1 Evaluation

In this chapter, the workflow explained in last chapter is evaluated and results are presented.

Before showing the evaluation, it is necessary to define training and testing conditions that can be easily used by others to verify the results.

Each evaluation has a setup and its corresponding result.

We have three datasets that can be used for training and evaluation.

1. Dataset 1 - Contains 100,000 raw data. It is collected in no traffic environment, doing straight driving without any sudden turning. The data is using San Francisco map and driven during afternoon. This dataset has only data representing centre camera pointed ahead, parallel to the ground and right camera pointed to the ground at an angle 20°. The control commands include acceleration, throttle, braking, and steering angle values.

2. Dataset 2 - Also contains 100,000 raw data. It is, however, collect with traffic where the cars stop at signal intersections for a longer time than dataset 3. This dataset is also collect in San Francisco map and during afternoon. It contains a centre camera, right camera like dataset 1, left camera similar to right camera by pointing at an angle 20° to the ground, depth camera sensors placed at centre, left and right just like RGB cameras. The control commands are same as dataset 1.

3. Dataset 3 - Contains 270,000 raw data. It is collected while driving around San Francisco. About 200,000 data is collected while driving in the afternoon. About 20,000 in different weather and light conditions. About 50,000 entries are collected in a different circular circuit map called CubeTown. In addition to RGB and depth cameras distributed just as dataset 2, a segmentation camera is kept next centre RGB camera facing forward, and a radar sensor just in front of the car near the hood also facing forward.

1.1 Evaluation setup

While evaluating, a testing parameter *episode* is used. Each episode lasts 30 seconds. A timer is started for 30 seconds and the model is tested for collisions. If a collision happens, the time at which collision happened is noted.

1.1.1 Setup 1 - Determine the best lighting conditions to test the model

All three datasets are used. The test is conducted in San Franciso map without traffic option switched ON. By varying the light conditions to morning, afternoon and evening, we observe how light influences the prediction of output. Only steering angle is predicted and a steady velocity of 3 meter per second is used. An episode length of 30s is used. When a collision is observed, the time of collision and the count are noted down.

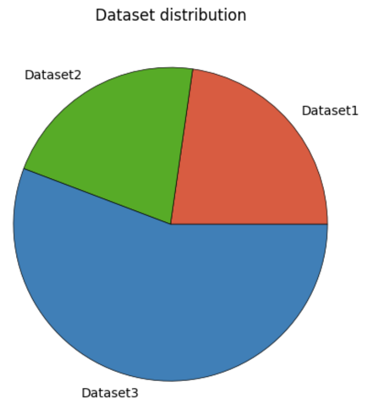

Figure 1.1: Datasets distribution

time(in 24 hrs standard)	Morning	Afternoon	Evening
	7:35	15:30	18:30

Table 1.1: Time of the day

It is seen from figure 1.2 that afternoon time provides the best light conditions for all the three datasets. Dataset 1 and 3 perform equally across the three lighting conditions.

If the percentage of number of collisions, as shown in figure 1.3, is calculated, dataset 3 performs the best among the datasets for morning and afternoon part of the day.

1.1.2 Setup 2 - How the datasets perform during afternoon if traffic is enabled?

All three datasets are again used. The time is fixed at 15:30. The traffic is toggled ON.

From figure 1.4, we can observe that all three datasets do well even in traffic. However, it is surprising to see dataset 1 which had no traffic while the dataset was collected, performs remarkably well when driven in traffic.

1.1.3 LSTM vs Non-LSTM

In the next stage of evaluation, acceleration in addition to steering is predicted. Acceleration in LGSVL contains positive values for forward throttle and negative values for braking.

Acceleration is relative to previous frame. Hence it is necessary to include past frames information while predicted it. LSTM is used for this purpose. When non-LSTM model is used to predict acceleration, it only predicts for the current frame which often results in vehicle being stationary.

For our setup, we choose a $timestep = 15$. That means acceleration of current time frame is predicted using previous 15 time frames.

1.1.4 Regression models - Acceleration and Steering

Determining the optimal LSTM output units

In the table 1.2, we can see that for 100 LSTM output units, though the trainable parameters(parameters that can be trained during backpropagation) is quite compared to other units, the

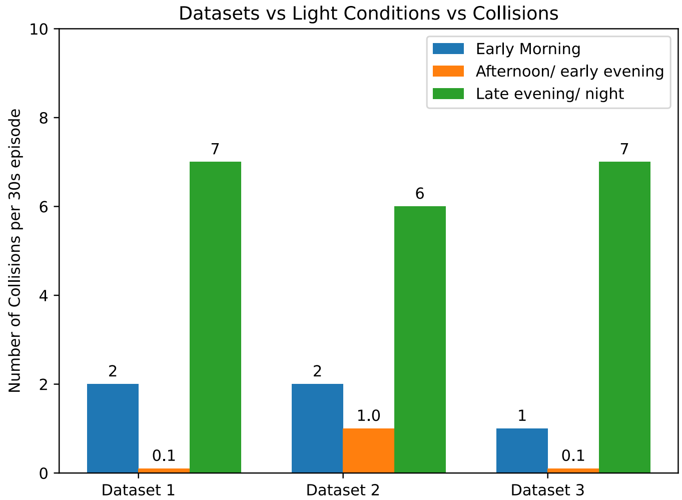

Figure 1.2: Datasets vs Light Conditions vs Number of Collisions

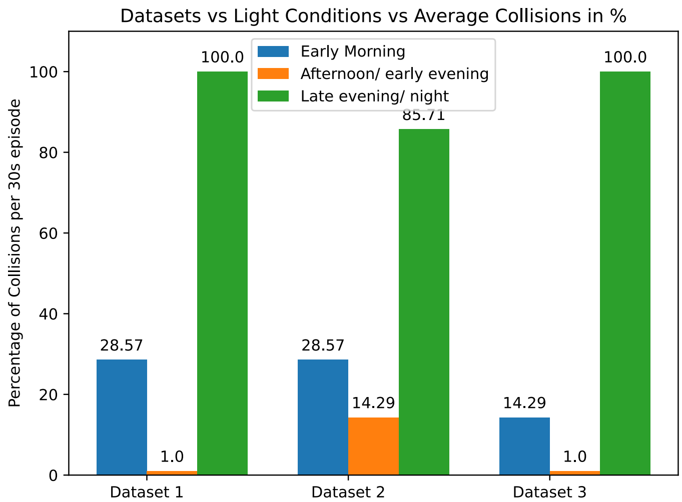

Figure 1.3: Datasets vs Afternoon vs Traffic vs Average number of Collisions

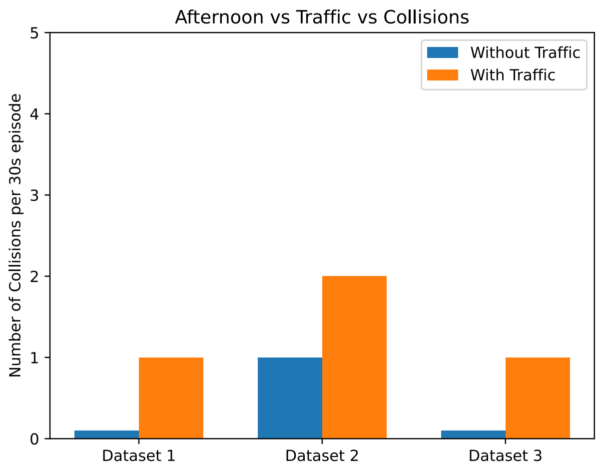

Figure 1.4: Datasets vs Afternoon vs Traffic vs Number of Collisions

LSTM Output Units	Trainable Parameters	Processing time needed
20	20000	1hr 44m
60	61000	1hr 42m
100	434000	1hr 40m

Table 1.2: LSTM Output Units vs Training time

training time is the least. Moreover on evaluation, it is observed that acceleration prediction is relatively good compared to other units' models. Hence, a LSTM unit of 100 is chosen.

1.1.5 Classification and Regression models

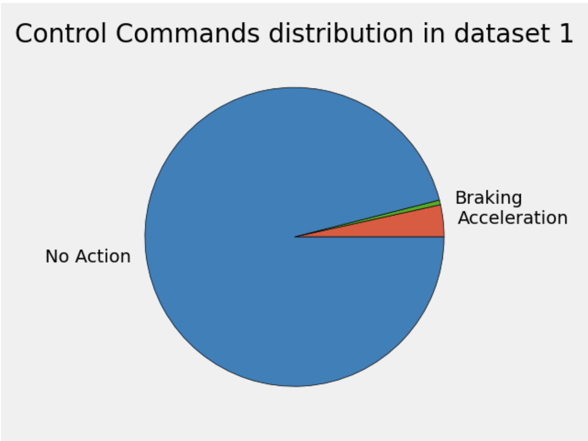

Figure 1.5: Dataset 1 Control commands distribution

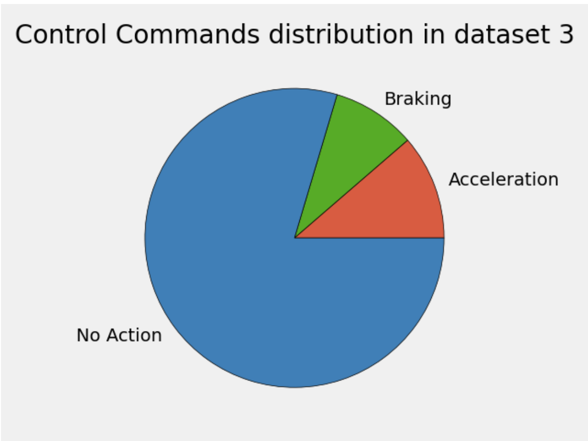

Figure 1.6: Dataset 3 Control commands distribution

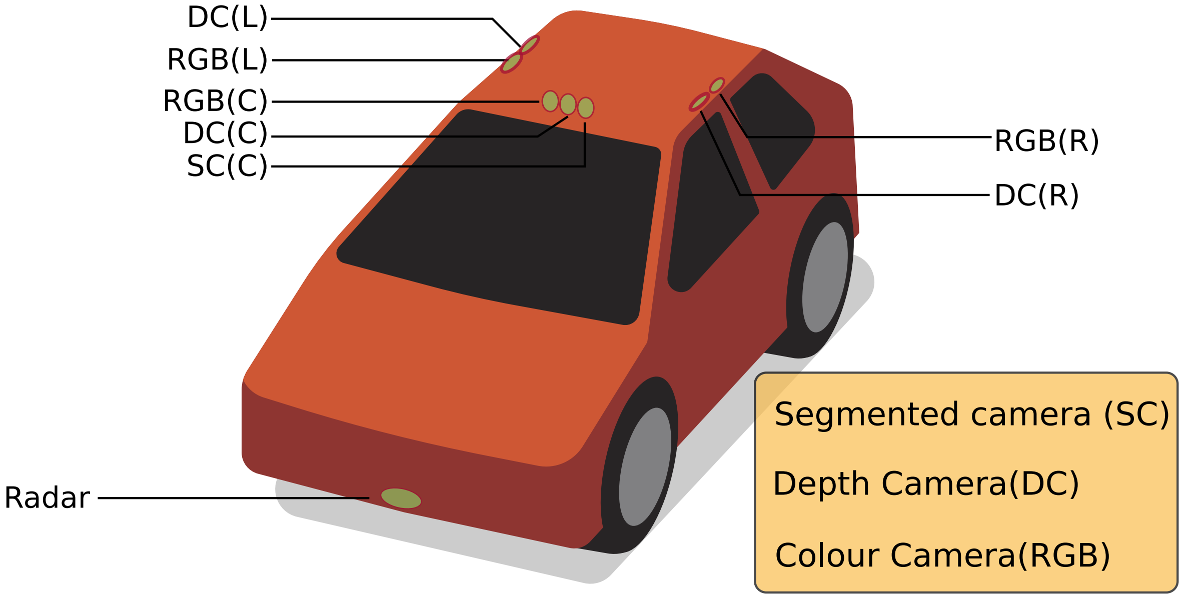

Figure 1.7: Sensor Constellation

Binary Crossentropy

Categorical Crossentropy - a basic model

Categorical Crossentropy - splitting at the dense layers

Categorical Crossentropy - splitting at the LSTM layers

Categorical Crossentropy - splitting at the flatten layers

Categorical Crossentropy - Using two different NN for acceleration and Steering

1.1.6 Velocity

List of Figures

References