, we would like to thank Final year Project committee for providing us necessary equipment and facilities required for our project progress and development.

REFRENCES

[1] "10 FACTS ABOUT BLINDNESS AND VISUAL IMPAIRMENT" FACT FILE. WHO, 10 july 2017

[2] "*World Report On Vision*" Prevention of Blindness and Visual Impairment. WHO, 10 july 2017

[3] Alexander Dreyer Johnsen, Tor-Morten Grønli, Bendik Bygstad*," Making Touch-Based Mobile Phones accessible for the Visually Impaired*", Norwegian School of IT, Publication: Norsk informatikkonferanse, ISBN: 9788232100132, 2012.

[4] Shao, Ling, et al., eds. Computer vision and machine learning with RGB-D sensors. Heidelberg: Springer, 2014.

[5] Chaudhry, Shonal, and Rohitash Chandra. "Design of a mobile face recognition system for visually impaired persons." *arXiv preprint arXiv:1502.00756* (2015).

[6] Thakoor, Kaveri A., et al. "Attention biased speeded up robust features (ab-surf): A neurally-inspired object recognition algorithm for a wearable aid for the visually-impaired." *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*. IEEE, 2013.

[7] Patil, Mrunmayee, and Ramesh Kagalkar. "An Automatic Approach for Translating Simple Images into Text Descriptions and Speech for Visually Impaired People." *International Journal of Computer Applications* 118.3 (2015).

[8] Mascetti, Sergio, et al. "Towards a natural user interface to support people with visual impairments in detecting colors." *International Conference on Computers Helping People with Special Needs*. Springer International Publishing, 2016.

[9] Poggi, Matteo, and Stefano Mattoccia. "A wearable mobility aid for the visually impaired based on embedded 3d vision and deep learning." *Computers and Communication (ISCC), 2016 IEEE Symposium on*. IEEE, 2016.

[10] Carney, B. (2017 April 7). *Ionic 3.0 has Arrived!* . Retrieved from http://blog.ionic.io/ionic-3-0-has-arrived/

[11] Christie, T. (2011*). Documentation for Django Rest Framework version 3*. Retrieved from http://www.django-rest-framework.org/

[12]*Documentation for Django framework*. Retrieved from https://www.djangoproject.com/start/overview/

[13]*Credits.* Retrieved from https://www.tensorflow.org (2017 September 28).

[14]*Credits*. Retrieved from https://www.tensorflow.org/tutorials/image_recognition June 19, 2017.

	Mooez ur Rehman is a Final Year B.S. student of Sir Syed University of Engineering and Technology in Computer Engineering Department.
	Muhammad Huzaifa is a Final Year B.S. student of Sir Syed University of Engineering and Technology in Computer Engineering Department.
	Qateba Moin Shahzaib is a Final Year B.S. student of Sir Syed University of Engineering and Technology in Computer Engineering Department.
	Syed Baber Kamal is a Final Year B.S. student of Sir Syed University of Engineering and Technology in Computer Engineering Department.
	Engr. Dur-e-Shawar Agha working as a Lecturer in the Department of Computer Engineering of Sir Syed University of Engineering and Technology, Karachi, Pakistan. She did her BS in Computer Engineering and MS in Computer Engineering (Specialization in Computer Networks) from Sir Syed University of Engineering and Technology, Karachi,

	Pakistan. Her research interests are Cryptography, Artificial Intelligence and Wireless Sensor Network.