THE ML COMMUNITY PROJECT

An Engineering Project in Community Service

Final Report

Submitted by

SHUBHAM GUPTA [19BCE10261]

in partial fulfillment of the requirements for the degree of

Bachelor of Engineering and Technology

VIT Bhopal University
Bhopal
Madhya Pradesh

April 2022

VIT®
BHOPAL
www.vitbhopal.ac.in

Bonafide Certificate

Certified that this project report titled **“The ML Community Project”** is the bonafide work of “(19BAI10070 DEV GUPTA, 19BAI10015 YUKTI SINGH, 19BAI10192 ANISH ANAND, 19BAI10180 OM KRISHNA YADAV, 19BCE10261 SHUBHAM GUPTA, 19BCY10101 ANKITH T, 19BCE10063, VARDHAN VISHNU, 19BEE10007, KARNATI SIVAMANIKANTA)” who carried out the project work under my supervision.

This project report (Phase I) is submitted for the Project Viva-Voce examination held on …22-04-2022………..

Satyam Ravi

Dr. Satyam Ravi
Chemistry, SASL
Supervisor

1. 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 will speak short text as it appears in front of the camera, provide audio guidance to capture a printed page, and recognize and narrate the text.

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

Our project will be an extended work of real-time object detection. We will implement 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.

1.1 Motivation

The main motivation for choosing this project is to help the visually impaired people who were facing day-to-day problems in society. Too frequently, blindness affects a person's ability to self-navigate the outside well-known environments and even simply walk 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 challenges limit a blind person's ability to meet people, and this only adds to low self-esteem.

In our modern world, there is a very large number of developments in machinery and electronic accessories but there are some components only that help visually impaired people. Our project will definitely help people who have eye defectiveness are also visually impaired. Our main is to help 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.

.

1.2 Objective

The project's aim is to help visually impaired people by using real-time object detection in a live video stream, converting the detected object to speech, and describing the position of the object. Apart from this, we will also implement the image-to-speech conversion. Mainly, Deep Learning will be used for Implementation. We will create a user interface to merge both 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.

2. Existing Work / Literature Review

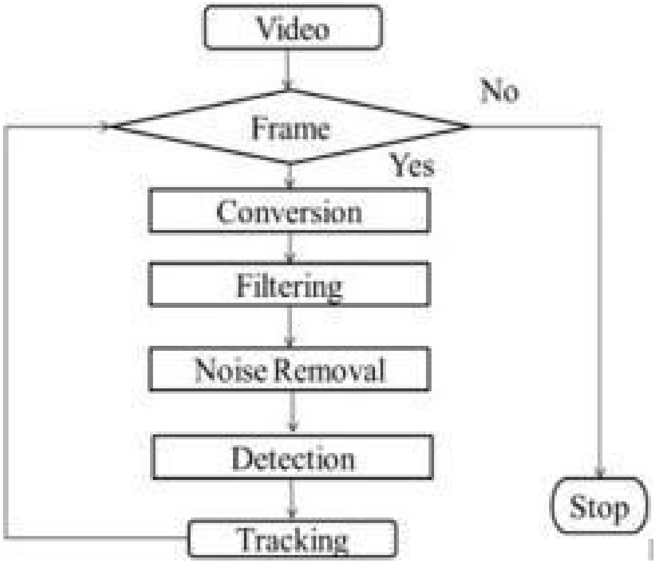
We could find readily pre-trained models/projects

Detection of objects is referred from this link.:
https://github.com/tensorflow/models/tree/master/research/object_detection

YOLO darknet and COCO API is referred from this link.:
https://pjreddie.com/darknet/yolo/

But the major problem was that none of the models converts the real-time detected objects to speech so with the help of Keras-yolov3 and pre-trained model yolov3.h5 we will create a module that can convert detected object to speech and describe the object's position.

Diagram :



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

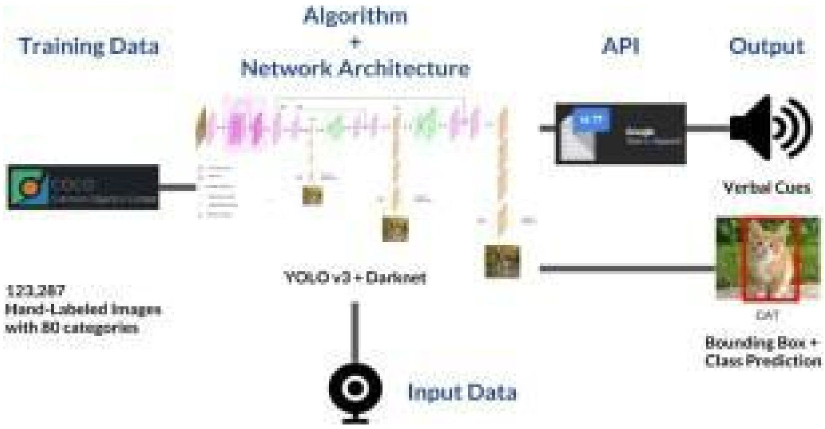
3. Topic of the work

System Design / Architecture

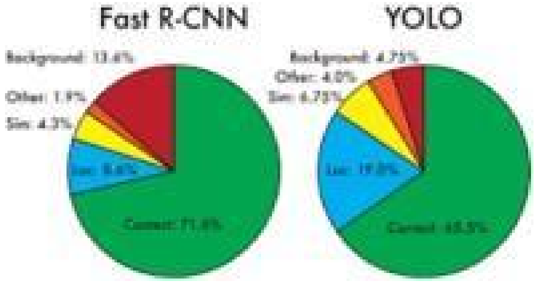
Most parts before the first review were spent on getting the data for real-time object detection and setting up coco API and installing modules in our system. We decided to use the Keras-yolov3 implementation as it has one of the best accuracies.

As real-time object detection requires a 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 will give the model input frames and run through yolo algorithm which detects objects using the bounding box method.



Working Principle



The figure shows the difference between Fast RCNN and YOLO

Experimental results show that Fast R-CNN has better accuracy than the YOLO model but we will still prefer YOLO as it is extremely fast. We will simply run our neural network on a new image at a time to predict the detections. The objects which will be detected using the YOLO algorithm

will be then converted to speech using **gTTS** (*Google Text-to-Speech*), a Python library, and a CLI tool to interface with Google Translates text-to-speech API. The gtts convert the text into speech.

Expected Results

The expected result would look like this :

4.TEAM CONTRIBUTION

- OM KRISHNA YADAV: In this project our main focus was to build the algorithm and train the dataset using the YOLO V3 model. All the data that will be captured through the ESP32 camera and after the frame distribution, the most important part of the project, that is processing each frame to determine the object in the image is handled in this part. YOLO V3 model is the latest version of YOLO which is easy to use, fast and more accurate then other versions which let us process the image efficiently.

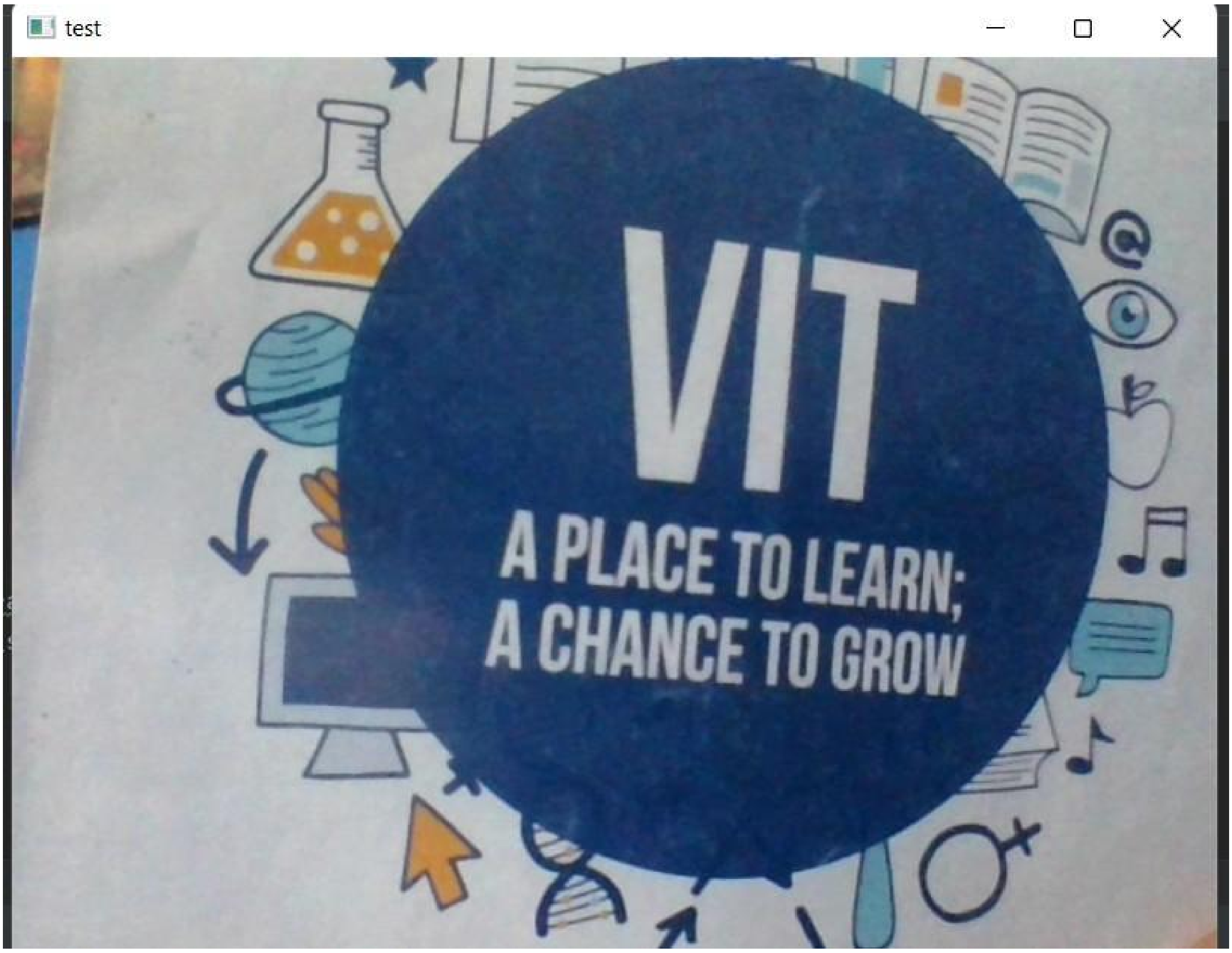
- SHUBHAM GUPTA & ANKITH T : In this project our main focus was to build the back end code for capturing the image from the ESP32 camera and process the image to extract the text from the image. This text is then converted into speech and played to the user. This is mainly done with the help of PyTesseract library and gTTS library. PyTesseract is a very helpful library which is used for converting image to string and store it into a .txt file while gTTS library is used for converting this .txt file into audio .mp3 format and play right after execution.
- ANISH ANAND & YUKTI SINGH : In this project our main focus is to build an interface from which we can connect our user to the project. We widely used the tkinter library of python to achieve the goal. Tkinter is an interface in python and available on almost all platforms including Windows,Linux and Mac. Our main focus is to display the project details and functionalities to the user in the most appropriate way.
- DEV GUPTA & VARDAAN VISHNU: In this project our main goal was to detect final objects from the processed image from the YOLO V3 model using the bounding box algorithm and OpenCV library. This object detection method is used for determining the objects in the image and then the image is sent to the YOLO model again for future processing.
- KARNATI SIVAMANIKANTA : He has worked on data collection and

further hardware implementation part if some possibilities are there with good outcome.

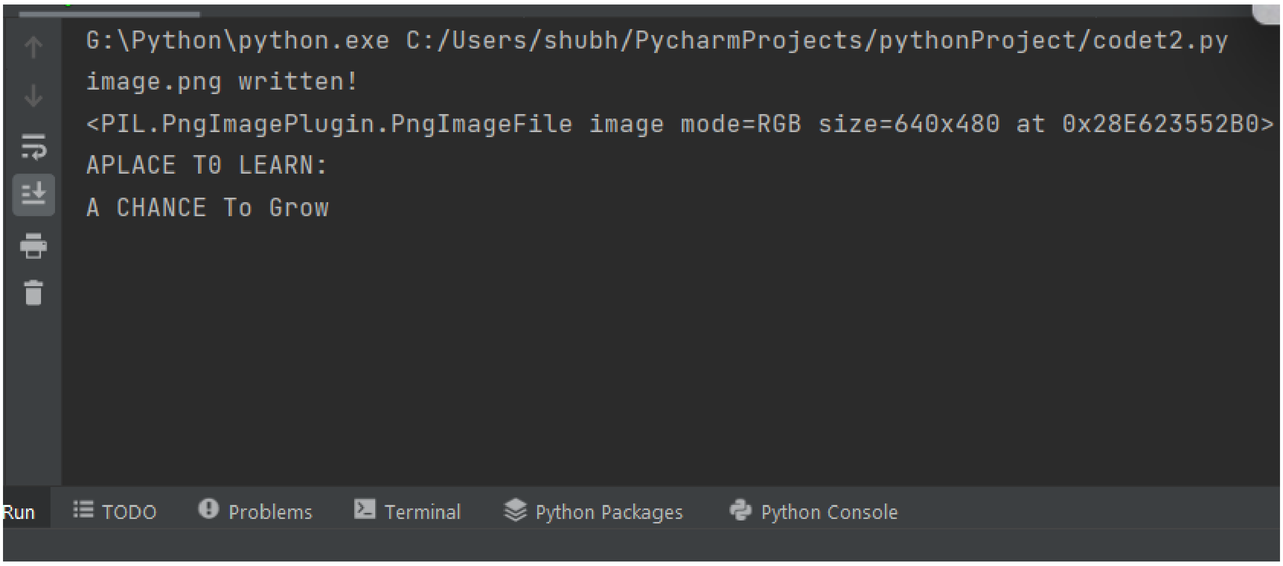
5.INDIVIDUAL WORK

In this project my main docs was to build the back end code for capturing the image from the ESP32 camera and process the image to extract the text from the image. This text is then converted into speech and played to the user. This is mainly done with the help of PyTesseract library and gTTS library. PyTesseract is a very helpful library which is used for converting image to string and store it into a .txt file while gTTS library is used for converting this .txt file into audio .mp3 format and play right after execution.

My co-worker 19BCY10101 ANKITH T has an important supportive role in building this.

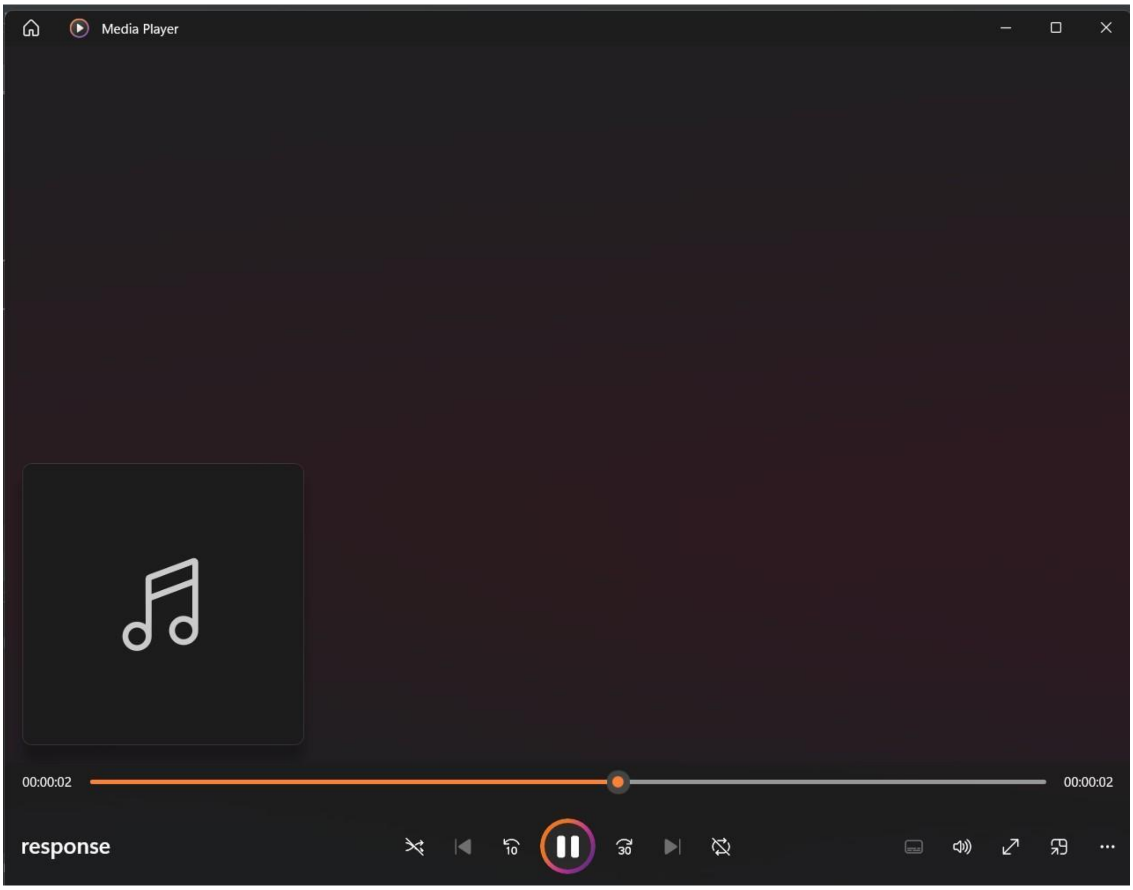


This code first captures the frames bounded into the text category by the Yolo model (Bounding Box Method) and then converts all the text present in the image frame into string and then store it into a text file.



```
G:\Python\python.exe C:/Users/shubh/PycharmProjects/pythonProject/codet2.py
image.png written!
<PIL.PngImagePlugin.PngImageFile image mode=RGB size=640x480 at 0x28E623552B0>
APLACE T0 LEARN:
A CHANCE To Grow
```

After successfully converting the image into text file the gTTS library is used to convert the text to speech which gives a mp3 file as an output and the file is opened right after execution and the audio plays to the hardware.



6.CONCLUSION

In our project, we will use object detection COCO API, which will use the **YOLO** algorithm and **PYTESSERECT** for image-to-speech conversion . The YOLO algorithm will take images/videos as input and will create 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 images to text and gets converts that text into speech. The purpose of using KERAS_YOLOV3 is that it gives better accuracy than other modules.

7. Reference:

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-pytho n-using-pytesseract-ec3c53e5fc3a**

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

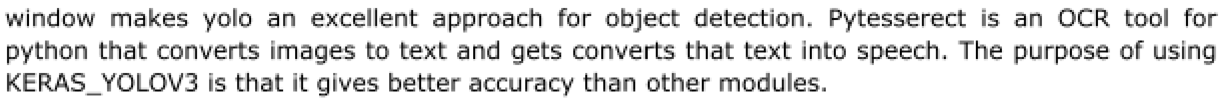
5. **https://pjreddie.com/darknet/yolo/**

Sentence

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 will speak short text as it appears in front of the camera, provide audio guidance to capture a printed page, and recognize and narrate the text. ● Recognize and locate the faces of people you’re with. ● Reads text quickly and gets audio guidance to capture full documents. Our project will be an extended work of real-time object detection. We will implement 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. 3 1.1 Motivation The main motivation for choosing this project is to help the visually impaired people who were facing day-to-day problems in society. Too frequently, blindness affects a person's ability to self-navigate the outside well-known environments and even simply walk 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 challenges limit a blind person's ability to meet people, and this only adds to low self- esteem. In our modern world, there is a very large number of developments in machinery and electronic accessories but there are some components only that help visually impaired people. Our project will definitely help people who have eye defectiveness are also visually impaired. Our main is to help 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. . 1.2 Objective The project's aim is to help visually impaired people by using real-time object detection in a live video stream, converting the detected object to speech, and describing the position of the object.

Apart from this, we will also implement the image-to-speech conversion. Mainly, Deep Learning will be used for Implementation. We will create a user interface to merge both 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. 4 2.

Existing Work / Literature Review We could find readily pre-trained models/projects Detection of objects is referred from this link.: https://github.com/tensorflow/models/tree/master/research/object_detection YOLO darknet and COCO API is referred from this link.: https://pjreddie.com/darknet/yolo/ But the major problem was that none of the models converts the real-time detected objects to speech so with the help of Keras-yolov3 and pre-trained model yolov3.h5 we will create a module that can convert detected object to speech and describe the object's position. Diagram : This flowchart in the above diagram explains how input is been taken as a video which converts to input frames. The input frames are then been pre-processed and run through yolo algorithm which detects objects using the bounding box method. 5 3. Topic of the work System Design / Architecture Most parts before the first review were 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 the Keras-yolov3 implementation as it has one of the best accuracies. As real-time object detection requires a 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 will give the model input frames and run through yolo algorithm which detects objects using the bounding box method. Working Principle The figure shows the difference between Fast RCNN and YOLO Experimental results show that Fast R-CNN has better accuracy than the YOLO model but we will still prefer YOLO as it is extremely fast. We will simply run our neural network on a new image at a time to predict the detections. The objects which will be detected using the YOLO algorithm 6 will be then converted to speech using gTTS (Google Text-to-Speech), a Python library, and a CLI tool to interface with Google Translates text-to-speech API. The gtts convert the text into speech. Expected Results The expected result would look like this : 4. INDIVIDUAL WORK In this project my main docs was to build the back end code for capturing the image from the ESP32 camera and process the image to extract the text from the image. This text is then converted into speech and played to the user. This is mainly done with the help of PyTesseract library and gTTS library. PyTesseract is a very helpful library which is used for converting image to string and store it into a .txt file while gTTS library is used for converting this .txt file into audio .mp3 format and play right after execution. My co-worker 19BCY10101 ANKITH T has an important supportive role in building this. This code first captures the frames bounded into the text category by the Yolo model (Bounding Box Method) and then converts all the text present in the image frame into string and then store it into a text file. After successfully converting the image into text file the gTTS library is used to convert the text to speech which gives a mp3 file as an output and the file is opened right after execution and the audio plays to the hardware. 9 6.CONCLUSION In our project, we will use object detection COCO API, which will use the YOLO algorithm and PYTESSERECT for image-to-speech conversion . The YOLO algorithm will take images/videos as input and will create 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 images to text and gets converts that text into speech. The purpose of using KERAS_YOLOV3 is that it gives better accuracy than other modules.

Report Title:	Report
Report Link: (Use this link to send report to anyone)	https://www.check-plagiarism.com/plag-report/72550b73ebc2fb8624ec547f7ffaee65994c91650652419
Report Generated Date:	22 April, 2022
Total Words:	1034
Total Characters:	6366
Keywords/Total Words Ratio:	0%
Excluded URL:	No
Unique:	**86%**
Matched:	**14%**

Sentence wise detail:

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. 4 2.

Existing Work / Literature Review We could find readily pre-trained models/projects Detection of objects is referred from this link.: https://github.

com/tensorflow/models/tree/master/research/object_detection YOLO darknet and COCO API is referred from this link.: https://pjreddie.

com/darknet/yolo/ But the major problem was that none of the models converts

the real-time detected objects to speech so with the help of Keras-yolov3 and

pre-trained model yolov3.

h5 we will create a module that can convert detected object to speech and describe the object&#039;s position.

Diagram : This flowchart in the above diagram explains how input is been taken as a video which converts to input frames.

The input frames are then been pre-processed and run through yolo algorithm which detects objects using the bounding box method. 5 3.

Topic of the work System Design / Architecture Most parts before the first review were spent

on getting the data for real-time object detection and setting up coco API and installing modules in

our system. (0) We decided to use the Keras-yolov3 implementation as it has one of the best

accuracies.

As real-time object detection requires a 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 will give the model input frames and run through yolo algorithm which detects objects using the bounding box method.

Working Principle The figure shows the difference between Fast RCNN and YOLO Experimental results

show that Fast R-CNN has better accuracy than the YOLO model but we will still prefer YOLO as it is

extremely fast. (1)

We will simply run our neural network on a new image at a time to predict the detections.

The objects which will be detected using the YOLO algorithm 6 will be then converted to

speech using gTTS (Google Text-to-Speech), a Python library, and a CLI tool to interface with

Google Translates text-to-speech API.

The gtts convert the text into speech.

After successfully converting the image into text file the gTTS library is used to convert the text to speech which gives a mp3 file as an output and the file is opened right after execution and the audio plays to the hardware. 9 6. (2)

CONCLUSION In our project, we will use object detection COCO API, which will use the YOLO algorithm and PYTESSERECT

for image-to-speech conversion .
The YOLO algorithm will take images/videos as input and will create 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 images to text and gets converts that text

into speech. (3) The purpose of using KERAS_YOLOV3 is that it gives better accuracy than

Match Urls:

0: https://context.reverso.net/traduccion/ingles-espanol/out+of+our+system

1: https://www.merriam-webster.com/dictionary/fast

2: https://www.amazon.com/96-bypass-door-hardware/s?k=96+bypass+door+hardware

3: https://www.stackshare.io/pypi-pytesseract

Keywords Density

One Word	2 Words	3 Words
object 3.97%	object detection 1.44%	real time object 0.9%
text 3.07%	text speech 1.08%	pre trained model 0.54%
yolo 2.89%	real time 1.08%	bounding box method 0.54%
convert 2.89%	yolo algorithm 0.9%	algorithm detects objects 0.36%
speech 2.53%	time object 0.9%	run yolo algorithm 0.36%

Plagiarism Report

By check-plagiarism.com

Satyam Ravi