**Super\_I\_Sense**

**Image recognition for specially made for Visually impaired people**

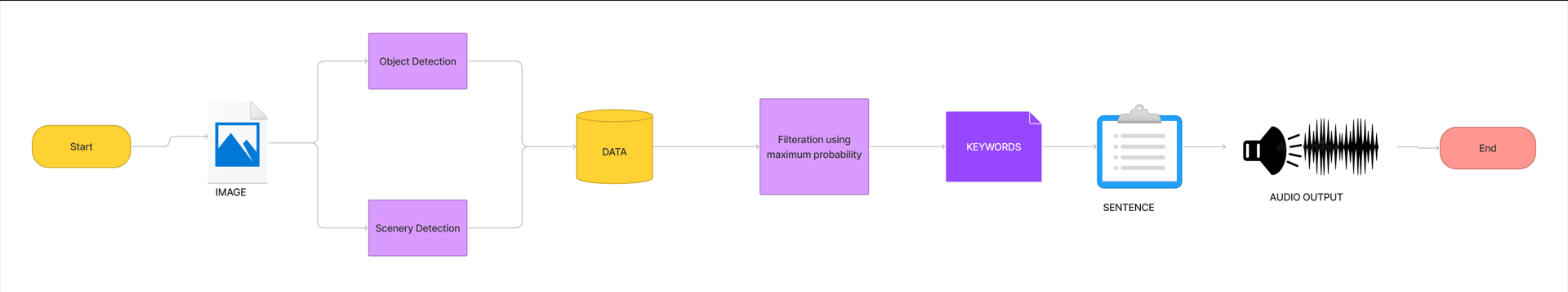
**Abstract :**

This idea is mainly focused to use technology in an efficient way and help through the betterment of the society. Visually impaired people are at a disadvantage as they are unable to perceive the necessary information. The concept of Super\_I\_Sense is to recognize the image in front of them and output the description as audio. This works by detecting the objects in the image, scenery, background. Every nook and corner of the image is being detected. The detected data is converted into keywords by filtering using the maximum probability values. Most appropriate key phrases of the image are obtained. Now these key phrases are formed into apt sentences thereby converting into speech. Many interfaces available today have difficulty in recognizing the image, they can only voice-out the filename and not the contents. “Super\_I\_Sense” can be integrated with the talkback feature available in today’s gadgets. This not only concentrates on classifying different images but also to precisely estimates the concepts and the locations of objects contained in the image. When a Visually impaired person gets to access Super\_I\_Sense , they can perceive their surroundings and feel included, relieving them from their isolated world.

**Introduction:**

Super\_I\_Sense represents a method where a blind person can get information about the shape of an image through speech signal. In this paper we have proposed an algorithm for image recognition by speech sound. Blind people face a number of challenges when interacting with their environments because so much information is encoded visually. The proposed method enables the visually impaired people to see with the help of ears. The novelty of this paper is to convert the image to sound using the methodology of edge detection.

**Workflow:**



1) Object Detection:

It detects the nook and corner of an image and recognizes objects present in it. This is done using YOLO - which is a single shot object detecting algorithm.

2)Scenary Detection:

the exact scenario is identified in which detected objects are actually present.

3)keyword Extraction:

 the detected objects and the identified scenarios are formulated into keywords by filtering the Maximum probable data.

4)Sentence formation :

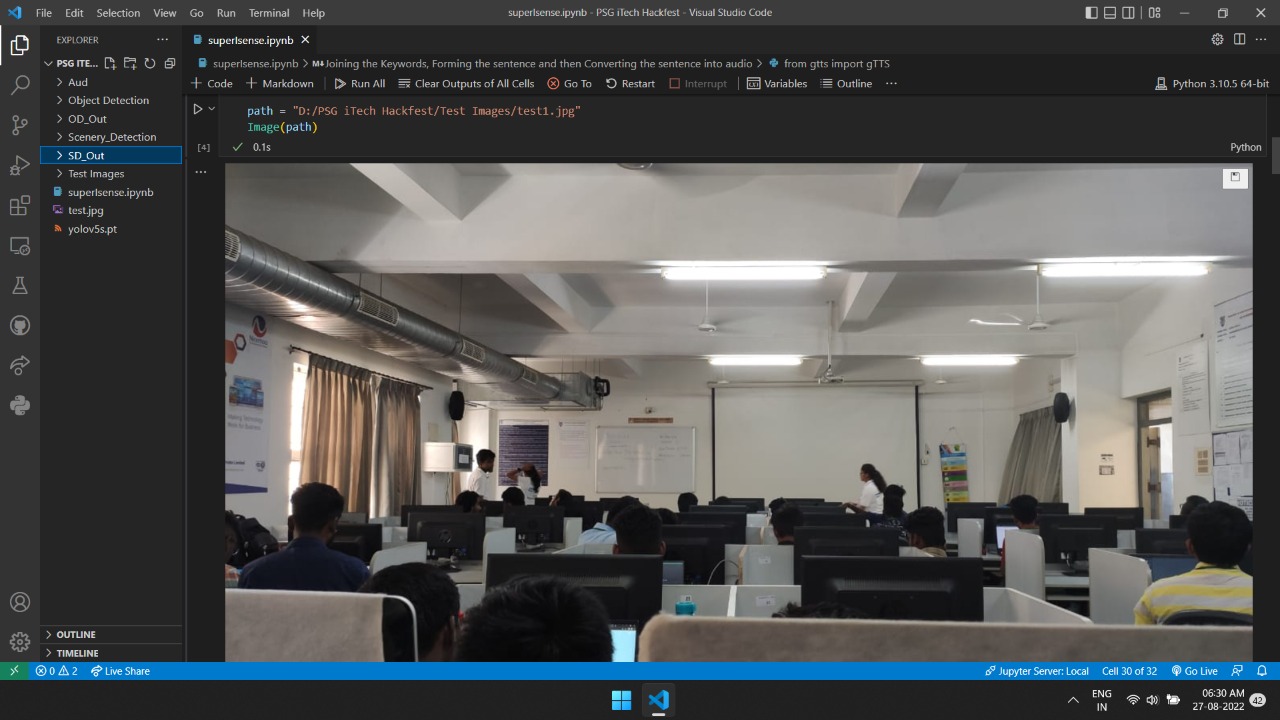
the generated keywords are converted into meaningful sentences using NLP (Natural language processing).

5)Audio Output :

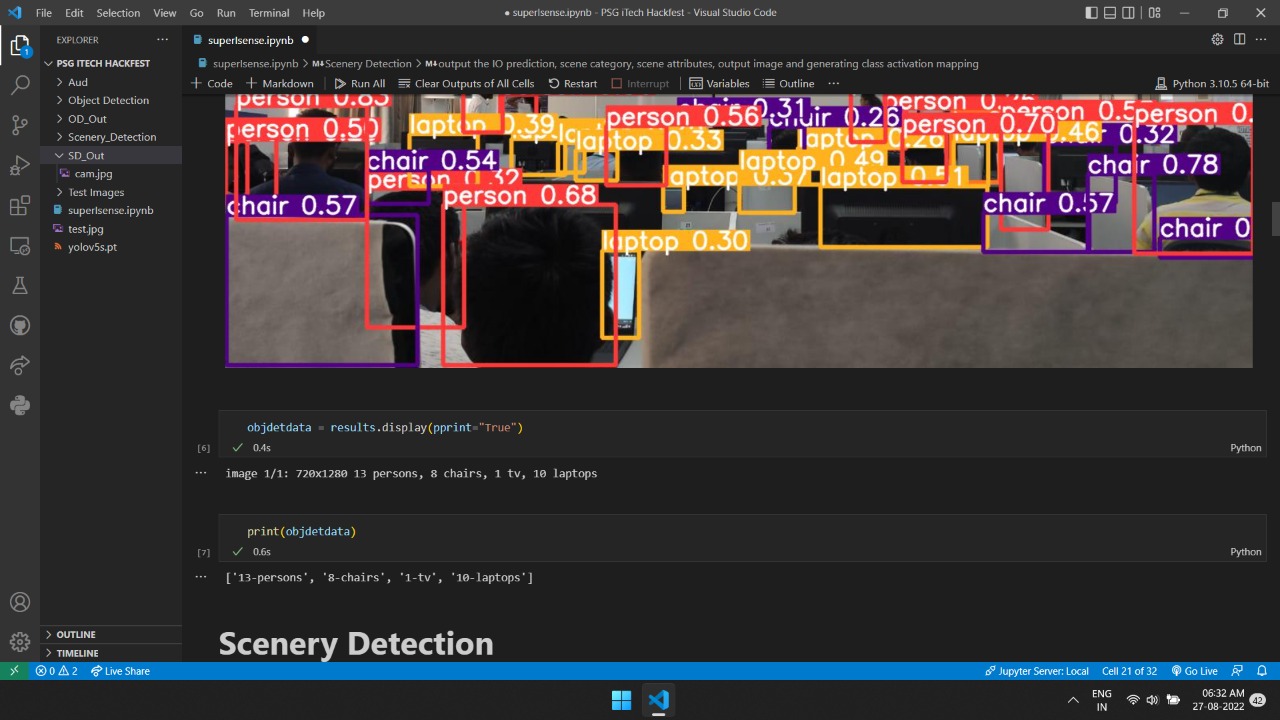
The Formed Sentences will be finally heard as speech.

**Results of each stage:**

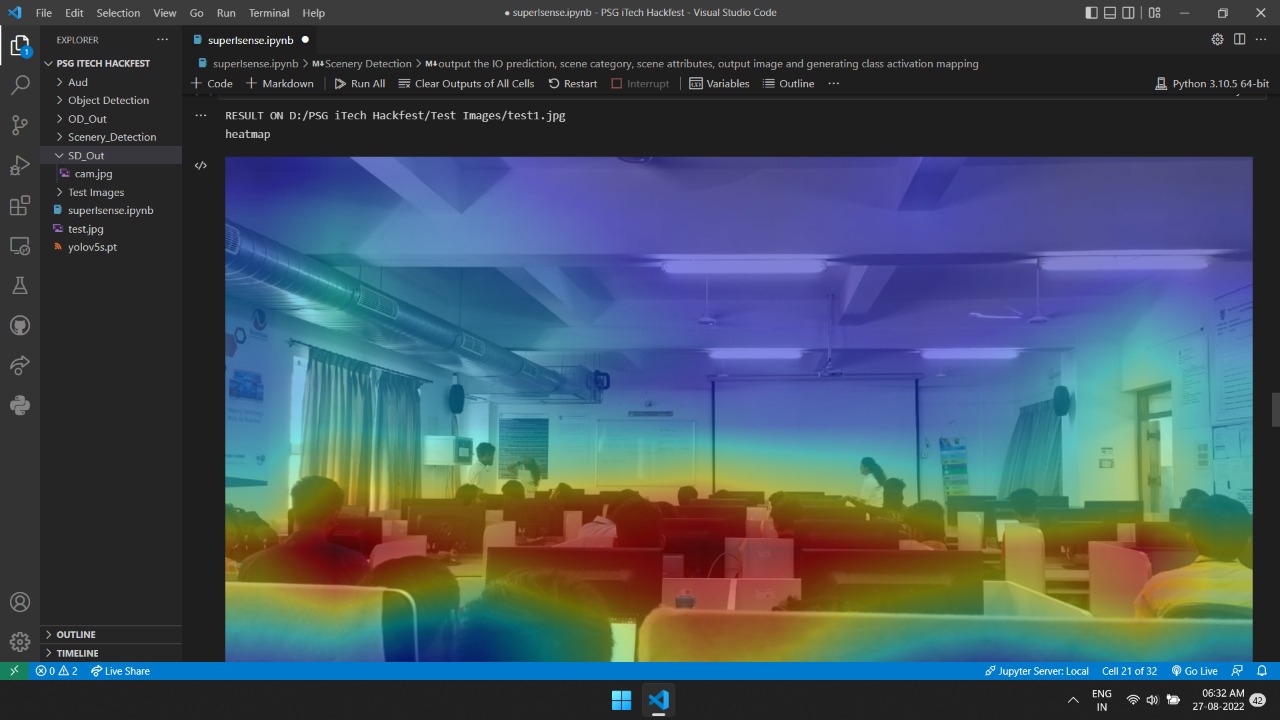
Input image



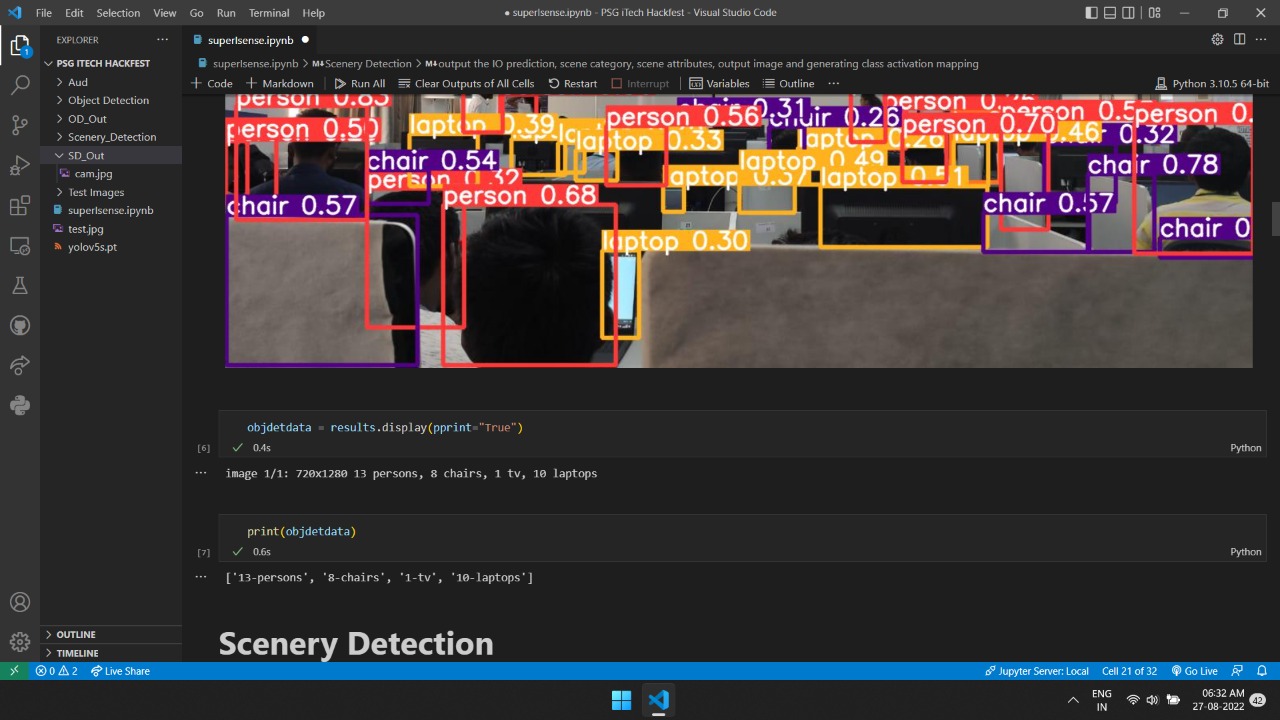
Object detection:



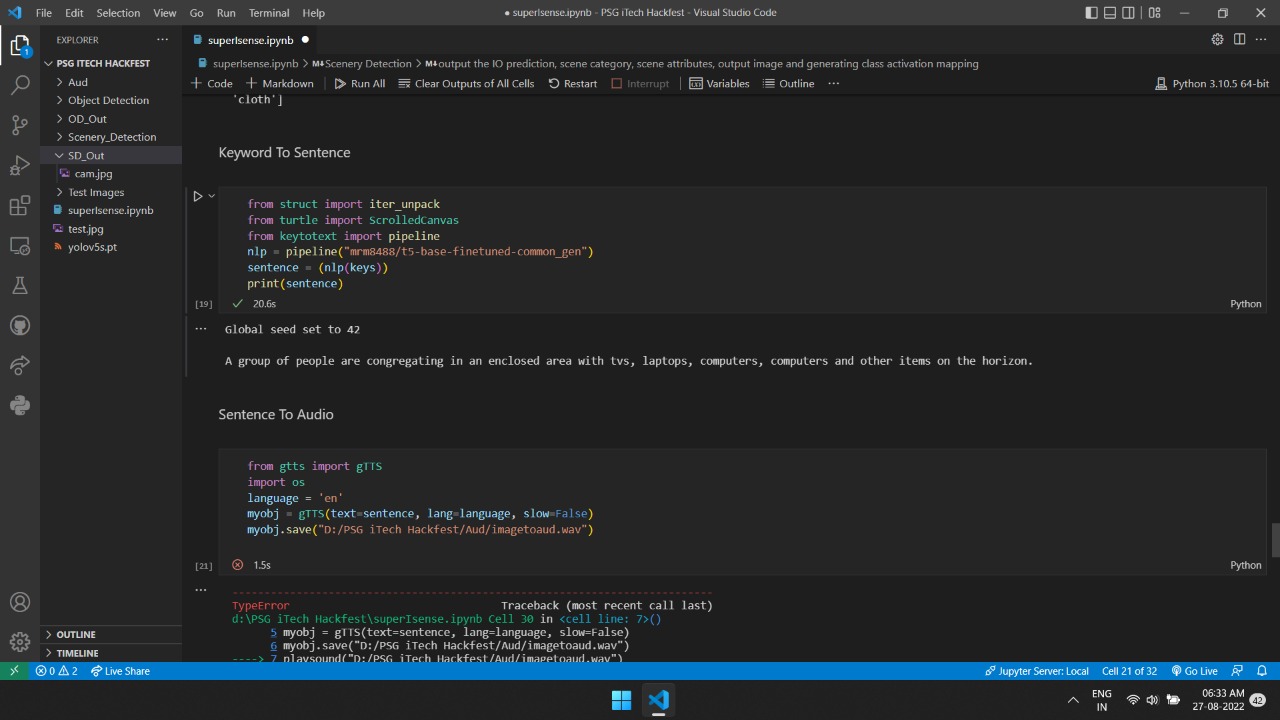
Scenery detection:



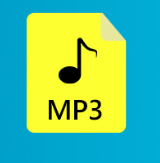
Keyword extraction:



Sentence formation:



Audio output:



Output.mp3

**Conclusion:**

From this analysis, it is therefore possible to substitute the eyes with ears to represent the physical world . Our system is being presented as a solution for the conversion of images into sound. It is designed to be portable and low power. We can see the system works great for all images . When a visually impaired person gets to access to Super\_I\_sense, they can perceive their surroundings and they can feel included which relieves them from their isolated world . So basically we are trying to use technology to help through the betterment of our society.

**References:**

Github link:

<https://github.com/phanda1326/Super_I_Sense>

demo video:

<https://drive.google.com/file/d/11zUJRBWO59fkiHrbbXCQeGM0a_PbcOqc/view?usp=sharing>

screenshots:

<https://drive.google.com/drive/folders/1t63dJThBj2WPELM4ctPlHW_iKsjLK9S6?usp=sharing>