

# PEPSI AND COCACOLA LOGO DETECTION MODEL-

## Approach Document for Video Processing and Object Detection Pipeline.

### Introduction-

This document explains the approach which I used to develop the pipeline for logo detection for brands of Pepsi and CocaCola. Simply, It has 3 parts:

- Training the model on the dataset of pepsi and Cocacola.
- Detection of the logo of Pepsi and Cocacola in the input video file.
- Giving the timestamps of occurrence of the above 2 logos in a json file.

### Overview of the Workflow:

- **Understanding the problem statement**-The problem statement was to detect the logo of pepsi and cocacola in the input video and return the timestamp of their occurrence in a json file.
- **Data gathering**- Firstly I tried to gather images from my own from Google and created a folder but that was a hectic process. Then I went to kaggle but didn't find relevant datasets. Then I searched on Google and found about roboflow. Then I gathered

datasets of pepsi and cocacola online from roboflow website.

- **Data Analysis-** I analysed the images in the datasets and found out that some images were misleading for the model training so I removed them.
- **Data Cleaning-** Since I found that some images were misleading so I removed them. Although there was not much of task for data cleaning because the online dataset already gave 3 folders of test, train and valid.
- **Model Selection-** The model is selected through the AUC curve, firstly I was working with YOLOV5 but then I shifted to YOLOV8 for improved performance, new features, and better python library support. In yolov5 I had to clone the repo for training which was hectic. I didn't use yolov2 because it is a older version and new versions are available in the market (Why I chose yolov8 over tensorflow- since I am a beginner and tensorflow and need the prototype quickly, it has a format for dataset like (<class\_id>, <x\_center>, <y\_center> <width> <height>) it is easy to setup for object detection. If I had to do something beyond object detection or my project would require complex or custom neural network architectures, I would have used tensorflow).
- **Training Data-** Then I trained the model from the online datasets available.
- **Object Detection, Timestamps and size and distance from the centre of the frame.**

## DETAILS OF THE WORKFLOW:

- Went to kaggle in search of dataset for cocacola and pepsi but could not found one so went to roboflow as suggested by Google.
- Downloaded the dataset of pepsi and cocacola logos from roboflow website.
- Counted the number of images on Jupyter notebook in all 3 folders of test,train and valid in respective pepsi and cocacola datasets for cross verification in the later stage.
- Now the class\_id of labels of both the datasets were 0 so I first converted class\_id of cocacola to 1.
- Then to train the model I transferred the whole dataset of cocacola to pepsi i.e it got combined in a single folder.
- **Training the model-**
  1. Installed ultralytics to import yolo for training the model.
  2. Changed the data.yaml file according to the class\_id with the corrected spelling.
  3. Trained the model with yolov8n not yolov8s under T4 GPU with epochs as 100 for good accuracy since the dataset was huge as it approximately consists of 1500 images.
  4. The model took approximately 1 hr to train and gave runs/detect/train.It had many things like graphs and matrices of all the training model like Precision-Recall Curve,F1 Confidence

curve, Precision Confidence curve , Recall Confidence curve, Confusion matrix etc.

5. Then I downloaded the best.pt( typically a checkpoint file that contains the model weights which performed best on the validation dataset during training) from runs/detect/train/weights.
6. Then downloaded the input youtube video of pepsi and cocacola war on system with the help of website [www.ssyoutube.com](http://www.ssyoutube.com) in .mp4 type.
7. Using the trained model(best.pt) to detect the input video file of pepsi and cocacola in each frame and generating the timestamps of the occurrences at which the logos appear and storing it in a json file.
8. Then screen recorded the output video file via loom.results1

## **MISTAKES:**

- Firstly, I cloned the yolov5 from ultralytics on vscode. I should have done proper research and used yolov8 because of better performance.
- With the libraries of av, pytorch, opencv-python... I was tried to train my model locally... that too on CPU.... that too on yolov5s not yolov5n. Then I shifted to Jupyter and yolov8.
- I didn't changed the class id of one of the dataset from 0 to 1.
- The dataset which I chose first for cocacola and pepsi were detecting their bottles and cans instead of their logos.

- In Jupyter also,I was trying to train the model without GPU.
- In labels of data.yaml I was typing the spelling of CocaCola as Cocacola which resulted in non-detection of CocaCola.

