TECHNICAL SOLUTIONS FOR VISUALLY IMPAIRED

LeadingIndia.ai

Submitted to :
Dr. Madhushi Verma.
Mrs. Kanak Manjari.

Submitted By : Team No – 9

1.) **Akshat Jaiswal - Amritha School of Engineering, Amrithapuri.**
2.) **Ashish Ranjan - National Institute of Technology, Manipur.**
3.) **Yash Pise - KIT's College of Engineering, Kolhapur.**
4.) **Kamala Raj T - Mahendra College of Engineering, Mallasamudram.**

Abstract

The projects aim is to help visually impaired people by using real time object detection in a live video stream, convert the detected object to speech and describe the position of the object. Apart from this we have also implemented image to speech conversion. Mainly, Deep Learning is used for Implementation. We have created a user interface to merge both the modules. We will discuss the progress made leading up to the current development scene and possible future enhancements that can serve as motivation for further work.

Introduction

Designed for the blind and low vision community, this research project harnesses the power of AI to describe people, text and objects. This project brings together the power of AI to deliver an intelligent system designed to help you navigate your day. Point your phone's camera, select a channel, and hear a description of what the AI has recognized around you.

With its intelligent system, just hold up your camera and hear information about the world around you. Our system can speak short text as it appears in front of the camera, provide audio guidance to capture a printed page, and recognizes and narrates the text.

- Recognize and locate the faces of people you're with.
- Reads text quickly and get audio guidance to capture full documents.

Our project is an extended work of real time object detection. We have implemented real time object detection using COCO API, which detects the object on live video stream and converts the objects to speech and give a gist of where the object is.

Motivation

The main reason for choosing this project is to help the visually impaired people who was facing day to day problems in the society. Too frequently, blindness affects a person's ability to self- navigate the outside well known environments and even simply walking down a crowded street. It affects a person's ability to perform job duties and also activities outside of the workplace, such as sports as well as academics. Many of these social challenging limit a blind person's ability to meet the people, and this only adds to low self -esteem. In our modern world there is a very large number of developments in machineries and electronic accessories but there are some components only which helps to the visually impaired people. Our project will definitely help for the people who has eye defectiveness are also visually impaired. Our main is to help the people who was visually impaired with the help of artificial intelligence in the way of detecting an obstacle, detecting direction, detecting person and also one more important is it converts text to speech.

Literature Survey

We could find readily pre trained models/projects

https://github.com/tensorflow/models/tree/master/research/object_detection

Detection of objects is referred from this link.

https://pjreddie.com/darknet/yolo/

YOLO darknet and COCO API is referred from this link.

But major problem was that none of the model converts real time detected object to speech so with the help to keras-yolov3 and pre trained model yolov3.h5 we created a module which can convert detected object to speech and describes the object's position.

Diagram

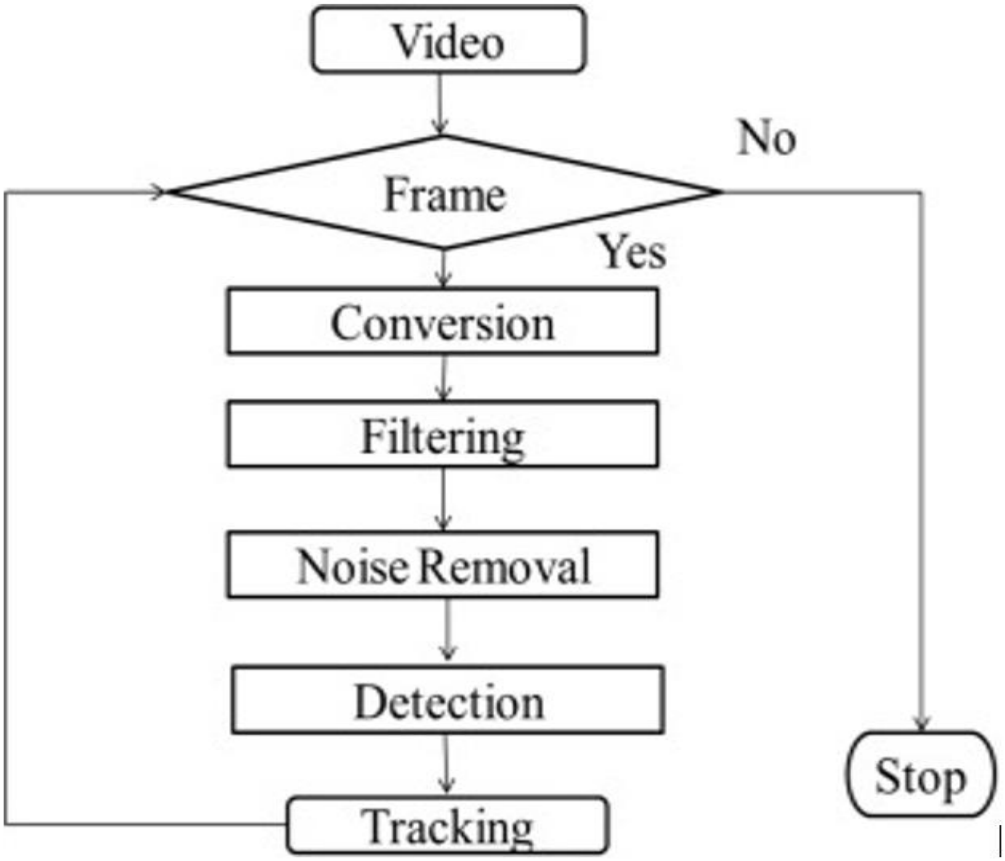


Figure 1.1

This flowchart in figure 1.1 explains how input is been taken as a video which converts to input frames. The input frames is then been pre-processed and run through yolo algorithm which detects objects using bounding box method.

Functionality

Most part of the first week was spent on getting the data for real time object detection and setting up coco api and installing modules in our system. We decided to use keras-yolov3 implementation as it has one of the best accuracy.

As real time object detection requires huge amount of classes which will take a lot of time to train so we choose to download pre trained model weight.

With the help of open-cv and webcam we gave the model input frames and run through yolo algorithm which detects objects using bounding box method.

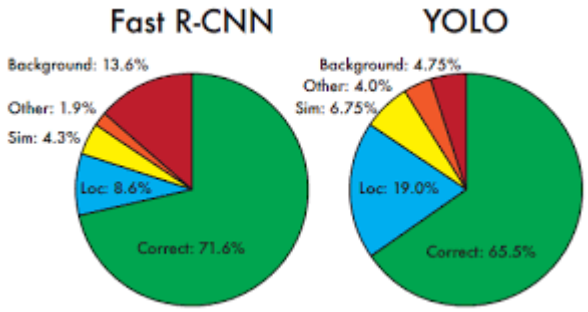


Figure 1.2

Figure 1.2 shows the difference between Fast RCNN and YOLO

Experimental results shows that Fast R-CNN has better accuracy than YOLO model but we still preferred YOLO as it is extremely fast. We simply run our neural network on a new image at time to predict the detections. Experimental results has shown that the network runs at 45 frames per second with no batch processing on a Titan X GPU and a fast version runs at more than 150 fps.

The objects which were detected using YOLO algorithm was then converted to speech using **gTTS** (*Google Text-to-Speech*), a Python library and CLI tool to interface with Google Translates text-to-speech API. The gtts converts the text into speech.

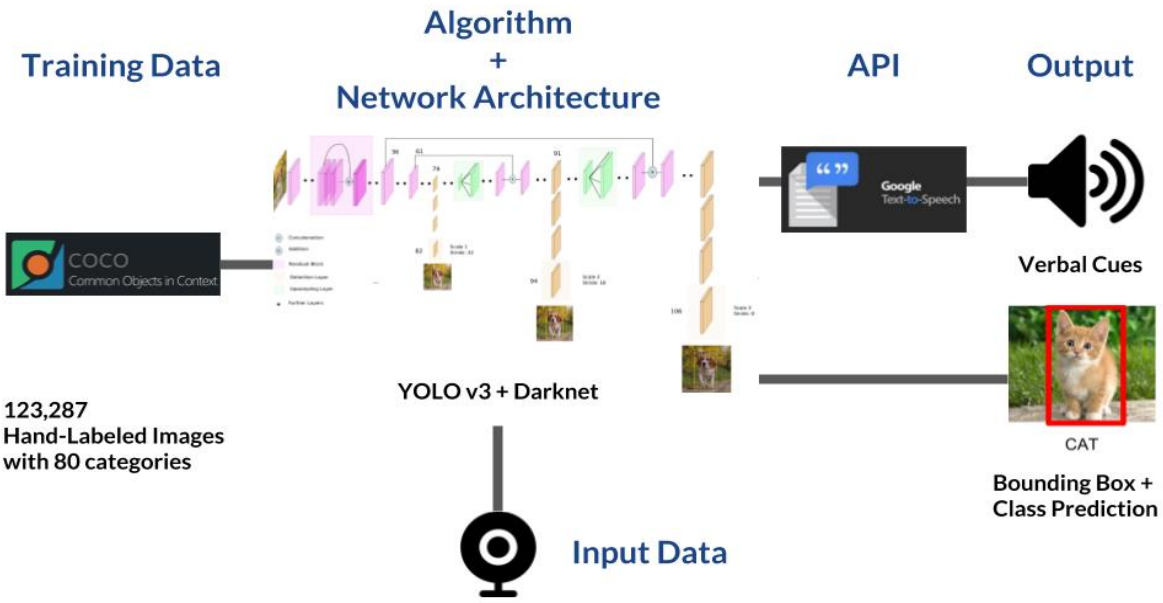


Figure 1.3 (Architecture)

In figure 1.4, we find the position of the object we divide the screen coordinates in 4 parts top left, top right, bottom left, bottom right and if the object lies between the region set by the coordinates we specify the object position and convert to speech using gTTs.

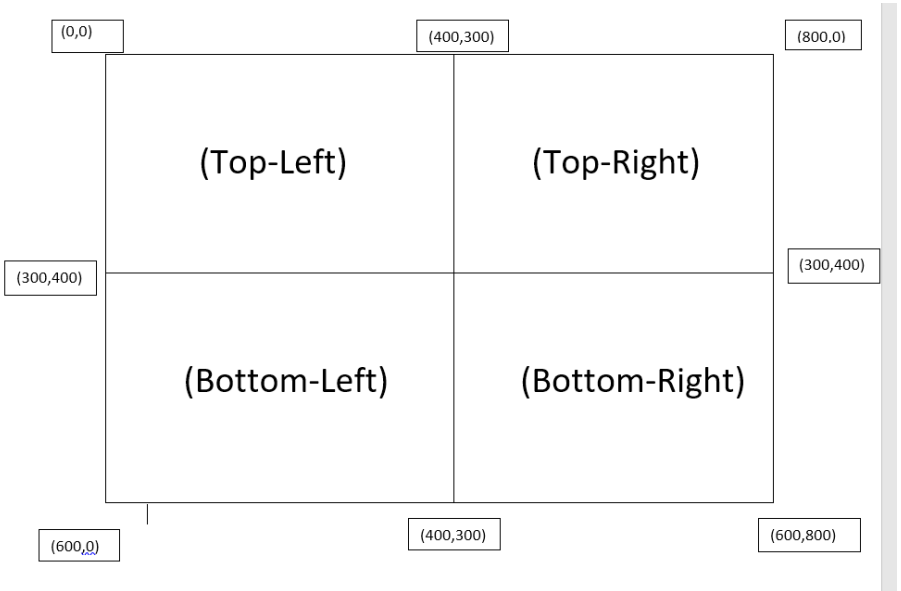


Figure 1.4

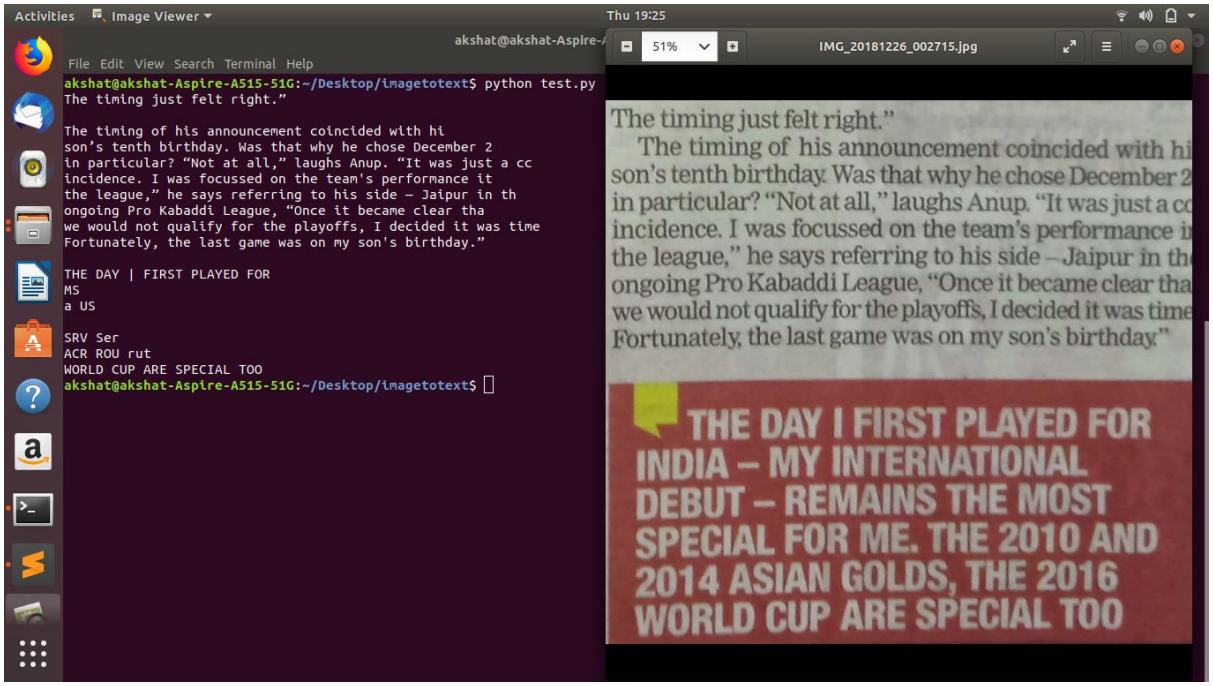
Our project continued as we tried to convert text to speech, for that we used Pytesseract library. Python- tesseract is an optical character recognition (OCR) tool for python. That is, it will recognize and “read” the text embedded in images. Python-tesseract is a wrapper

for Google's Tesseract-OCR Engine. It is also useful as a stand-alone invocation script to the tesseract, it can read all image types supported by the Python Image Library, including images which are in the form of jpeg, gif, bmp, tiff, and others, whereas tesseract-ocr by default only supports tiff and bmp.

That pytessrect reads the live input images and then it converts into text. After that for converting the text into speech we use **gTTS** (*Google Text-to-Speech*), a Python library and CLI tool to interface with Google Translates text-to-speech API. The gtts converts the text into speech.

The drawback of this model is that it can't detect handwritten text and need a very good image quality with no background noise.

Converts printed text to images:



Testing images

The output of the object was something like this:

Finally we created a UI interface to merge bout the api which runs on speech to text application.

It detects the voice of the user and gives the results accordingly for example option 1 for real time object detection and option 2 as image to speech classification and according to the options the runs the following module.

Conclusion

In our project we have used object detection COCO API, which uses **YOLO** algorithm, and **PYTESSERECT** for image to speech .The YOLO algorithm takes image/videos as an input and creates bounding boxes according to the trained data. This convolutional implementation of the sliding window makes yolo an excellent approach for object detection. Pytesserect is an OCR tool for python that converts image to text and gtts converts that text into speech. The purpose for using KERAS_YOLOV3 is that is gives better accuracy than other modules.

Limitations

Object Recognition Problem: Object detection and tracking remains an open research problem even after research of several years in that field. A robust, and high performance approach is still a great challenge today. This difficulty level of this problem highly depends on how one defines the object to be detected and then tracked. If only a few visual features (e.g. colour) are used as representation of an object, it is not so difficult to identify all pixels with same colour as the object. It is desirable that the background model adapts to gradual changes of the appearance of that environment. For example in outdoor settings, the light intensity typically varies during the day. Sudden illumination changes can also occur in that scene. This type of change occurs for example with sudden switching on/off a light in a indoor environment. This may also happen in the outdoor scenes (fast transition from cloudy to bright sunlight). Illumination strongly affects the appearance of background, and cause false positive detections.

Text to speech conversion

In this method also some drawbacks are observed, the handwritten text will not be recognised by the module and nor the flipped images and the rotating images are detected by the module. The most important thing is that quality of the image is playing main role in image detection.

Future scope

Object Detection : Can train or more classes and removal of background noise which can enhance the quality of the image Text to speech :Instead of using a predefined model we can use deep learning methods such as CNN-RNN-CTC scanning for text and handwritten written text, removal of background noise using opencv to enhance the quality of the image.

Video & Github Link

https://youtu.be/hcshEgfMG_4

https://github.com/ak9969/Technical-Solutions-For-Visually-Impaired

References

1. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

2. https://pdfs.semanticscholar.org/8f98/ad7509032de3a19458287ee4cbfc858064a6.pdf

3. https://medium.com/@MicroPyramid/extract-text-with-ocr-for-all-image-types-in-python-using-pytesseract-ec3c53e5fc3a

4. https://ieeexplore.ieee.org/document/4669755

5. https://pjreddie.com/darknet/yolo/