

Bosch Assignment - Nikhil Singh

Data Analysis

Observations

1. Data Imbalance: There is a significant data imbalance across many parameters in the training and validation data.
 - a. Day and night time data is more prominent than dawn/dusk.
 - b. Clear skies overshadow other types of weather variations with foggy being close to non-existent.
 - c. City street, highway and residential have reasonable data-points while other classes are present in marginal numbers.
 - d. Classes have a significant data imbalance with classes like cars, traffic signs, traffic lights, and person having relatively more instances than other classes.
 - e. Occluded cars are almost double in number than not occluded, this is because of the cars that are present further away in the lane and hence occluded by the car or any other vehicle present between the observer and that car.
2. Image Quality in the dataset: The images in the dataset sometimes suffer from glare coming from the sun or headlights of other vehicles coming from the front. This results in a loss in information for the model and may result in missed predictions.
3. Bounding box sizes: The bounding box sizes vary across the classes, for example -
 - a. A car present immediately ahead of the recording vehicle will have a bounding box that almost covers the entire frame while one in the distance will only span across a few pixels.
 - b. For the above reason, people, cars and vehicles in general have a higher variation in the sizes of the bounding boxes.
 - c. Traffic lights and signs have relatively smaller sizes as compared to the frame size.

Training

Model Choice

I've used YOLOv11m for this task because of my previous experience with YOLO and the Ultralytics library, which makes it easy for fast prototyping. The task here is object detection for possible use in some sort of driver assistance, in which case, latency will play a huge role. YOLO family of detection models have always been the most obvious choice in low latency use cases.

Experimentation

Three iterations of the model were trained for the task. Default values for training and augmentations from ultralytics were chosen for experiments.

1. Entire dataset, including the occluded images and truncated images.

2. Entire dataset, excluding the occluded and truncated images. *(The assumption here is that, any object that is occluded is behind something else and our priority in a driver assistance system would be to focus on the nearest object. In case of truncated objects, the idea is that they're partially out of frame, which in most cases indicate that they're out of the path of the vehicle.)*
3. Class agnostic training, model detects objects from the background. This experiment also does not include occluded and truncated objects.

Other Possible Experiments

Few other possible experimentations that were not done due to limited time but could lead to better performance in detection on this kind of dataset.

1. Training different models for objects of different sizes i.e. traffic signs, lights, etc are detected by one model, another model for vehicles and another one for rider, bike, person, etc.
2. Splitting the dataset for static and moving items in the frame and training separate models for each. Fixed items such as traffic signs, light, road markings are handled by one model while another one for vehicles and pedestrians.

Evaluation

Quantitative Method

For quantifying the model performance, we can take the mAP on the validation set along with a confusion matrix for all the classes. This can be used to compare various models and approaches.

Qualitative Method

Evaluating the model performance qualitatively is tricky. The model predictions can be broadly classified into

