Object Information Retrieval in Real-time

Group name : Now You SEE Me

Shrushti Gupta, Sai Dharshini Akula, Abhishek Veeravelli, Palacherla Bhaskar Praveen
{sgupta32, sakula8, aveerave, bpalache}@gmu.edu

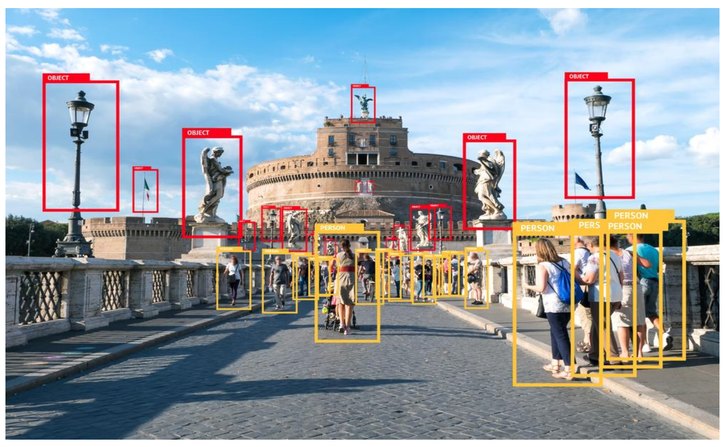
Dataset

https://www.kaggle.com/datasets/awsaf49/coco-2017-dataset

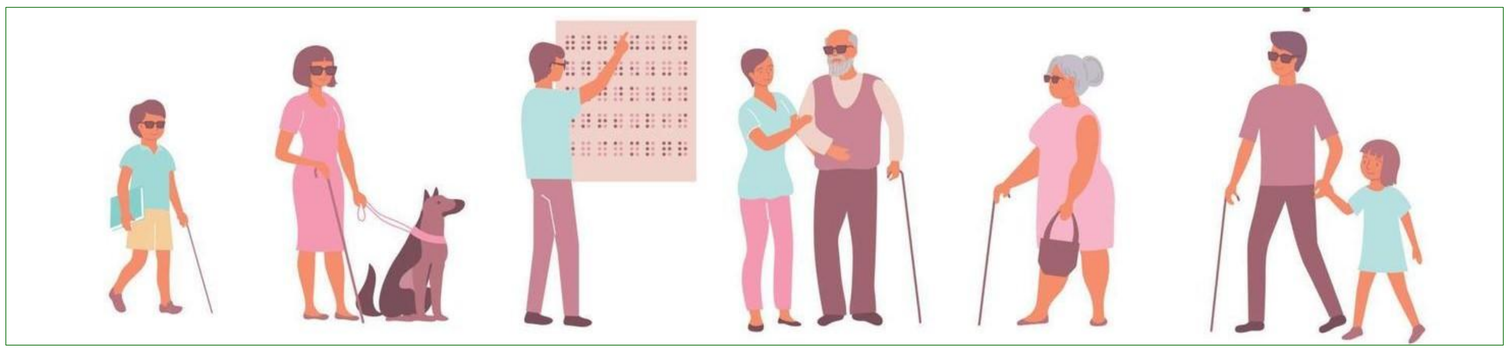
Common Objects in Context (COCO) is a high-quality dataset primarily used in neural networks.

COCO has several features:

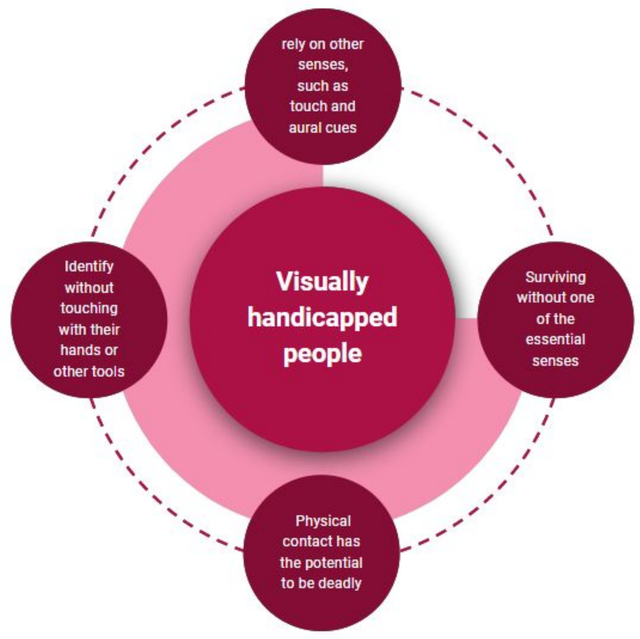
- 328K images (>200K labeled)
- 1.5 million object instances
- 80 object categories



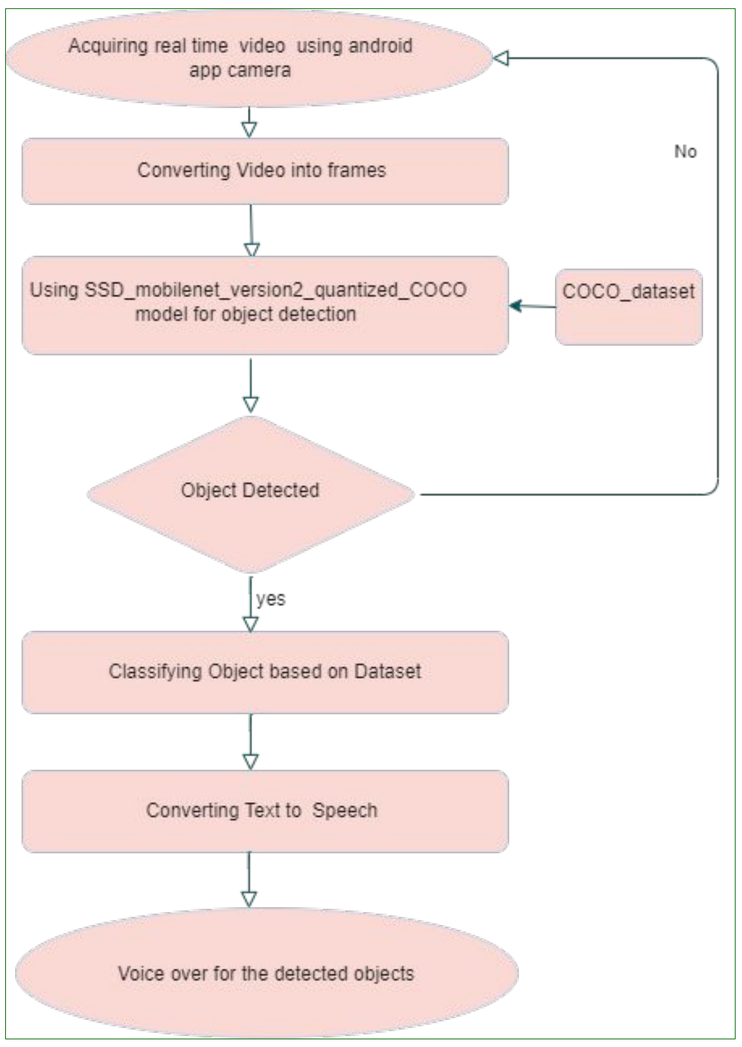
Abstract



According to the World Health Organization survey from 2021, there are 1 billion people in the world who have moderate or severe distance vision impairment. The project's main objective is to develop an android application for object detection for the blind. This is implemented for the COCO dataset, which has 80 classes of objects that the user will probably utilize daily. Android Studio is used to make the app and the UI. The model is trained using TensorFlow and converted into TensorFlow Lite. The project is carried out using a camera device on the client's side that records live video, converts it into frames, detects the objects and it's position and gives audio output.



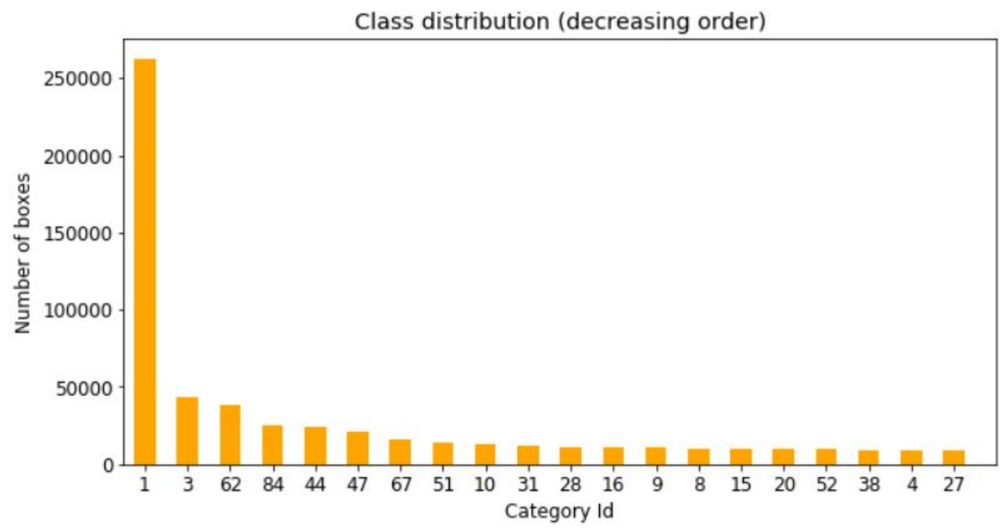
Methodology



1. The user opens the android application built using Android Studio and accesses the camera to detect the object in real-time.
2. The model converts the real-time video and divides it into a series of frames.
3. Now using the ssd_mobilenet_v2_quantized model, the objects are detected in the frames, if there are any, and classified into different classes available in the COCO dataset; if no object is found, the process keeps on running until it finds the object.
4. The detected label of an object with a spatial location is converted to speech.

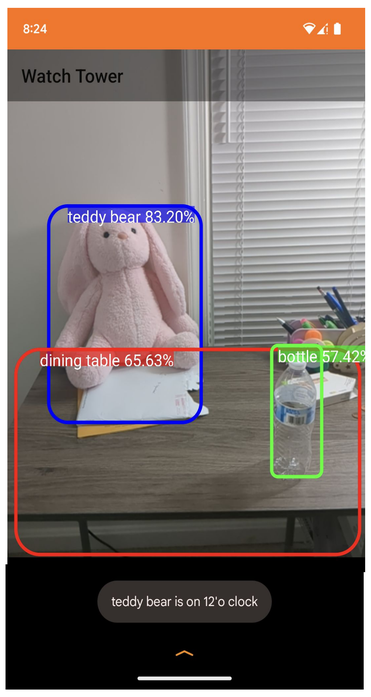
Platforms Utilized

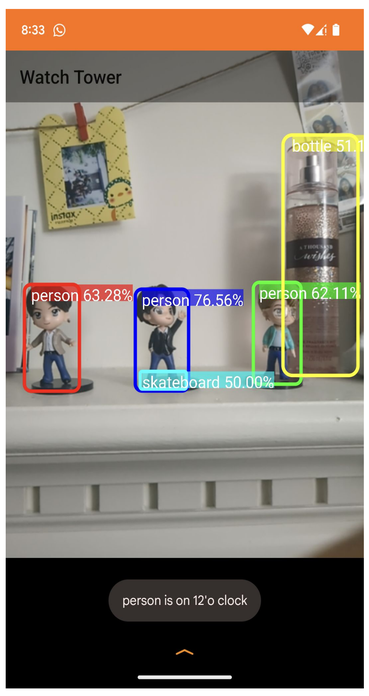
Analysis



Result

We chose the best model for our project and converted it to a TensorFlow Lite model, after which we built an Android app and integrated the TensorFlow Lite model. The developed model detects objects and their locations in the frame. It also displays the confidence of each detected object. To name a few limitations, the app can only detect 10 objects at a time and has difficulty identifying objects in low lighting. Furthermore, the app keeps detecting the same object. We added a play and stop function to the code to start and stop the audio output whenever the screen is touched.





Future Work

- Predict the situation of the surroundings based on the objects detected and their positioning.
- Quantizing the model for increasing the fps and decreasing the inference time.