

OCR Oriented Reading System for Blind People

Pooja Shree H R

Dept. of Computer Science & Engineering, DSU,
Kudlu Gate, Bangalore.
poojashree.hr0204@gmail.com

Dr. Revathi

Dept. of Computer Science & Engineering, DSU,
Kudlu Gate, Bangalore.
revathi-cse@dsu.edu.in

Abstract - The current engendering of the digital phones has prominent hardware endowment and a speedier operation that is robust and ample to evolve applications that facilitate the user to be a part of and act with the Universe at their own convenience. this method is an OCR scanning system that uses a camera application present in your smartphone consolidated with Optical Character Recognition (OCR). OCR may well be a mechanism that converts pictures of manuscripts or printed content into machine-encoded text. During this system, the application will scan and convert the content that is written in English (West Germanic language) accent into a speech format. By victimization of the Text to Speech Module speech output is generated. The seek of coming up with the type of voice/speech is to give out datum that is bestowed in the given image to the visually impaired people.

Keywords: *OCR, TTS, CNN, Optical Character, Azure API*

I. INTRODUCTION

Digital pictures are becoming terribly modern rapidly. each day, several images are generated by many teams like students, engineers, and doctors, in step with their variable needs. They'll access images and build their primitive options or associated text. Text present in such images will return up with purposeful information. we tend to style to fetch the content and description the visual information impromptu from images. associate optical character recognition system that involves many algorithms are needed for this purpose. Tesseract is presently the foremost correct optical character recognition engine that was developed by power unit Labs and is currently in hand by Google. We propose a system in which the input image is read technologically and it is converted into speech output in order to assist blind people in getting vital datum. Our proposed application is prepared to recognize the text and produce the speech of the translated text or content that is entrapped by our portable camera and it is going to display the same interpretation result back onto the screen of the mobile phone. In this project, the ML model makes use of OCR (Optical Character Recognition) to read and acknowledge the pre-processed content that is visible within the given input image. As we know this project is proposed mainly for blind people therefore the ultimate outcome will be an audible speech that will read out loudly the content admitted by the OCR module, and Text-To-Speech (TTS) module is getting used for delivering this voice output. This proposed application system will read every special character like comma, exclamation, full stop, the question mark is taken into consideration, also it gives right pause whenever any of those is encountered and it read anything that is acknowledged by any numerical integer value and the West Germanic.

II. LITERATURE SURVEY

In [1] uses each CNN and LSTM deep learning classifiers for image classification and text extraction. A

pitfall isn't any program is intended. In [3] uses the beano TU app to scan and skim aloud the lottery variety with high accuracy. Debit is it works solely on lottery tickets. In [4] robot Studio is employed to develop the appliance but, minus purpose isn't any cloud deployment. In [5] totally different digitization ways are used these digitized pictures are pre-processed piecemeal. The drawback is that scanning degrades the quality of the document and thus the accuracy and this OCR system are very dependable on the quality and nature of the text data. In [7] The MATLAB technique is employed to convert text to speech and thus the minus purpose isn't any cloud service is provided. In [8] used here, for segmentation of text-line and words are 2-stage algorithm. the primary stage can logically be analyzing fuzzy text-line and word boundaries victimization DC co-efficient and at intervals, the ordinal stage, AC co-efficient of selected DCT blocks are acquainted with extract actual line and word boundaries but, the matter is paper worked on solely JPEG compressed written text documents. In [14] Scan.it is a web application applied to beat the language barriers among states and code is enforced victimization Node.js however, the inconvenience is that temperament is detected for expanse and thus equivalent letters are detected otherwise this paper presents future work for cloud deployment. In [16] used here are K-Means agglomeration for background separation and SIFT technique to want out key points of the merchandise but, the matter is it acknowledges solely encompassing objects but not the characters.

Common Drawback of Existing System

No cloud deployment and it recognise only text present in the image but not the image.

III. PROPOSED SYSTEM

Mobile gadgets are getting tremendously popular, especially smartphones. Because of these analysts are inventing copious requisitions (applications) for the clients i.e. users which might be used on Smartphones. Our proposed system is going to be at one's fingertips to clients in the configuration of Android applications that will be present in the Application Store in the downloadable format. Thus, developing smartphone-based requisitions within some clicks and triggering an inbuilt camera application system present within one's phone aid blind people to read the indispensable information. Since blind people can't readily able to capture an ideal picture of the image, the pre-processing task of the clicked image is going to be done automatically as all features are going to be provided by Optical Character. The ML model makes use of OCR (Optical Character Recognition) to read and acknowledge the pre-processed content that is visible within the given input image. As we know this project is proposed mainly for blind people therefore the ultimate outcome will be an audible speech that will read out loudly the content admitted by the OCR module, and Text-To-Speech (TTS) module is



getting used for delivering this voice output. This proposed application system will read every special character like comma, exclamation, full stop, the question mark is taken into consideration, also it gives right pause whenever any of those is encountered and it read anything that is acknowledged by any numerical integer value and the West Germanic.

System Architecture

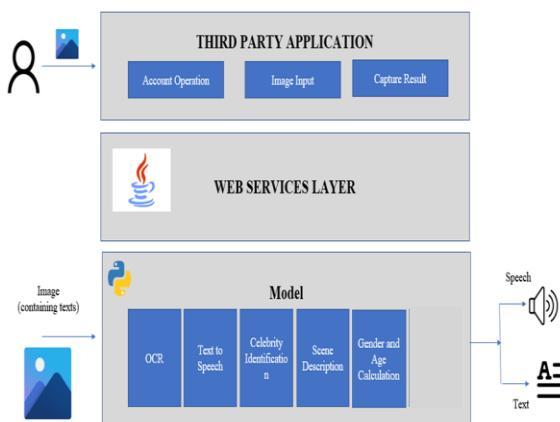


Fig. 1. Architecture

A. Third Party Application (TPA)/ UPO

Here, the client can dedicate various activities to their profiles. First and foremost, the client can enrol a brand-new account and thus get admittance to the portal. Then the client can log on to their accounts using the enrolled email ID and password to way in various other dissections within the portal.

The client can then opt to delete their accounts just in case they want to know not to access our portal and can also better update their profile by handing over the new values to the fields that are provided during the registration phase, or the client can alter their password by feeding earlier password and brand-new password.

Model Configuration: Here the client must configure the model by providing the context root of the model where it's being deployed, hostname, application name, and also the port number.

Model Input: Here, the user will need to provide an input image for the model. The input is usually accepted within the variety of a hyperlink that points to the image which can be went to execute the OCR algorithm.

Model Result: Here, the users are presented with a view that contains the results of the model. This view also contains a hyperlink to download the MP3 for the Speech output.

B. Web Service Layer

By exploiting the web service layer, we tend to be aiming to expose solutions to the surface world. that anybody around the globe from any location will invoke our web service layer by exporting pictures as input and the web service layer can later forward that exact image to the model has input get the result and provides the response back to the application.

C. Model

Azure Computer Vision Model for OCR using CNN and Tesseract:

Azure's Computer Vision Project is An AI service that examines the content in pictures and videos. As a region of this project, Azure provides APIs to extract printed content from images with the English language and writing styles. Latest Azure Computer Vision (CV) API scans the most recent OCR technology (learn what's new) that extracts printed content (English only), currency symbols, and digits from images. In the text-heavy image, it is optimized to extract the content with mixed languages. Here, the core algorithm is implemented to acknowledge the objects within the input image.

Microsoft Azures' Speech API for Text to Speech Conversion:

This is Azure's Speech carrier operation that converts textual content to real-looking speech. Validate fluid, natural-sounding textual content to speech that contests the layout associated pitch of human voices. Here, the OCR output generated within the preceding module is volumed as an entry to the services of the ones. These Azure Services can examine the textual content statistics and assemble a Speech in the style of an MP3 audio record to look at the textual content and deliver a downloadable link for the record.

D. Cloud Deployment

Here, for the cloud deployment process, we use either of Amazon web service (AWS) or the Google Cloud and the model has been deployed on a cloud server to form the answer accessible in and across geographical areas.

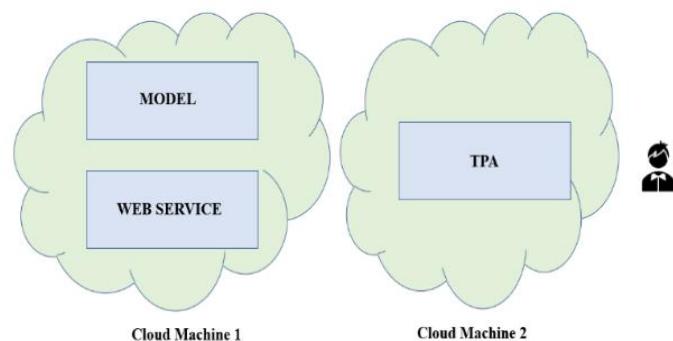
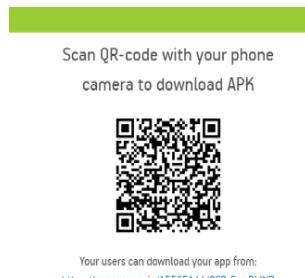


Fig. 2. Deployment Diagram

E. Creating App

The URL (uniform resource locator) of this project is regenerated to an APK file victimization a web or an internet appsgeyser onset here, we get both APK files as well as AlphaBB (ABB) files. APK file will be accustomed to directly installing the android app and AAB could be a business enterprise format that you just got to undergo the Play Store for publishing your app.



IV. RESULTS

We are going to create an ML model that can do all of these things. To run this model, we need to run it standalone, where we need to load images on a computer and get the output on the same computer, that's how this ML model shall run. But we didn't want that, so we wanted our model to be used by many other applications. For example, tomorrow someone might decide to develop an Android app. This android app captures images and invokes our project by sending this image as input and our project outputs this output this android app reads the output of our solution on their mobile phone so someone developing a smart app

tomorrow can. for the blind where a small programming chip is built into glasses like google glasses google specs concept google eyes concept where every time these glasses are embedded with a small camera and every time blind people where that specifies automatically takes the pictures in front of them, so if someone this deploys project tomorrow, he has the image as input, but he doesn't have that solution ready, he needs to deploy that solution from scratch, so you can capture the image and reuse our solution. We don't build a smart app for the blind. We're creating a solution for the blind that someone else will find useful in the future, that is, we're making the solution available for other people to use. For this to be possible we need to expose a solution to the outside world using the web services layer. This web services layer makes this solution available to the outside world. That is, anyone anywhere in the world can call our web service layer from anywhere by sending the image as input, and the web service layer passes that image to the model you input, fetches the result, and returns the result to the application return.

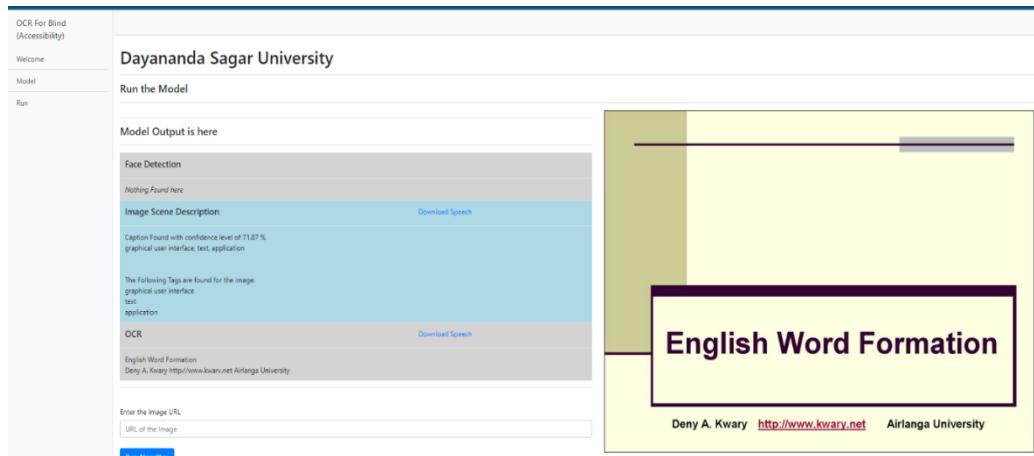


Fig. 3. Output 1

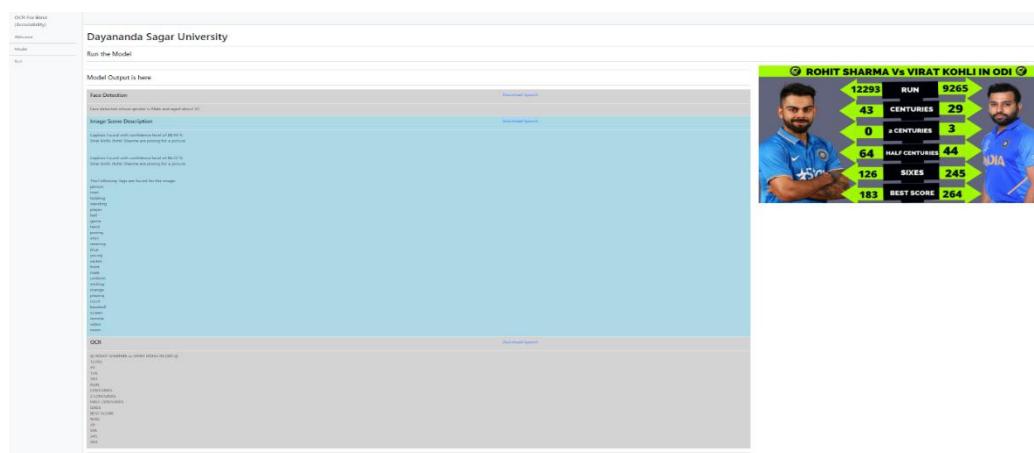


Fig. 4. Output 2

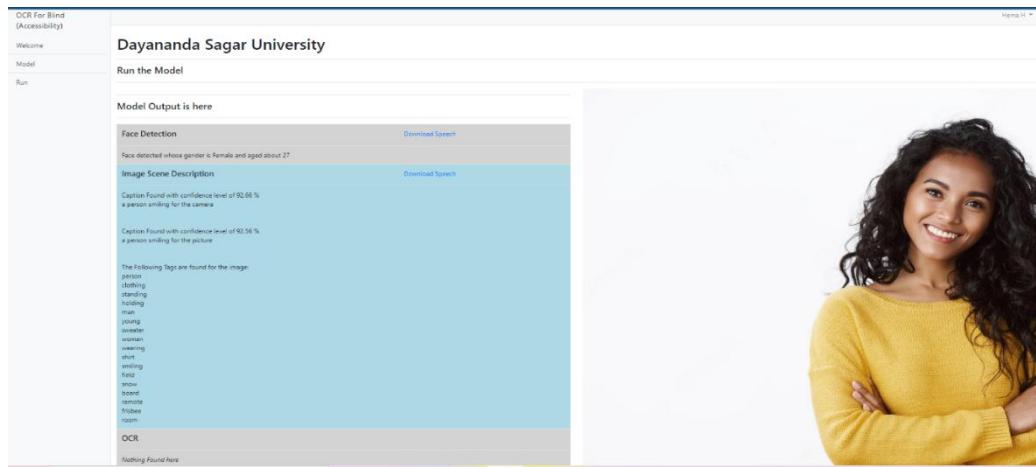


Fig. 5. Output 3

V. CONCLUSIONS & FUTURE WORK

An (Artificial Intelligence) i.e., AI-based reading system was developed to scan the detected content present in the image and convert it into digital text which is recognized by the system and displays the translated content and provides voice/speech output.

A. Future Work

The present project solution can be implemented further can develop smart application for blind where is small programming chip will be installed in their spectacles like Google lens, Google specs concept, Google eyes concept, where whenever those spectacles embedded with a small camera and whenever blind people where that specs it automatically captures the images in front of them.

REFERENCE

- [1] R. Deepa and Kiran N Lalwani, "Image Classification and Text Extraction Using ML", June 2019
- [2] Jisha Gopinath, Aravind S, Pooja Chandran, Saranya S S, "Text to Speech Conversion System using OCR", International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 1, January 2015.
- [3] Nawin Somyat and Songyot Nakariyakul, "Thai Lottery Number Reader App for Blind Lottery Ticket Sellers", 2018
- [4] Abhishek Mathur, Akshada Pathare, and Prerna Sharama, "AI-Based Reading System for Blind Using OCR", September 2019
- [5] Chandni Kanndiya, Diksha Chawala, and Yatin Chopra, "Automate Text Extraction from Images Using OCR System", March 2019
- [6] Ashwani Kumar, Ankush Chourasia, "Blind Navigation System Using Artificial Intelligence", March 2018
- [7] Sneha. C. Madre and S.B. Gundre, "OCR Based Image Text to Speech Conversion Using MATLAB", 2018
- [8] Bulla Rajesh, P. Nagabhushan, Mahammed Javed, and Watanabe Osamu, "Segmentation of Text-Lines and Words from JPEG Compressed Printed Text Document Using DCT Co-Efficient", 2020
- [9] Aaron James S, Sanjana S, Monisha M, "OCR based automatic book reader for the visually impaired using Raspberry PI", Vol. 4, Issue 7, January 2016.
- [10] Christopher G Relf, "Image Acquisition and Processing with LabVIEW", CRC Press, 2004.
- [11] Raviraj S Patkar, S. Pratap Singh, Ms. Swati V. Birje, "Marker Based Augmented Reality Using Android OS", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013.
- [12] Sonia Bhaskar, Nicholas Lavassar, Scott Green, "Implementing Optical Character Recognition on the Android Operating System for Business Cards".
- [13] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, "A Robust Algorithm for Text Detection in Images".
- [14] Minal Acharya, Priti Chouhan, and Asmita Deshmukh, "Scan. It-Text Recognition, Translation and Conversion", December 2019
- [15] Chucai Yi, Yingli Tian, "Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration", IEEE Transactions on Image Processing, Vol. 23 No. 7, July 2014.
- [16] Vikky Mohane, Chetan Gode, "Object Recognition for Blind People Using Portable Camera", IEEE 2016.