

ASSIGNMENT REPORT

Goal: Build a program that separates text and visual elements from image.

Technologies used:

- **Programming Language:** Python
- **Graphics:** Tkinter
- **Generative AI API:** Gemini 1.5 Flash

Approach

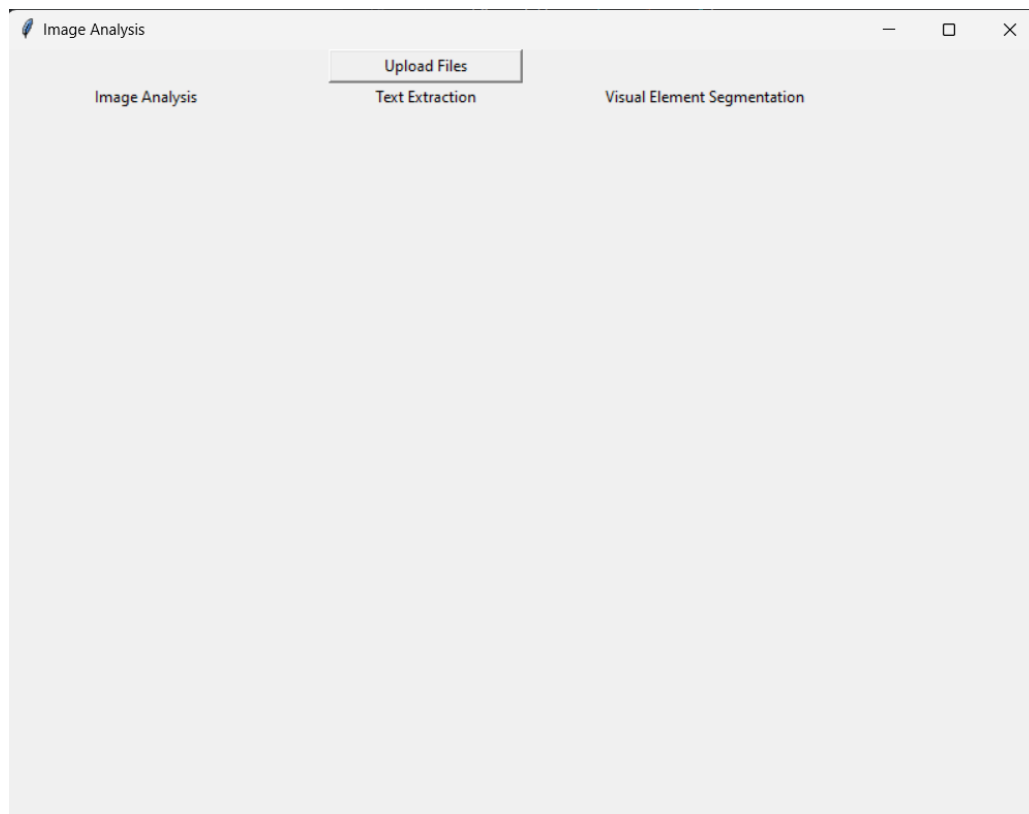
The program is coded by me in Python and I have used Tkinter for graphics.

I have created the following:

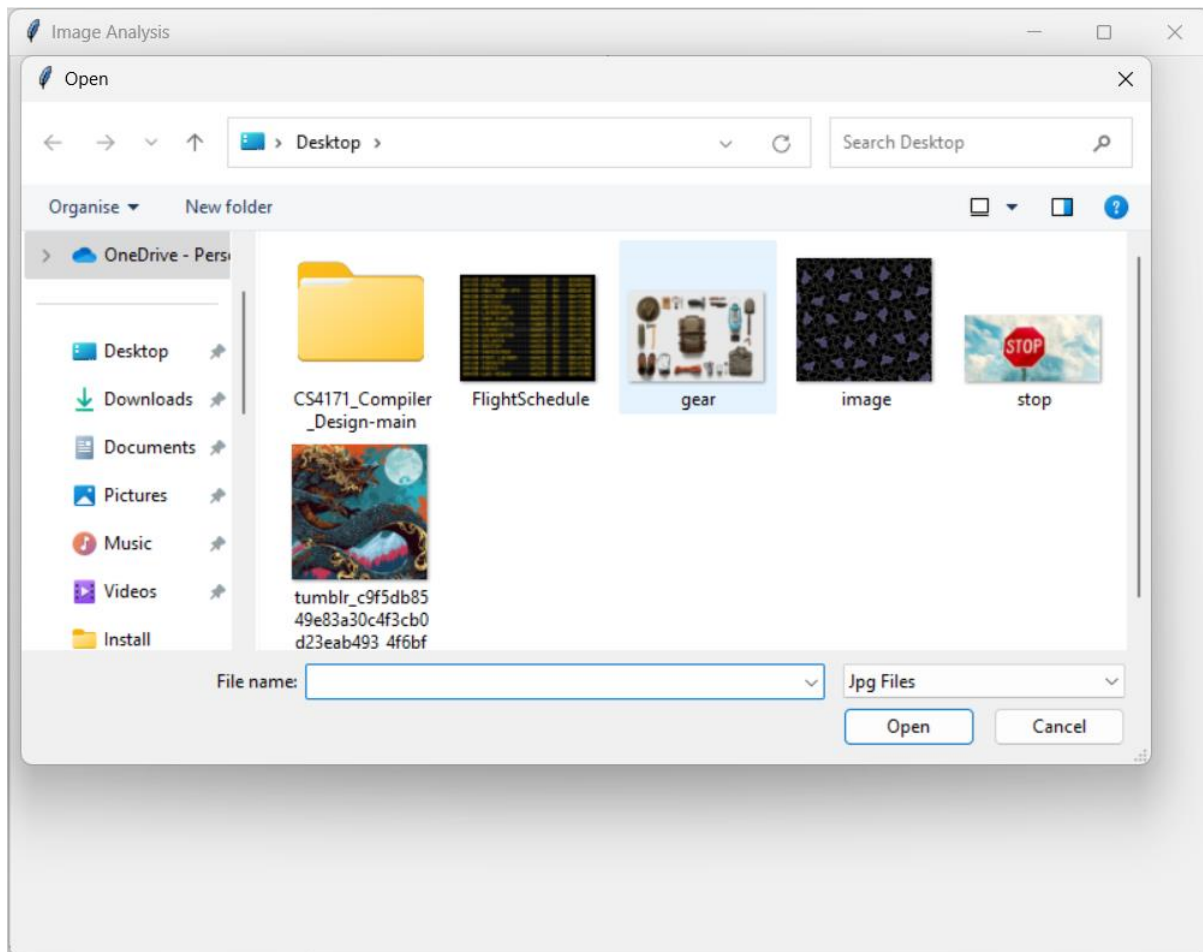
Button: 'Upload Files'

Label: 'Image Analysis', 'Text Extraction', 'Visual Element Segmentation'

I have used the grid structure for postioning.



I will be uploading an image by clicking the button. Next, it will show me the following prompt.

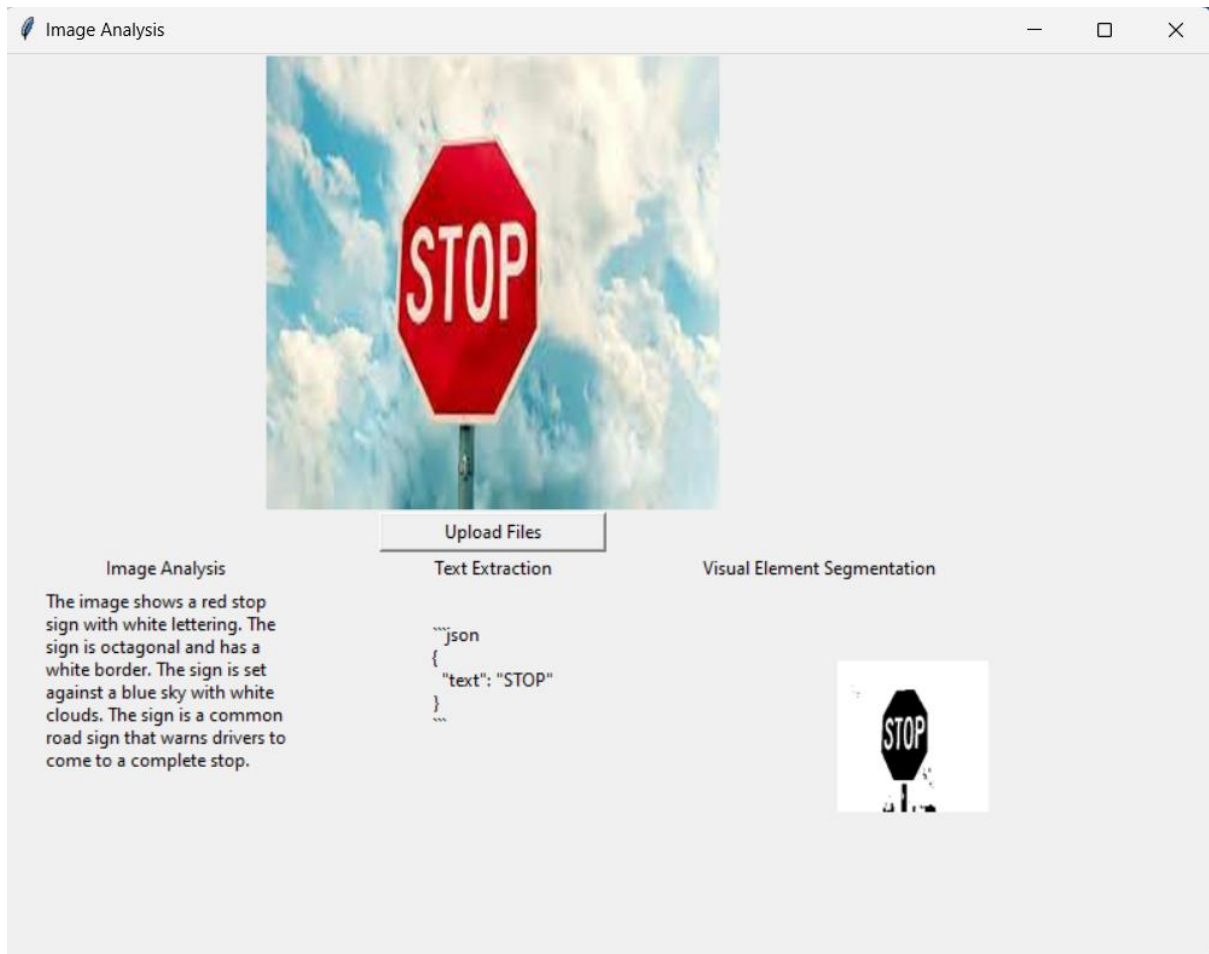


After selecting the image, the results will be displayed. I will be using google Generative AI, Gemini 1.5 Flash to create outputs.

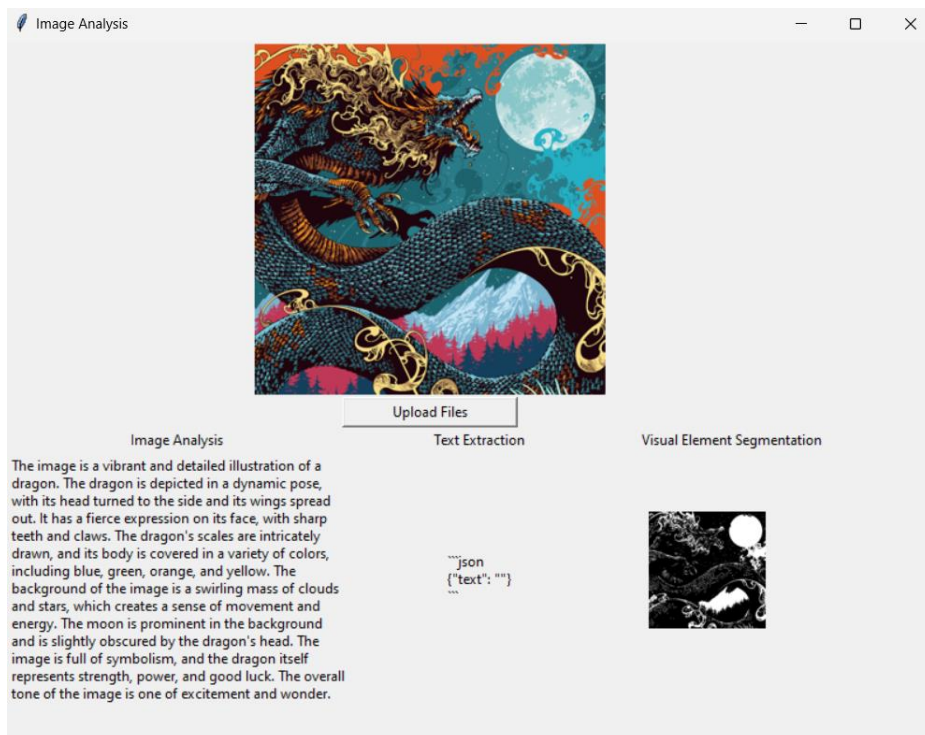
The Image Analysis section uses the following prompt: "Analyse the image" and passes the image as the parameter. The output is displayed using Message(a component in Tkinter).

The text extraction feature uses the following prompt: "Extract text from the image. Provide output as a json file" and image is passed as a parameter. The output is displayed using Message.

The Image Segmentation feature segments the image by a very basic operation such as thresholding. This is a very basic method of image segmentation.



I can switch the image by again clicking in the 'Upload Files' Button.



Challenges Encountered

While completing this assignment, I had to brush up on my previous experience of TKinter. One of the challenges I had encountered was to find an AI which was free to use and could complete the requirements.

As per the assignment, I had to use a Vision AI, however, I couldn't find any free API implementations of them, so I had to use the google generative AI, which is free and could complete the same job.