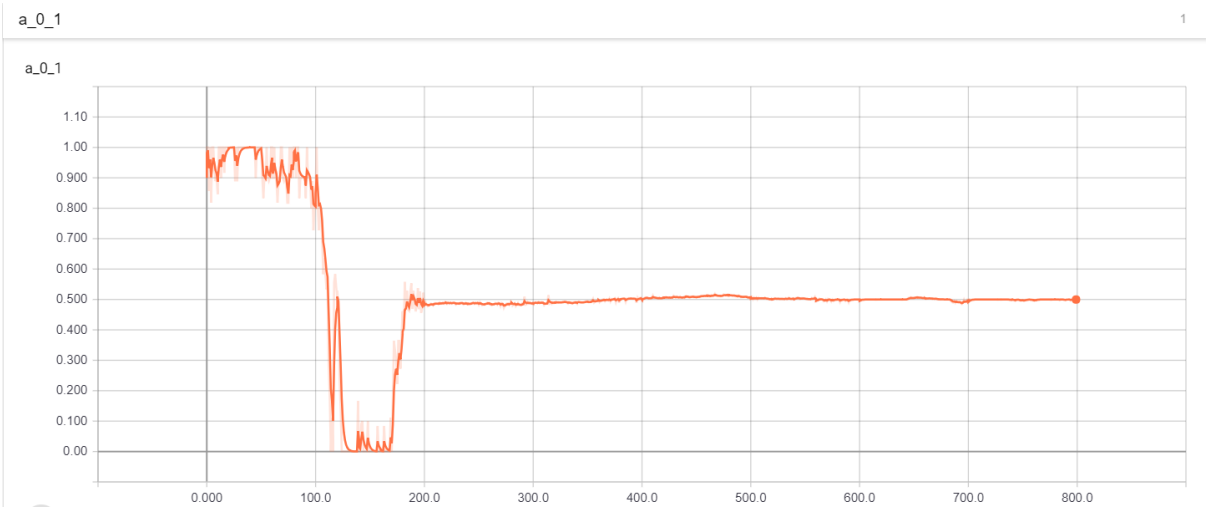
REPORT

Task 2.1 – Cartpole

We implemented the cartpole task using DQN, trained this with different hyperparameter configurations and obtained the required mean test score of 200+ over 15 episodes. Below are the details:

LR	Episodes	Epsilon	Tau	Test Reward
5e-4	800	0.05	0.01	344
5e-4	800	0.05	0.01	709



Here we have also done a combined plot of the running and the average value of the rewards obtained during training. The value of mean reward after every 20 episodes is plotted in the same graph (hence the spikes in the values at every 20th step). Final test reward is 344 for this training example.

Task 2.2 – Car Racing

For Car Racing we have implemented a 3-layer CNN with 2 stride, 0 padding for both Q-network and the target network with LR as 1e-4 and Tau value as 0.01. Initially the reward values were around 400-500, so to improve this we made use of all the suggested tricks – set higher probabilities for straight and acceleration, skipped frames and made use of adaptive max_timesteps.

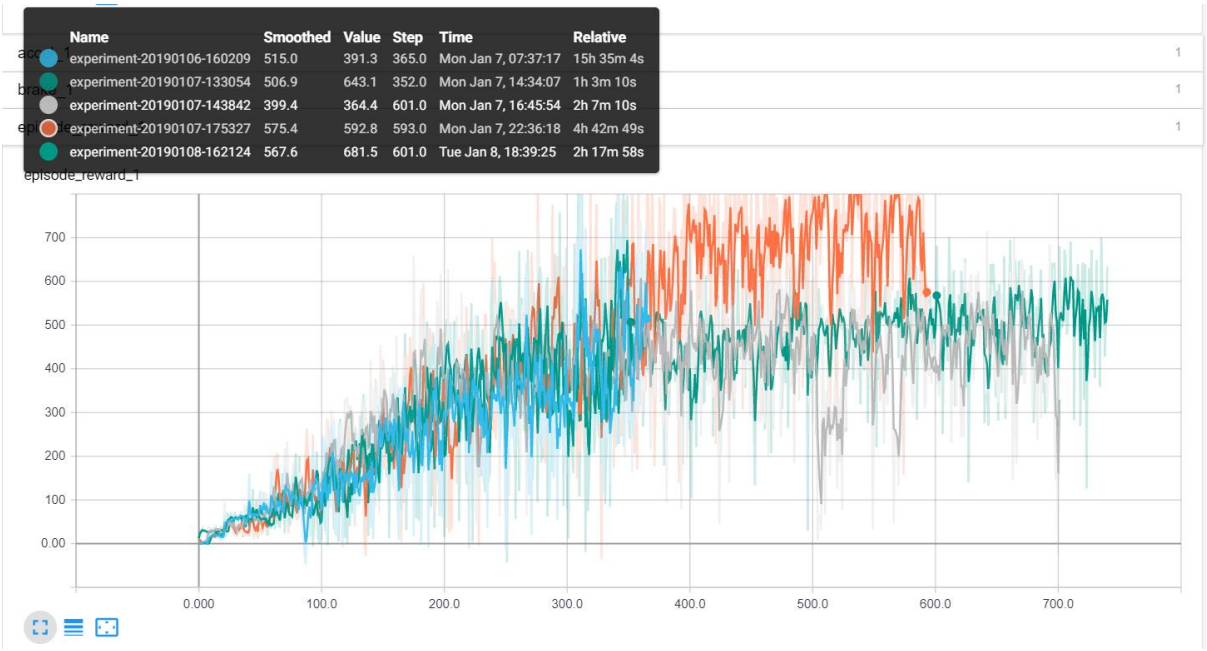
- Exploration: Tried the following combinations of probabilities for the actions

S	L	R	A	B	Reward
0.30	0.15	0.15	0.30	0.10	399
0.35	0.10	0.10	0.35	0.10	785
0.40	0.075	0.075	0.40	0.05	600
0.35	0.10	0.10	0.40	0.05	680

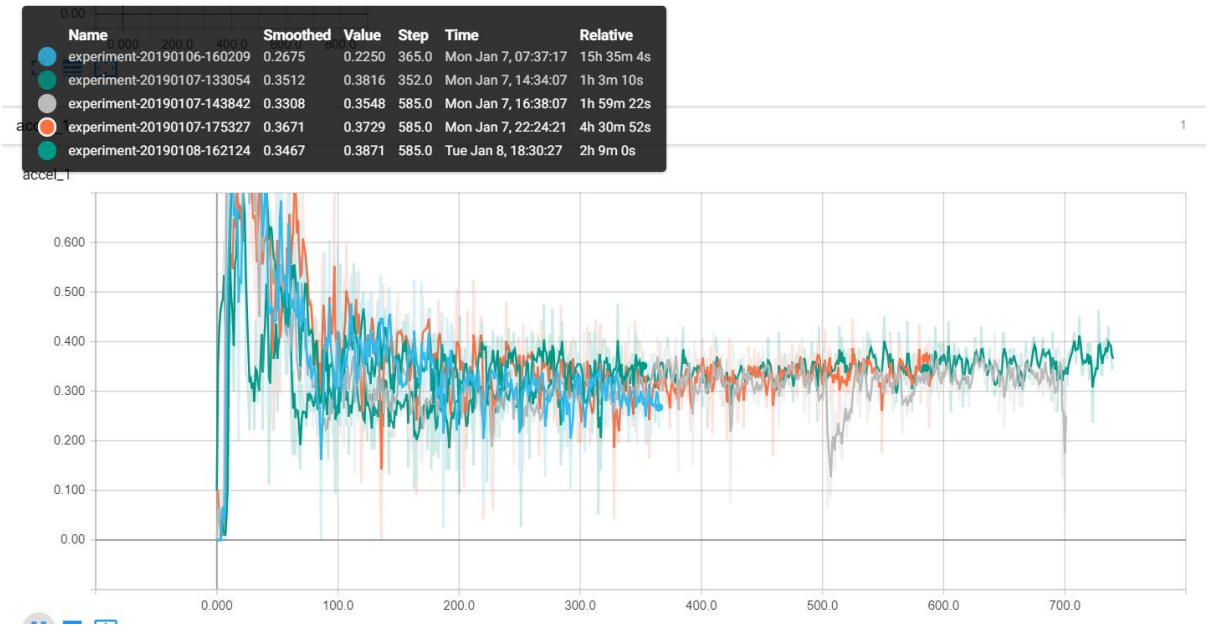
- Frame Skipping: Frame skipping was set at 5.
- Shorter Episodes: For a total episode run of around 700-900, we varied the value of max_timesteps from 500 to 750. We set the value of max_timesteps initially to 50 episodes and increased it by 50 episodes after every 20 episodes, [50, 100, 150 …]

The combination that worked best for us (mean reward of around 800) is:
Exploration probabilities: [S 0.35, L 0.10, R 0.10, A 0.40, B 0.05]
Frame Skip: 5
max_timesteps: 50 through 500, incremented by 50 after every 20 steps, for a total 700 steps.

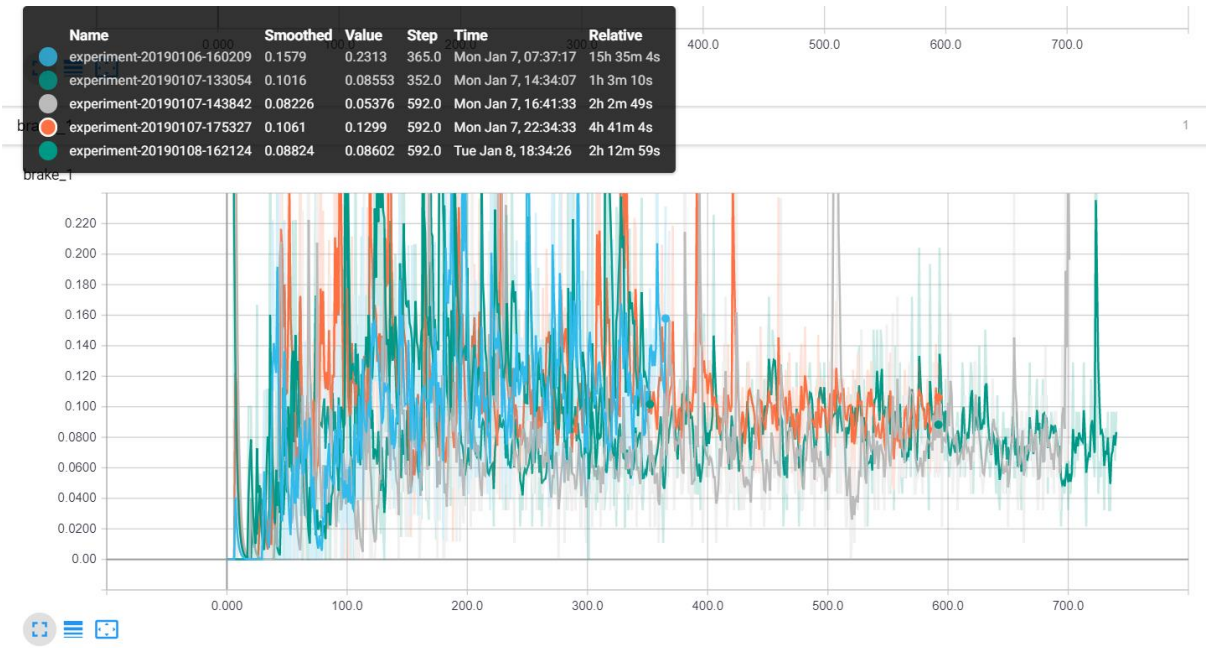
Episode_Rewards



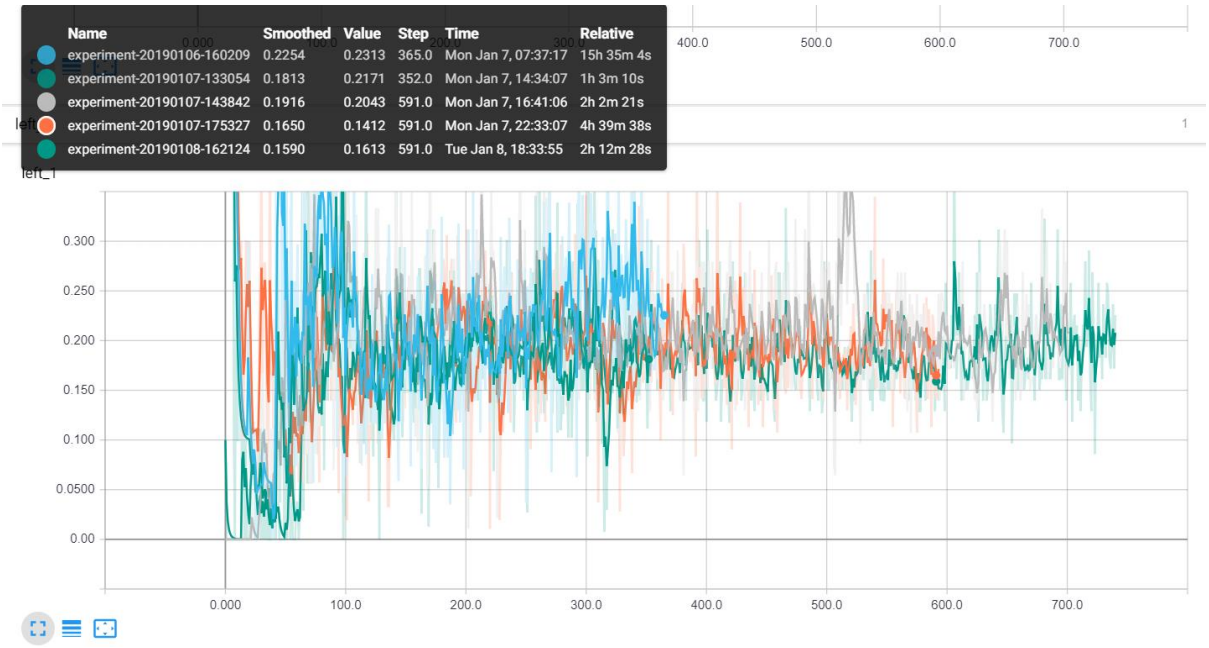
Accelerate



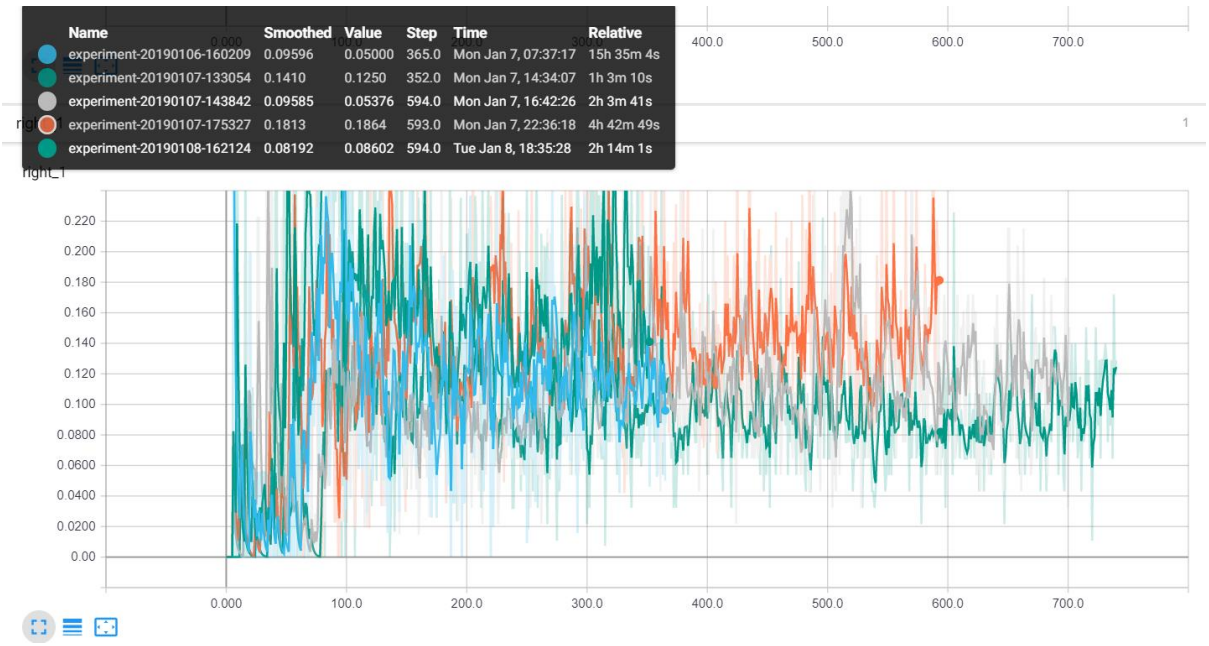
Brake


Name Smoothed Value Step Time Relative
experiment-20190106-160209 0.1579 0.2313 365.0 Mon Jan 7, 07:37:17 15h 35m 4s
experiment-20190107-133054 0.1016 0.08553 352.0 Mon Jan 7, 14:34:07 1h 3m 10s
experiment-20190107-143842 0.08226 0.05376 592.0 Mon Jan 7, 16:41:33 2h 2m 49s
experiment-20190107-175327 0.1061 0.1299 592.0 Mon Jan 7, 22:34:33 4h 41m 4s
experiment-20190108-162124 0.08824 0.08602 592.0 Tue Jan 8, 18:34:26 2h 12m 59s
brake_1
0.220
0.200
0.180
0.160
0.140
0.120
0.100
0.0800
0.0600
0.0400
0.0200
0.00
0.000
100.0
200.0
300.0
400.0
500.0
600.0
700.0

Left


Name Smoothed Value Step Time Relative
experiment-20190106-160209 0.2254 0.2313 365.0 Mon Jan 7, 07:37:17 15h 35m 4s
experiment-20190107-133054 0.1813 0.2171 352.0 Mon Jan 7, 14:34:07 1h 3m 10s
experiment-20190107-143842 0.1916 0.2043 591.0 Mon Jan 7, 16:41:06 2h 2m 21s
experiment-20190107-175327 0.1650 0.1412 591.0 Mon Jan 7, 22:33:07 4h 39m 38s
experiment-20190108-162124 0.1590 0.1613 591.0 Tue Jan 8, 18:33:55 2h 12m 28s
left_1
0.300
0.250
0.200
0.150
0.100
0.0500
0.00
0.000
100.0
200.0
300.0
400.0
500.0
600.0
700.0

Right



Straight