THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22- Issue-14-December-2019

Application for OCR & Navigation Assistance for Blind

Sowmya A.M[1], Sachin KumarS[2], Sandhya.S[3], Tejasri.K[4], Dhananjayan.S[5]

[1]Assistant Professor, Department of Computer Science & Engineering, Sri Sairam College of Engineering, Bengaluru.

[2345]UG Scholars, Department of Computer Science & Engineering, Sri Sairam College of Engineering, Bengaluru.

ABSTRACT---Visually impairment and illiterate, or have a learning disability is one of the biggest drawbacks for humanity, especially in this day and age when information and people is interconnected a lot by text messages (electronic and paper based) rather than talking. There is a need for convenient text reader that is reasonable and readily available to the Blind Community. The work in this research is images are converted into audio output. It is mainly used in the field of research in Character Recognition and Product Recognition, AI and Computer Vision. This device consists of three modules, image processing for text detection, object detection and convert detected things to voice. And then assisting the blind people to navigate by using the object detection API recognizing the objects and spelling it to the user. Button Camera/ Smart Phone Camera are designed to help Visually impaired in Navigation.

KEYWORDS---OCR, Tensor Flow, Tesser act, pyttsx.

1. INTRODUCTION

A common use of machine learning is to identify what an image represents. For example, we might want to know what type of animal appears in the following photograph. The task of predicting what an image represents is called image classification. An image classification model is trained to recognize various classes of images. For example, a model might be trained to recognize photos representing three different types of animals: rabbits, hamsters, and dogs. When we subsequently provide a new image as input to the model, it will output the probabilities of the image representing each of the types of animal it was trained on. Tensor flow is an open-source deep learning framework created by Google Brain. Tensor flow's Object Detection API is a powerful tool which enables everyone to create their own powerful Image Classifiers'

The Google Cloud Vision API enables developers to create vision-based machine learning applications based on object detection, OCR, etc. without having any actual background in machine learning. The Google Cloud Vision API takes incredibly complex machine learning models centred on image recognition and formats it in a simple REST API interface. It encompasses a broad selection of tools to extract contextual data on your images with a single API request. It uses a model which trained on a large dataset of images, similar to the models used to power Google Photos, so there is no need to develop and train your own custom model. Using the TensorFlow object Detection API and the Google Vision API can be deployed on a portable mobile device generally an android OS supported mobile phone. The visually challenged user can be guided through his navigation and OCR with speech output.

2.PROPOSED ALGORITHM

The below image (2.1) is a popular example of illustrating how an object detection algorithm works. Each object in the image, from a person to a kite, have been located and identified with a certain level of precision.

1.First, we take an image as input

2.Then we divide the image into various regions

3.We will then consider each region as a separate image.

4.Pass all these regions (images) to the CNN and classify them into various classes.

ISSN:0971-1260
Vol-22- Issue-14-December-2019

5.Once we have divided each region into its corresponding class, we can combine all these regions to get the original image with the detectedobjects.

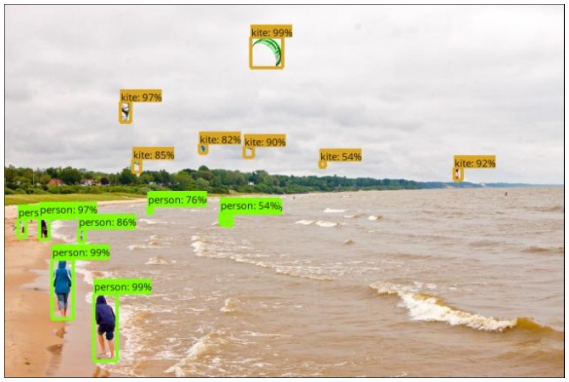


Fig (2.1)

3.OCR and Object detection

a. Google Vision

Google Cloud's Vision API offers powerful pre-trained machine learning models through REST and RPC APIs. Assign labels to images and quickly classify them into millions of predefined categories. Detect objects and faces, read printed and handwritten text, and build valuable metadata into your image catalog.

b. TensorFlow Object Detection API

For assisting the user, the system has to first identify the objects through the camera and spell it as a speech output. When the user is walking the obstacle is called that object which comes near and in your way. The video frame will be containing so many objects but we will only look at some objects that are in our way. For that ROI is defined. This function will black the other part in the frame except the coordinates which we will give. Hence, extra objects will not be detected. For that the logic is to create two boundaries in ROI named left and right boundary. Hence it will split out the ROI into 3 parts. If the obstacle is in the right part, we will tell the user to move left and if it is into right then we will tell the user to move right.

c. Text recogniser

All of these systems take in data – often an image – from an unknown source, analyse the data in that input, and attempt to match them to existing entries in a database of known images. Object recognition does this in three steps: detection, Text print creation, and verification or identification. When an image is captured, computer software analyses it to identify where the text are in, say, a crowd of words. In a photo, for example, cameras will feed into a software to identify text in the video feed.

d. Speech Output

To convert speech to on screen text or a computer command, a computer has to go through several complex steps in this implementat6ion we use Text to Speech (TTS) library for Python 2 and 3.Works without internet connection or delay. Supports multiple TTS engines, including Sapi5, nsss, and espeak. pyttsx is a cross-platform text to speech library which is platform independent. The major advantage of using this library for text-to-speech conversion is that it works offline. TTS includes two different model implementations which are based on Tacotron and Tacotron2. Tacotron is smaller, efficient and easier to train but Tacotron2 provides better results, especially when it is combined with a Neural vocoder.

e. Android Application

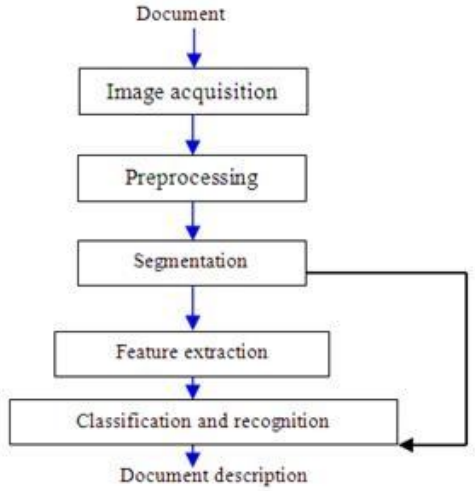
An Android app is a software application running on the Android platform. Because the Android platform is built for mobile devices, a typical Android app is designed for a smartphone or a tablet PC running on the Android OS. Both the API are integrated into the application the user can use both OCR and navigation assistance system there is an option to switch between the modes.

4.Implementation of Algorithm OCR

OCR is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image. Through the scanning process a digital

ISSN:0971-1260
Vol-22- Issue-14-December-2019

image of the original document is captured. In OCR optical scanners are used, which generally consist of a transport mechanism plus a sensing device that converts light intensity into gray-levels. Printed documents usually consist of black print on a white background. Hence, when performing OCR, it is common practice to convert the multilevel image into a bilevel image of black and white.Often this process, known as thresholding, is performed on the scanner to save memory space and computational effort. The thresholding process is important as the results of the following recognition is totally dependent of the quality of the bilevel image.



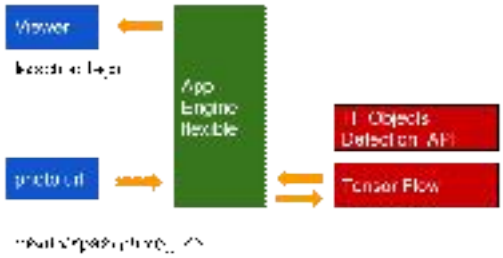
Fig(4.1)-OCR Flow Chart

Still, the thresholding performed on the scanner is usually very simple. A fixed threshold is used, where gray-levels below this threshold is said to be black and levels above are said to be white. For ahigh-contrast document with uniform background, a prechosen fixed threshold can be sufficient. However, a lot of documents encountered in practice have a rather large range in contrast. In these cases, more sophisticated methods for thresholding are required to obtain a good result. Segmentation is a process that determines the constituents of an image. It is necessary to locate the regions of the document where data have been printed and distinguish them from figures and graphics. For instance, when performing automatic mail-sorting, the address must be located and separated from other print on the envelope like stamps and company logos, prior to recognition.

5.Object Detection API

To train a robust classifier, we need a lot of pictures which should differ a lot from each other. So, they should have different backgrounds, random object, and varying lighting conditions. In order to label our data, we need some kind of image labeling software. LabelImg is a great tool for labeling images. With the images labeled, we need to create TFRecords that can be served as input data for training of the objectdetector.Namely,the *xml_to_csv.py* and *generate_tfrecord.py* files.

To train the model we use faster_rcnn_inception, which just like a lot of other models will start with sample config. trained model need to generated an inference graph, which can be used to run the model



Fig(5.1)-TF API

a.TensorFlow Lite

TensorFlow Lite is a set of tools to help developers run TensorFlow models on mobile, embedded, and IoT devices. It enables on-device machine learning inference with low latency and a small binary size.

TensorFlow Lite is designed to execute models efficiently on mobile and other embedded devices with limited compute and memory resources. Some of this efficiency comes from the use of a special format for storing models. TensorFlow models must be converted into this format before they can be used by TensorFlow Lite. Example Fig (5.1)

Converting models reduces their file size and introduces optimizations that do not affect accuracy.

ISSN:0971-1260
Vol-22- Issue-14-December-2019

The TensorFlow Lite converter provides options that allow you to further reduce file size and increase speed of execution, with some trade-offs.
TensorFlow Lite provides tools to optimize the size and performance of your models, often with minimal impact on accuracy. Optimized models may require slightly more complex training, conversion, or integration.

The Model Optimization Kit is a set of tools and techniques designed to make it easy for developers to optimize their models. Many of the techniques can be applied to all TensorFlow models and are not specific to TensorFlow Lite, but they are especially valuable when running inference on devices with limited resources.

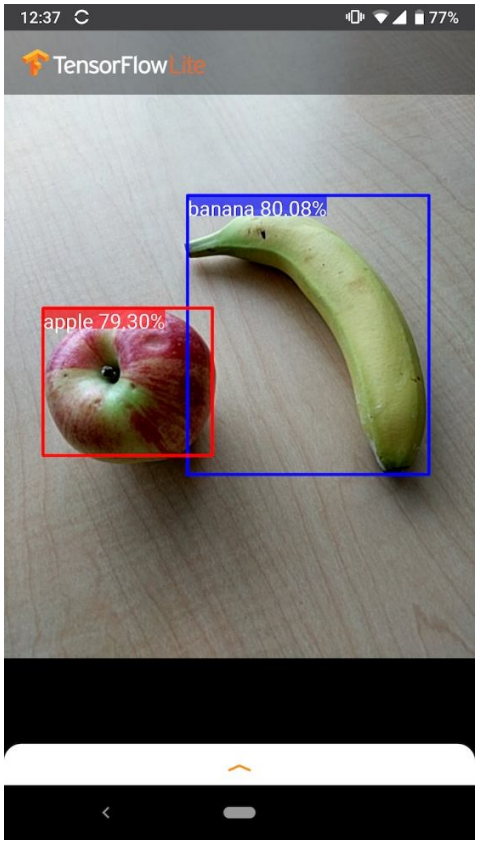


Fig (5.1)-Application TF Lite

ADVANTAGES:

- Works on offline Mode.
- Can be used in a daily driver.
- Reduced operational cost
- Improved security and customer experience
- User friendly
- Voice assistance enabled
- Offers mobility
- Ensures the highest accuracy

DISADVANTAGES:

- Additionally, requires a headphone
- Camera should point the user's way.
- Comparatively high computing.
- Consumes more battery.
- The database has to be explicitly present.

CONCLUSION AND FORESEEK:

This research is to develop the Mobile Application which is used in assisting the visually challenge people to navigate easily and read out the text. This preliminary study will be composed the system design, analysis, testing, implementation, and performance evaluation. The research illustrates the efficiency, accuracy, quantity of subjects, and good results. Providing both OCR and navigation assistance stand out a complete feature rich package in single application for the user.

REFERENCES:

[1]. Case Study-Mustamo P. (2018) Object detection in sports: TensorFlow Object Detection API case

ISSN:0971-1260
Vol-22- Issue-14-December-2019

study. University of Oulu, Degree Programme in Mathematical Sciences. Bachelor's Thesis, 43 p.

[2]. Shaikh, F. (2017, May 18). Why are GPUs necessary for training deep learning models? Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2017/05/gpus-necessary-for-deep-learning/

[3].TensorFlow. (2017, November 2). tf.Tensor. API Documentation.

[4]. H. Hosseini, B. Xiao, and R. Poovendran, "Deceiving google's cloud video intelligence api built for summarizing videos," arXiv preprint arXiv:1703.09793, 2017.