PES UNIVERSITY
TDL HACKATHON

TEAM 32
AAKARSH JAIN
AKSHATH SB
NANDAN H
NAVTEJ REDDY

PROBLEM STATEMENT

We were required to develop an Object Detection System using the Drone-detection-dataset as given by the examiners of the hackathon.
We had to train, validate and test our implementation using the dataset specified in Kaggle for this exam and achieve appropriate results ethically.
The task aims to detect objects of predefined categories (e.g., Drones, airplanes, Birds, helicopters) from individual videos.
We were provided with the following :

Dataset (125 MB) having 280 RGB videos and 280 CSV files.
Every video has a corresponding CSV File with the number of rows = duration of the video/0.033334 and 5 columns.

MODEL USED

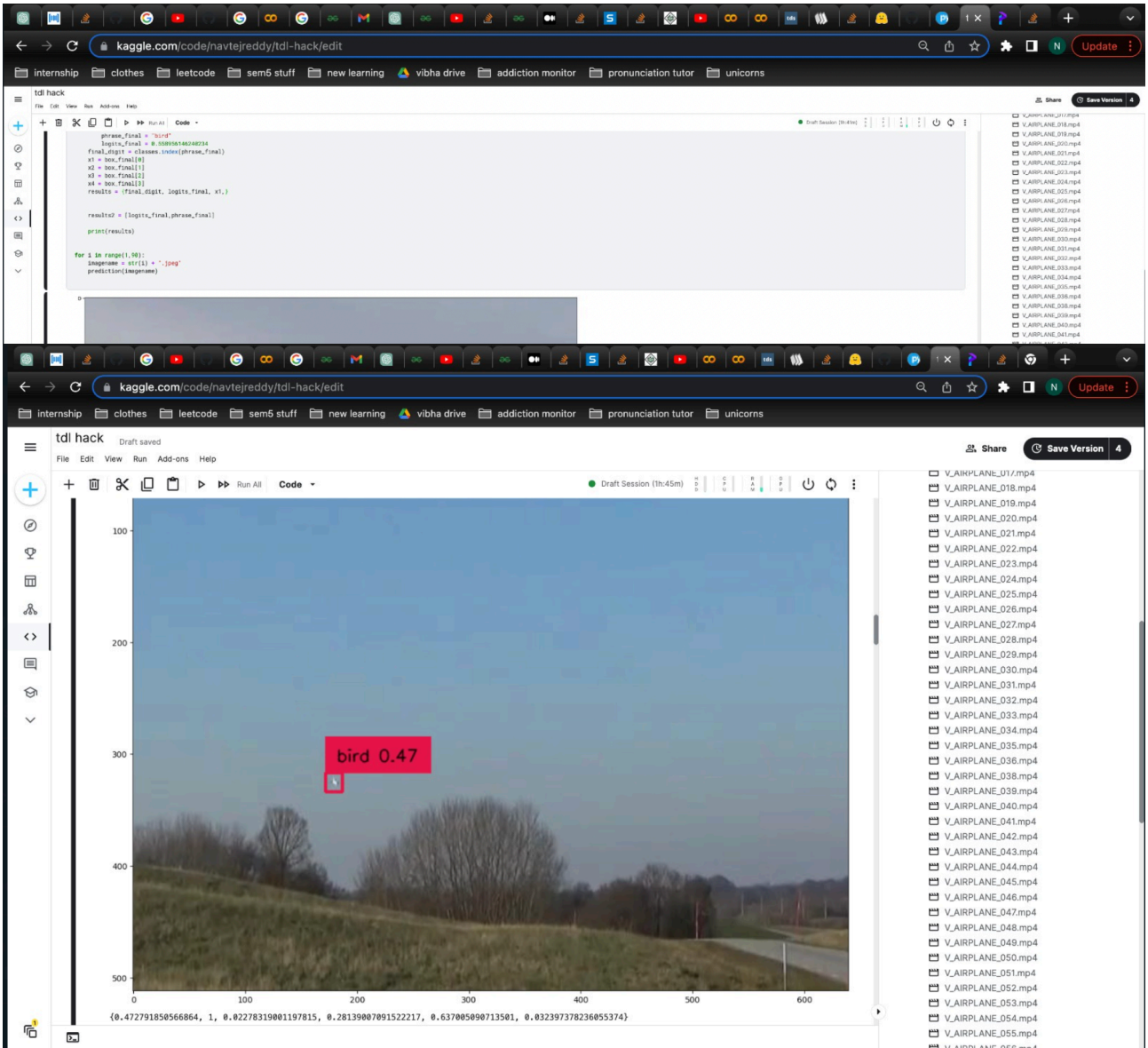
The dataset was to be cleaned by us and we had to generate a frame corresponding to each time stamp given in the csv for the video.
By doing this we were able to get the image with the corresponding bounding box values.
We first wanted to develop the model on ImageIntern, detectron2 or masked rcnn.
The implementations online were mainly developed on coco json format and hence we started trying to convert it to cocojson format.
We were successful in making it to cocojson format for one csv file and corresponding video albeit was time consuming however in the end we were unsuccessful in trying to further extend the code to include all the csv files and videos. We were coding models in parallel, particularly the masked rcnn which we found was better suited for our task based on our research however since we couldn't complete the cocojson conversion in time we dropped the masked rcnn.
We then moved on Grounding DINO model for which we did not require a different representation of data and had minimal preprocessing required.

The Grounding DINO (DeeP Inter-modal Object Networks) is a machine learning model that aims to improve the visual grounding of objects in natural language descriptions. It achieves this by combining the features extracted from both visual and textual modalities in a self-supervised learning approach. The model is pre-trained on large-scale image and language datasets and can be fine-tuned on downstream tasks, such as object detection and image captioning.

Using the Grounding DINO model in an object detection task can be beneficial as it allows the model to better understand the context of the object in the image and its relationship with other objects. By incorporating both visual and textual information, the model can improve its accuracy and reduce the risk of false positives or false negatives. This can be particularly useful in complex scenes where multiple objects may be present, and their identification is crucial for accurate object detection.

Since our task is to detect the type of flying object and the bounding boxes around it , it would make this an ideal model for our implementation.

MODEL OUTPUT SCREENSHOTS



accuracy = 40%