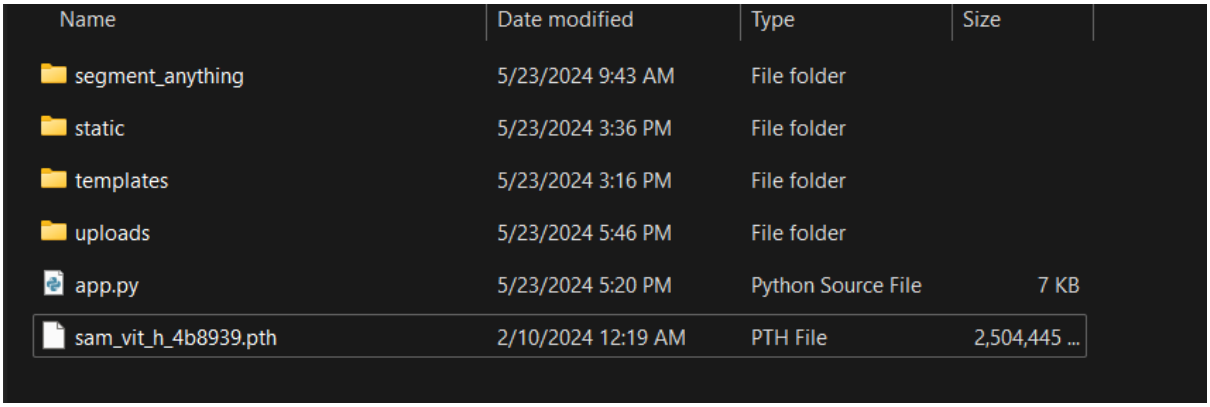
Text & Object Extractor

Install:

- Make a virtual conda environment and python version 3.8 or above.
- Install pytorch, torchvision from their attached links.
- Extract my intern zip file. And in cmd do: `cd intern`
- Clone https://github.com/facebookresearch/segment-anything and extract if zip is downloaded. Save this extracted folder into my intern folder.
- In cmd do: `cd segment-anything; pip install -e .`
- `pip install opencv-python pycocotools matplotlib onnxruntime onnx flask`
- Download the below file into intern folder: https://dl.fbaipublicfiles.com/segment_anything/sam_vit_h_4b8939.pth

Intern folder structure with all files/folders in it:



Explanation:

I used Meta's Segment Anything Model (SAM) to segment the objects and items from the image. I fine-tuned the parameters in SAM so that it's suitable for this project on trial-and-error basis method. It has features of auto segmentation, prediction, click segmentation, labelling, etc. I used the SamAutomaticMaskGenerator to generate masks for the objects which are identified and then saved the objects individually as .png files.

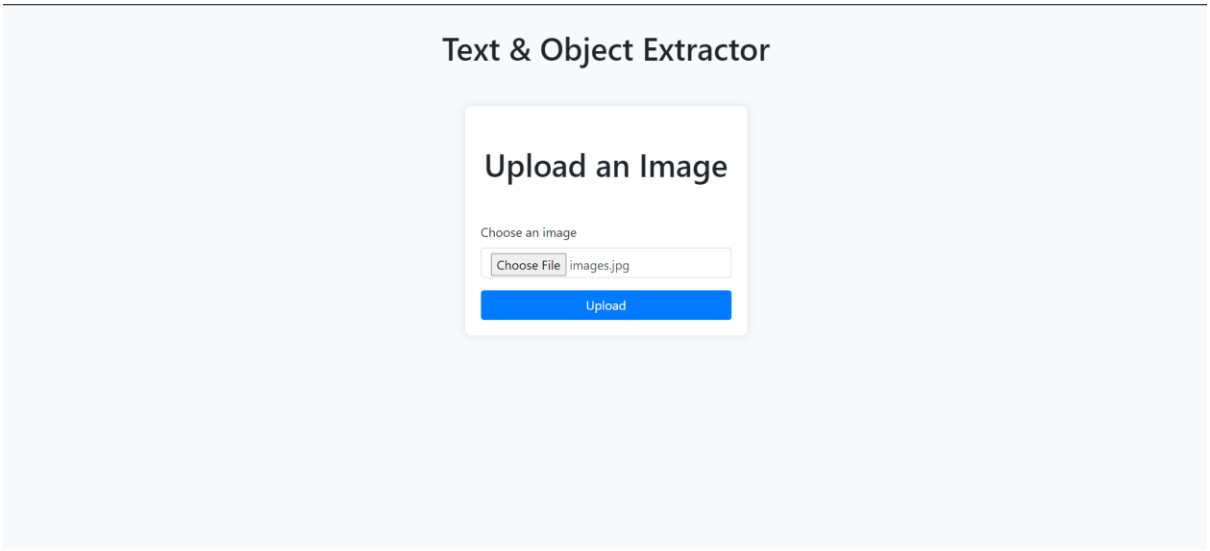
Next, I used OpenAI API to identify the text from given image. I used gpt-4o model with vision capabilities to extract the text from the image. Then refined this text using openai text completion feature to get better sentences.

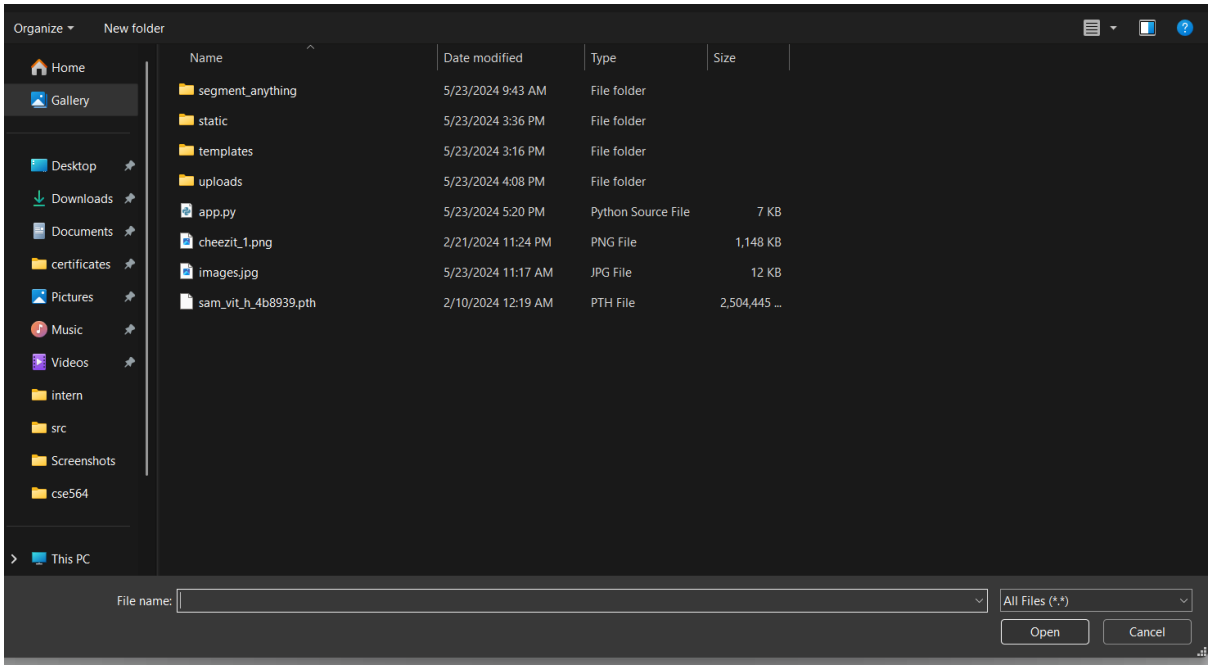
Finally build a Flask web application where user can give input image and it's processed and the output is the extracted text along with the individual pics of objects in the given image.

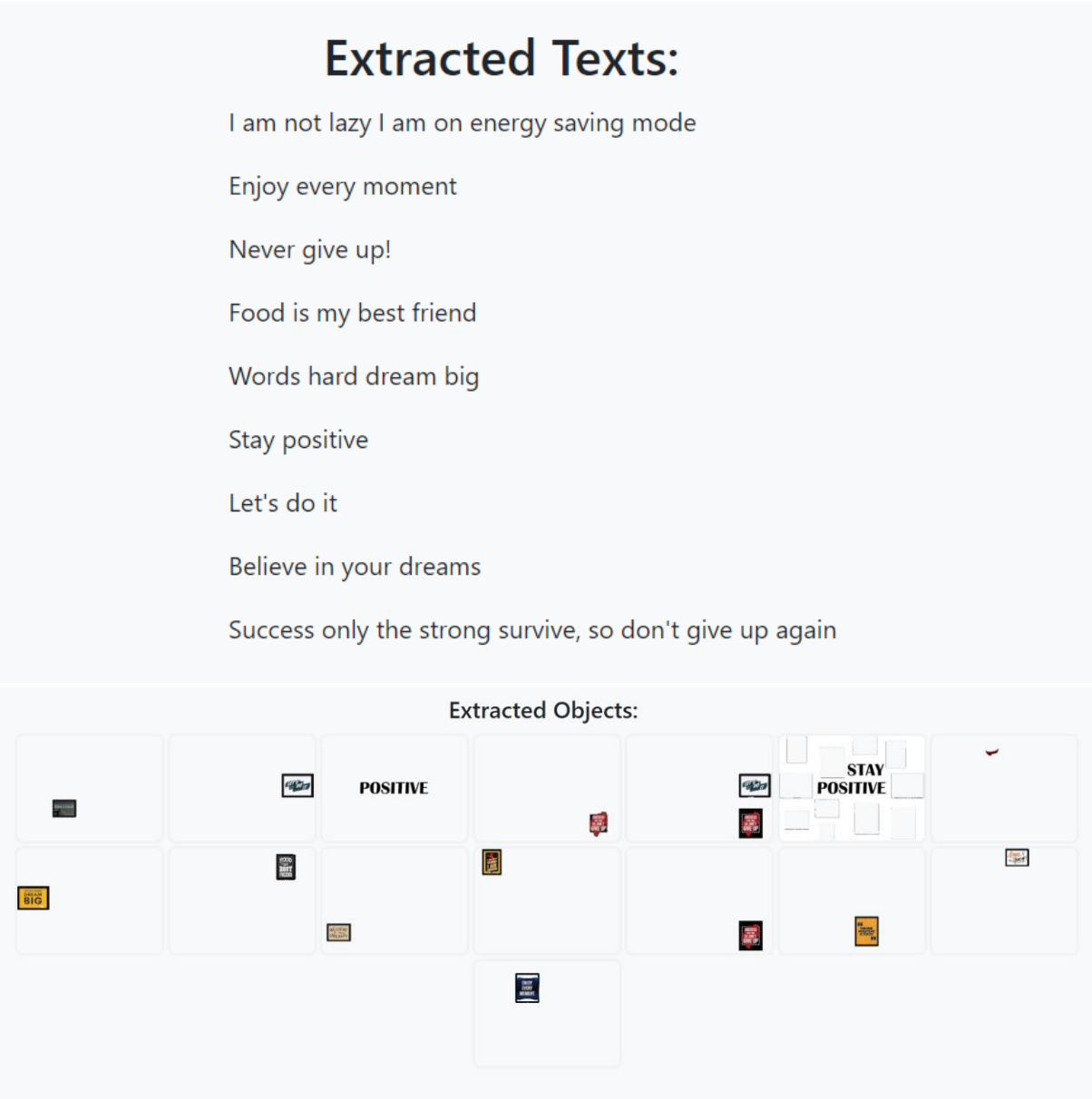
Running:

- Activate the conda environment
- Run program: python app.py
- Go to webpage: http://127.0.0.1:5000/
- Choose an image with any objects or text in it.
- Click on upload button.

Demo Usage pics:




Organize
New folder
Home
Gallery
Desktop
Downloads
Documents
certificates
Pictures
Music
Videos
intern
src
Screenshots
cse564
This PC
Name
Date modified
Type
Size
segment_anything
5/23/2024 9:43 AM
File folder
static
5/23/2024 3:36 PM
File folder
templates
5/23/2024 3:16 PM
File folder
uploads
5/23/2024 4:08 PM
File folder
app.py
5/23/2024 5:20 PM
Python Source File
7 KB
cheezit_1.png
2/21/2024 11:24 PM
PNG File
1,148 KB
images.jpg
5/23/2024 11:17 AM
JPG File
12 KB
sam_vit_h_4b8939.pth
2/10/2024 12:19 AM
PTH File
2,504,445 ...
File name:
All Files (*.*)
Open
Cancel



Challenges faced:

Not much, whole project done in 4hrs 30min from scratch.

Have previous experience in using OpenAI API for my projects in master's and free time.

Have previous experience using META Segment Anything Model in master's degree where I am currently working as a research assistant.

Can improve it or add extra features if required.

References:

- https://github.com/facebookresearch/segment-anything
- https://platform.openai.com/docs/guides/vision