Submission Template
Github Link:
https://github.com/hablu1/Udacity_SDC_Project1.git

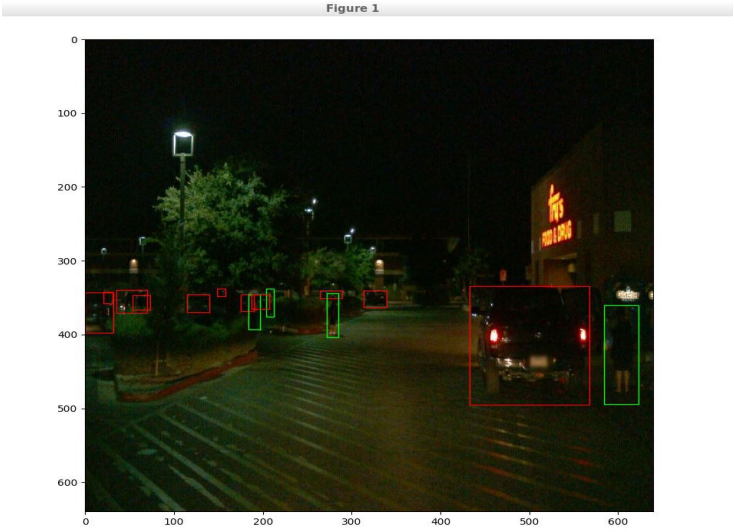
Project overview
This section should contain a brief description of the project and what we are trying to achieve. Why is object detection such an important component of self driving car systems?

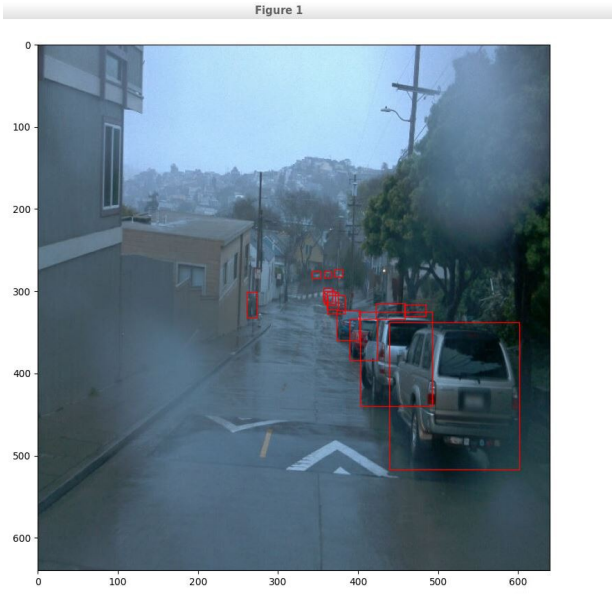
In the first project the task is to detect object in images (cars, byciclists, pedestrians). The Waymo open dataset serves as training and validation data. The goal is to first analyze the data and then split the data into a training, test and calidation set. Further, the model is trained on the training data and model performance is evaluated on the evalutation split. Different config files are used to analyze the influnce of the adaptions to the config file as the config file is the main configuration in tensorflow's object detection api. The task of object detection is a crucial part for self driving systems as it enables the system to detect vulnerable road users (pedestrians, byciclists), traffic signs and other cars in the scene. These objects serve as one of the inputs to the path planning of the system

Set up
This section should contain a brief description of the steps to follow to run the code for this repository.

For this project I used the provided workspace from Udacity. Nevertheless I will not use workspace in further projects as I had a lot of troubles (e.g. no space on device, even if I deleted all checkpoints and unneccesary files). This is also the reason why I could not analyse the data in more detail and I could not export the model. I will try a local setup for the next projects. Please take this into account when you are reviewing my project.

Dataset
Dataset analysis
This section should contain a quantitative and qualitative description of the dataset. It should include images, charts and other visualizations.

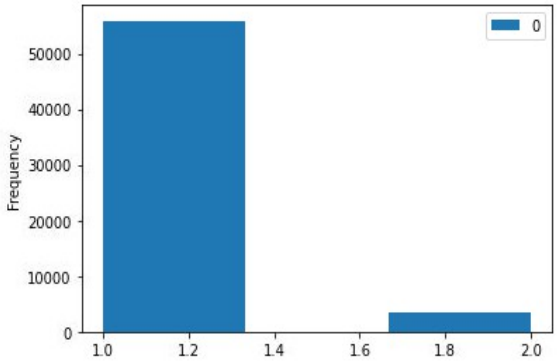




The training dataset shows different urban and rural scences (intersection, parking cars) during day and nighttime and changing weather conditions as can be seen in the figures above. Such a realistic split with differnet influencing factors is usefull to have a good performing model during all possible environmental conditions. The distribution of classes (e.g. cars, pedestrians, byciclists) seems to be rather imbalanced which most likely will effect the model's performance on the different classes.

Additional DA

```
0    55889
dtype: int64
0    3485
dtype: int64
```

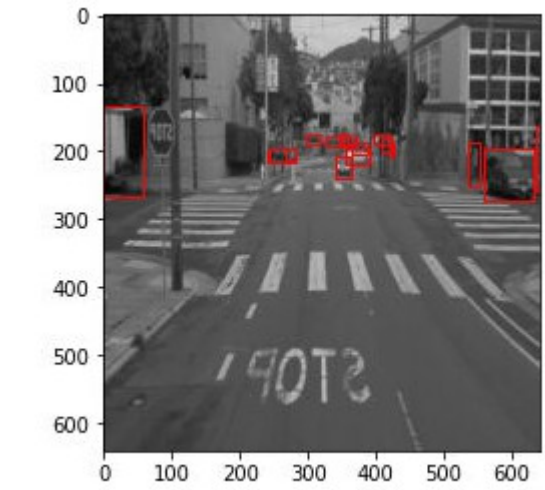


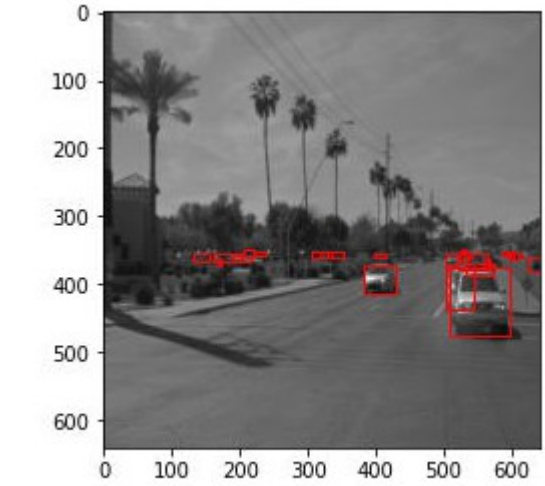
In the additional data analysis it can be seen that the car class is by far the most occuring class, followed by the pedestrian class. In this case 5000 images from the training split where inspected. Here no byciclist were present. That indicates a highly imbalanced dataset.

Data Augmentation

The notebook you provided was broken, so I had to guess what you want me to do here. I added random image cropping and random grayscaling.

```
  data_augmentation_options {
    random_crop_image {
      min_object_covered: 0.0
      min_aspect_ratio: 0.75
      max_aspect_ratio: 3.0
      min_area: 0.75
      max_area: 1.0
      overlap_thresh: 0.0
    }
  }
  data_augmentation_options {
    random_rgb_to_gray {
      probability: 0.95
    }
  }
```

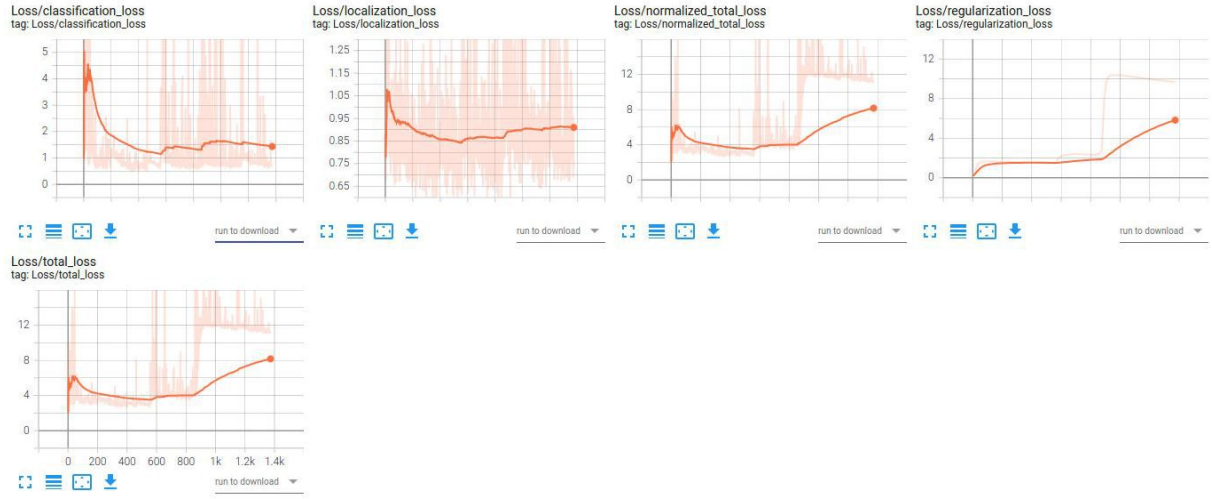




Training

Reference experiment

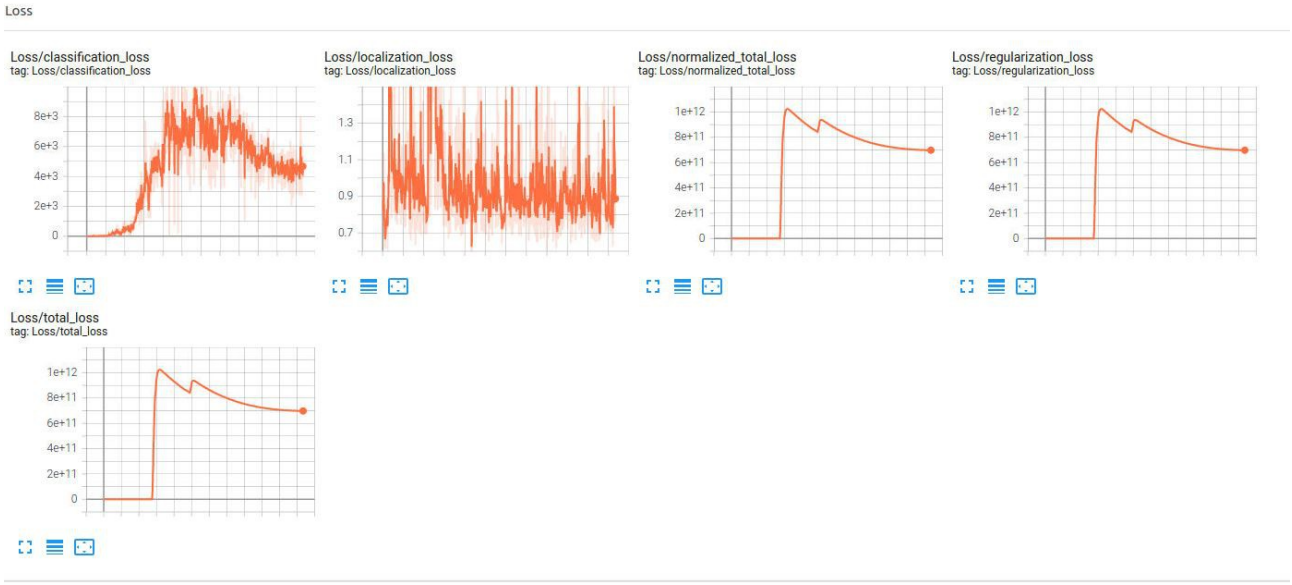
This section should detail the results of the reference experiment. It should includes training metrics and a detailed explanation of the algorithm's performances.



Using the config file provided and training the model for for 1400 epochs, it can be seen that classification and localization loss are still very high. Moreover the total loss seems to be raising. Most likely the model should hav been trained for more epochs. This was not possible due to space limitations.

Looking at a side by side evaluation of the model, it can be seen that none of the cars is detected and therefore all metrics are 0.

For the second config the learning_rate_base was set from 0.04 to 0.09 in order so speed up training. The learning_rate_base decayed and the model stoped trainign at an very early stage. This can be seen in the figure above, the trainign loss is still very high.



Improve on the reference

This section should highlight the different strategies you adopted to improve your model. It should contain relevant figures and details of your findings.

Because of the mentioned workspace issues, I could not evaluate more experiments. The main finding was, that increasing the learning_rate_base to 0.09 is defenetly to high as training stops to early. I was planning to change the optimizer from momentum optimizer to adam optimizer.