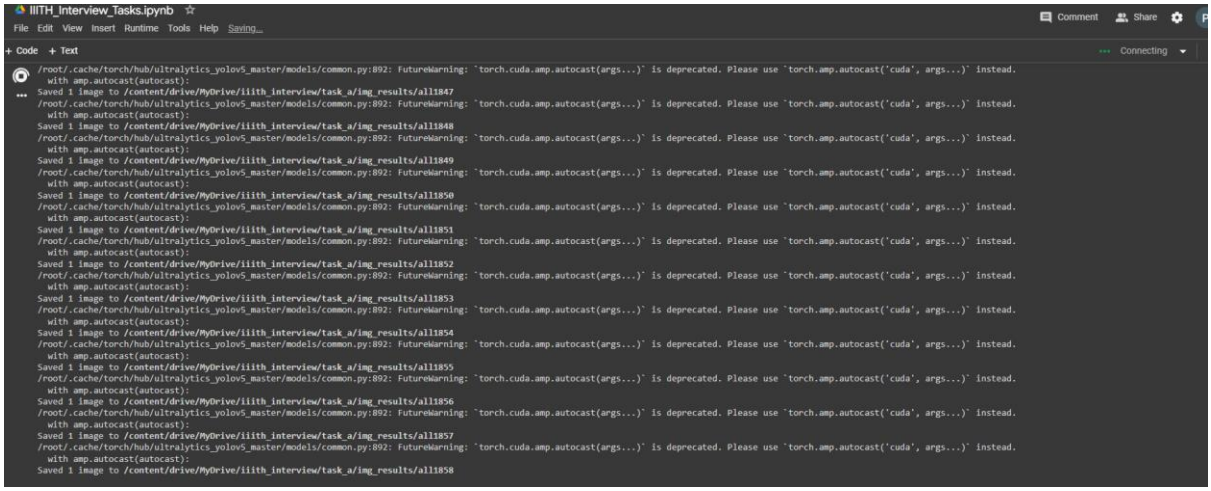
Task Report

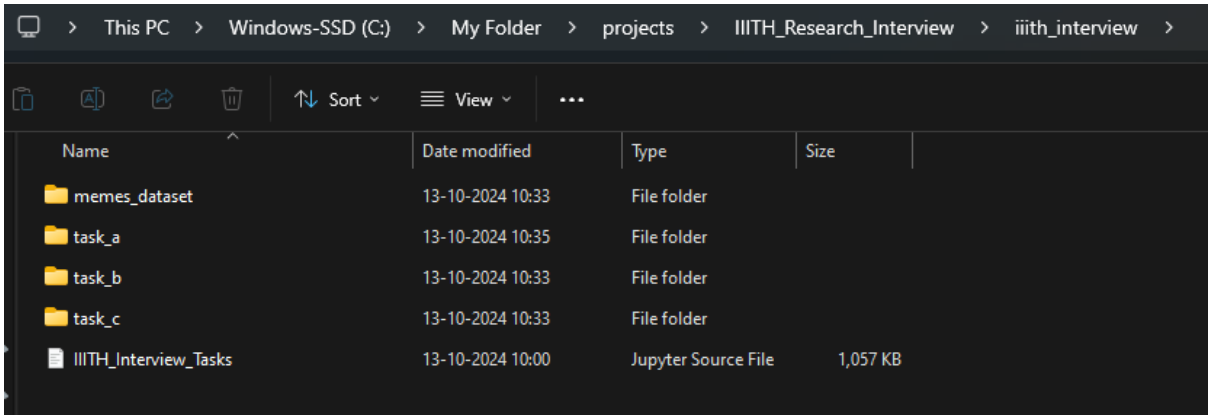
Project – Analysing Hateful Memes

I have taken this project, since I had some familiarity with the subject. It took me around 11 hours to upload the images to the drive. I am using Google Colab for this task. I have been trying to run the code for the past 3 days, but somehow it gets stuck at 1803rd or sometimes at 5013th image, but it is not reaching its conclusion.



Therefore, I have taken a smaller subset of 10 images, to make sure that each of the tasks are completed on the first part. So, this report shows how each task has been performed with respect to the smaller subset of the original data.

Directory Structure



- The subset of 10 images has been stored in /iiith_interview/memes_dataset.
- Task A results has been saved to /iiith_interview/task_a/img_10_results. Since this was the first task, and hence, object detection results were to get saved. They have saved in /iiith_interview/task_a/img_10_results/all. Further, all the csv and txt files are also saved in the same folder.
- Task B results has been saved to /iiith_interview/task_b/img_10_results.
- Task C results has been saved to /iiith_interview/task_b/img_10_results.

- The most important python file, where all the computations were performed IIITH_Interview_Tasks.ipynb file is also present under /iiih_interview.

TASK A – Object Detection

- To perform the object detection task on the meme's dataset from the Hateful Memes Dataset, I have used YOLOv5 model. This model is one of the best open-source vision AI models, further, it is fast, accurate, and easy to use. And finally, it is easy to integrate in Google Colab.
- I have uploaded the smaller subset of data to the Google Drive. Installed the YOLOv5 repo from the GitHub.
- Next, I specified the output directory. And checked one image by displaying it so that I am certain I have reached the right directory.
- Further, I ran the object detection model on 10 images and stored it in the output directory. Also, this is the place where my computer is hanging each time, when I am running the same code on the original dataset.
- Once, I had all the results, I imported Pandas library to for creating csv files and extracting the count of each object from the results in the object_counts.csv.
- Here is the frequency table of the 10 images dataset –

```
Frequency of Detected Objects:
  Object Name  Count
0      person     11
1       sheep      9
2        bear      1
3         tie      1
4  teddy bear      1
5         dog      1
```

- Further, we have plotted a bar chart to show the relative number of objects detected, which is saved in the output directory.

TASK A – BONUS

- In the bonus task, I had to label the images as toxic or non-toxic. Then develop a simple classification system using this catalogue.
- I have used a very simple classification system to label each object as either toxic (1) or non-toxic (0).
- Defined a function, with certain objects as toxic in a list. Iterated over the object_counts.csv file and created a new column "toxicity." If the object was toxic, it was assigned 1 and if non-toxic, the toxicity score of 0 was assigned.

```
Updated Data with Toxicity Labels:
  Object Name  Count  toxicity
0      person     11         0
1       sheep      9         1
2        bear      1         0
3         tie      1         1
4  teddy bear      1         0
```

- Finally, saved the new results to the output directory.
- After this part, I am stuck on the part of how to assign the toxicity score to an image, what threshold to take. Since, there might be images which has both sheep and a person, now what basis do I have to label it. Hence, moved to the next task.

TASK B – Caption Impact Assessment

- In this task, I had to analyse whether or not the text overlay on the images influences the object detection process.
- The approach for qualitative analysis – I have visually examined the object detection results before and after applying the caption hiding techniques. I will explain later what process I have used for the caption hiding.
- For quantitative analysis, I have used simple metrics to measure the performance.
    1. Number of objects detected before and after masking the captions.
    2. Object confidence score
- For computing the object detection, we are using the same model (YOLOv5) on both the original and caption-removed images.
- I have used pytesseract, which is a python OCR tool. It helps to recognize and read the text embedded in images.
- Once the mask is identified in the text regions, we have used OpenCV's inpainting technique to remove the text by filling the detected regions.
- As per the analysis, the average number of objects detected per image with caption is 2.4. After inpainting the captions, the average drops to 1.5. Hence, so does the average confidence. So, the total number of objects that I lost after removing the captions is 9.

image_path	num_objects_orig	num_objects_inpain	mean_conf_orig	mean_conf_inpaint
/content/drive/MyDri	2	1	0.3840302974	0.3566207588
/content/drive/MyDri	2	0	0.4182769656	0
/content/drive/MyDri	2	1	0.6321300566	0.4366572201
/content/drive/MyDri	1	2	0.4416429996	0.4562041312
/content/drive/MyDri	1	1	0.7626440525	0.6685587168
/content/drive/MyDri	2	0	0.3385677338	0
/content/drive/MyDri	1	0	0.773912251	0
/content/drive/MyDri	0	0	0	0
/content/drive/MyDri	12	9	0.5820193763	0.4914768007
/content/drive/MyDri	1	1	0.8944702744	0.3040029407

- So, the result clearly demonstrates how the captions are hindering the object detection. While inpainting to remove the text, it has introduced distortions, thus, reducing the model's ability to detect the objects accurately.
- Even in the small dataset of 10 images, the average confidence score of the model has dropped significantly from 52.8% to 27.1%.

TASK C – Classification System Development

- I had to develop a system that classify the images based on something non-trivial. Since the given dataset is already too large, I am not considering any other dataset in the give task.

- System on which I want to classify the dataset - I am using binary classification methodology. If the given meme is a having at least 1 person, then it will be to one class. On the other hand, if a given meme does not have a single person. Then, it will belong to the second class.
- For each image, we have first of all detected the objects using the same YOLOv5 model. If the person detection confidence is greater than 0.5, I have assigned such images to class 1, otherwise, class 2.
- I have built a simple classification report for the same. Since the current working dataset is too small, I have reached accuracy score of 1.

```
Accuracy: 1.00
Classification Report:
                           precision    recall  f1-score   support

   Class 1: Person Detected       1.00      1.00      1.00         5
Class 2: No Person Detected       1.00      1.00      1.00         5

                  accuracy                           1.00        10
                 macro avg       1.00      1.00      1.00        10
              weighted avg       1.00      1.00      1.00        10
```

- Further, due to small dataset, I have checked the intricate results as well by manually checking all the images.
- Further, I have used Standard Scaler to normalize the data on the same scale and simple Logistic Regression to build a classification system.
- Since, the test size of the dataset has been chosen as 0.2 (of 10), the support for the system is too low.

```
Accuracy: 1.00
Classification Report:
                           precision    recall  f1-score   support

   Class 1: Person Detected       1.00      1.00      1.00         1
Class 2: No Person Detected       1.00      1.00      1.00         1

                  accuracy                           1.00         2
                 macro avg       1.00      1.00      1.00         2
              weighted avg       1.00      1.00      1.00         2
```

- But I am quite confident, we will have positive results in the case of entire dataset as well.

Paper Reading Task

- For this task, I downloaded the paper from the given website.
- Built a simple presentation with 3 slides containing the 3 major strengths, 3 major weaknesses, and 3 scopes of improvement in the give paper.