Park It Right!

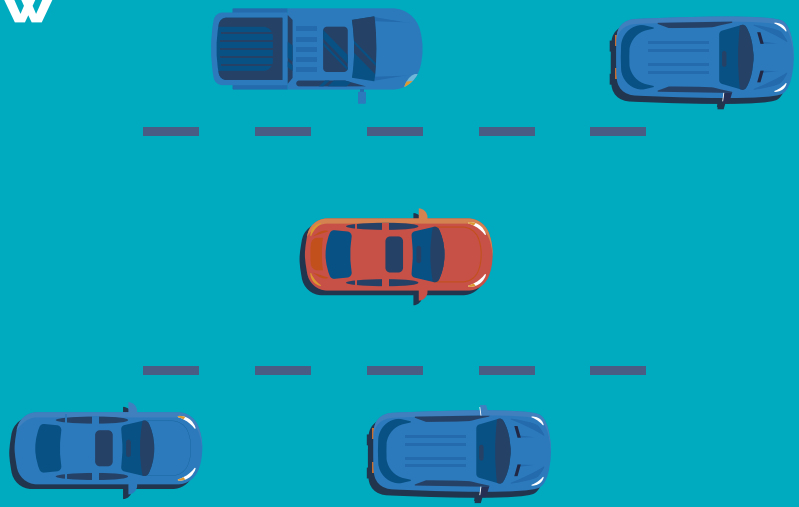
Keerthana Nandanavanam
nandanav@usc.edu
Krishna Manoj Maddipatla
km69564@usc.edu
Nidhi Chaudhary
nidhicha@usc.edu
Sumanth Mothkuri
mothkuri@usc.edu

Game Overview

Car Parking Simulation

Level 1 - Static obstacles & 1 floor

Level 2 - Moving obstacles & 2 floors



Game Modifications

SCOREBOARD

Parking Score,
Obstacle Hit Score,
Wall Hit Score,
Cumulative Reward

NAVIGATION

Keyboard navigation instead of Touchscreen

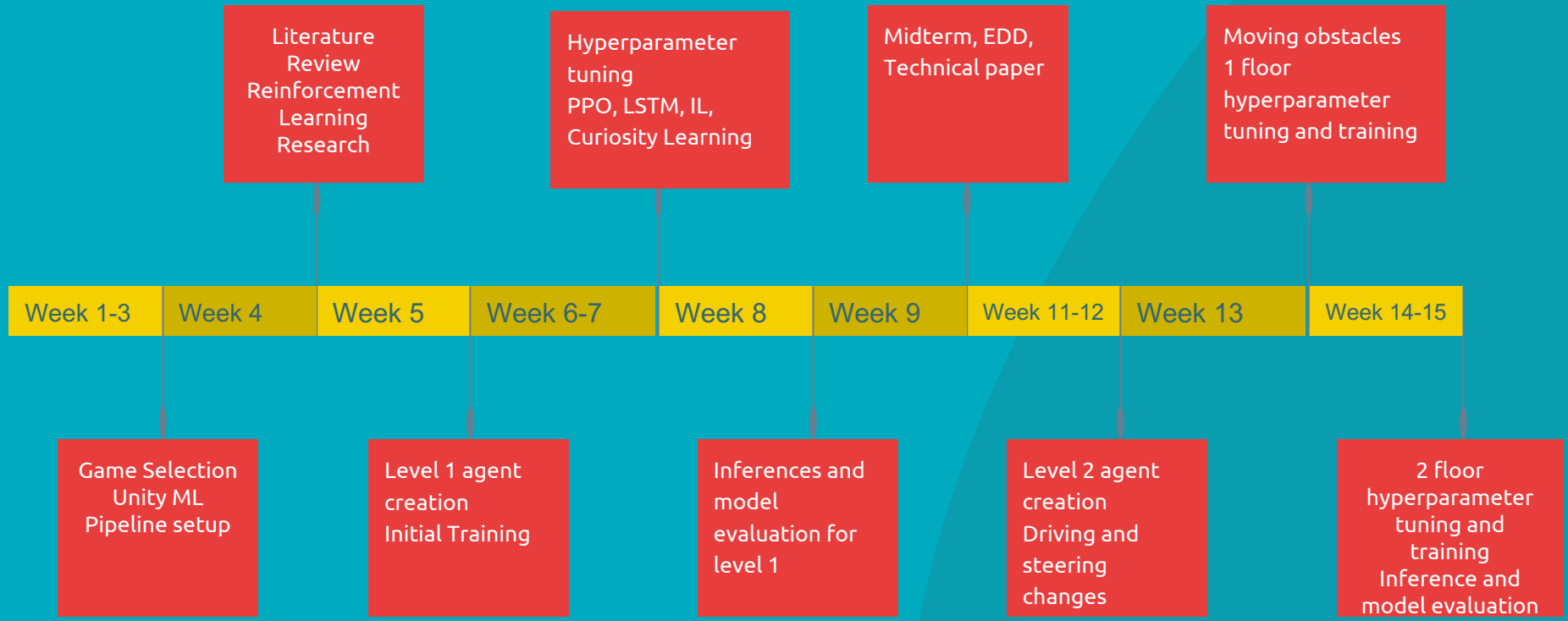
OBSTACLES

Converted boundaries and walls to collision objects

PARKING SPOTS

Randomly assign a parking spot from the set of available spots

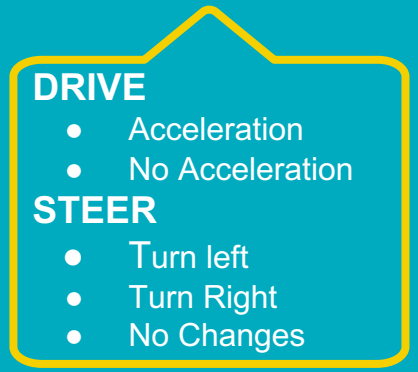
Project Timeline



Model Architecture

Observations

Action Set



Game Environment
Car Parking Arena

Environment State Changes

Read Observations
- Ray Perceptron 3D
- Collect Observations

GAIL Discriminator

Discriminator compares agent observations/actions with demonstration and rewards agent for being close

Expert Demonstrations

Perform action on environment from action set

Agent

Assign rewards based on action

Reward System

S.No	Condition	Level 1 Reward [PPO]	Level 1 Reward [PPO + GAIL]	Level 2 Reward [PPO + GAIL]
1.	**Hit the wall [Episode Ends]**	-0.5	-0.5	-0.5
2.	**Hit an obstacle [Episode Ends]**	-0.5	-0.5	-0.5
3.	**Car Parked [Episode Ends]**	+5	+5	+5
4.	**Within 2.5 units of distance to the goal location**	+0.00008	+0.00003	+0.00003
5.	**Best current distance to the goal location**	+0.00002	+0.00002	+0.00002
6.	**Moving towards the goal but not the best distance**	-0.00004	+0.00001	+0.00001
7.	**Moving away from the goal**	-0.00008	-0.00002	-0.00002
8.	**Within 2 units of distance to the wall**	-0.005	-0.005	-0.005
9.	**Within 2 units of distance to the obstacle**	N/A	-0.005	-0.005
10	**Move through portal towards target**	N/A	N/A	+0.5
11	**Move through portal away from target**	N/A	N/A	-0.1

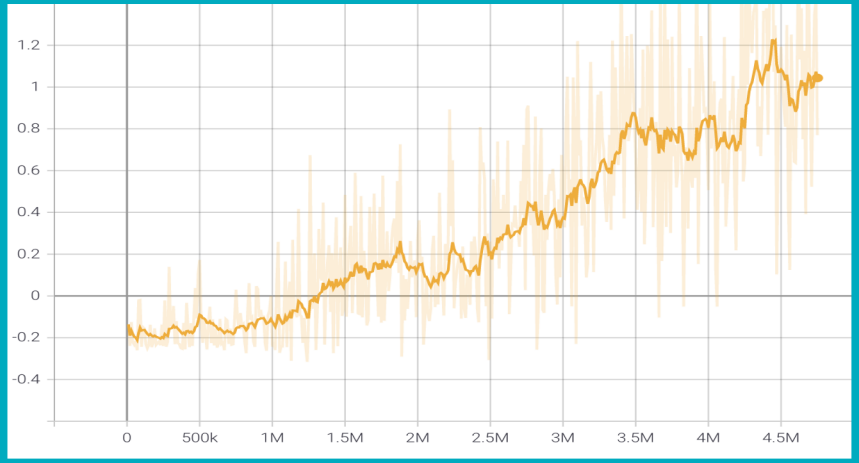
Hyperparameters - Level 1

- Performed training on 9 different hyperparameters*
- Low learning rate of 1e-05, high batch and buffer size for stability

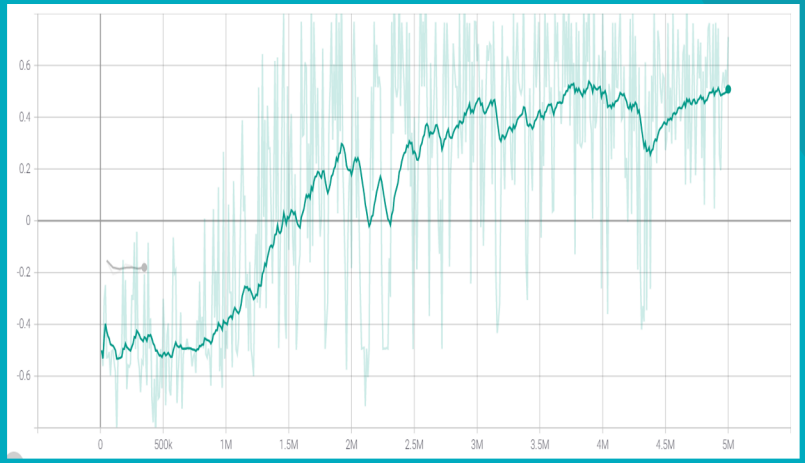
PPO + LSTM
batch size = 512 buffer size = 10240 beta = 0.001 epsilon = 0.3 hidden units= 64 Number of layers = 2 Normalize = True lambd=0.92

PPO + LSTM + GAIL
batch size = 256 buffer size = 20480 beta = 0.03 epsilon = 0.1 hidden units= 64 Number of layers = 2 Normalize = False lambd=0.92 **Gail strength = 0.7**

Cumulative Rewards - Level 1



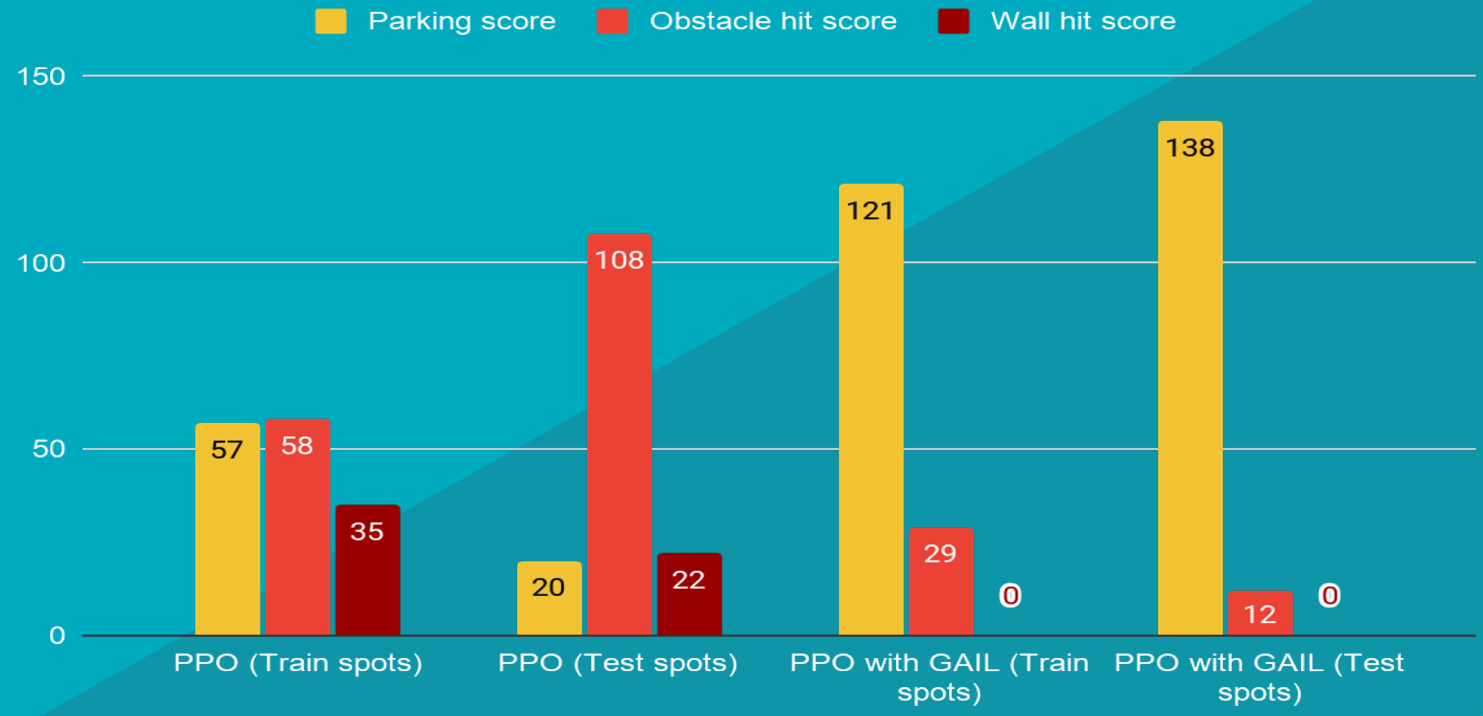
PPO



PPO with GAIL

- Cumulative rewards keep on increasing with the number of steps for both PPO and PPO with GAIL (5M steps).
- Entropy decreases for both as well!

Inference Statistics - Level 1



GAIL with PPO performs much better!

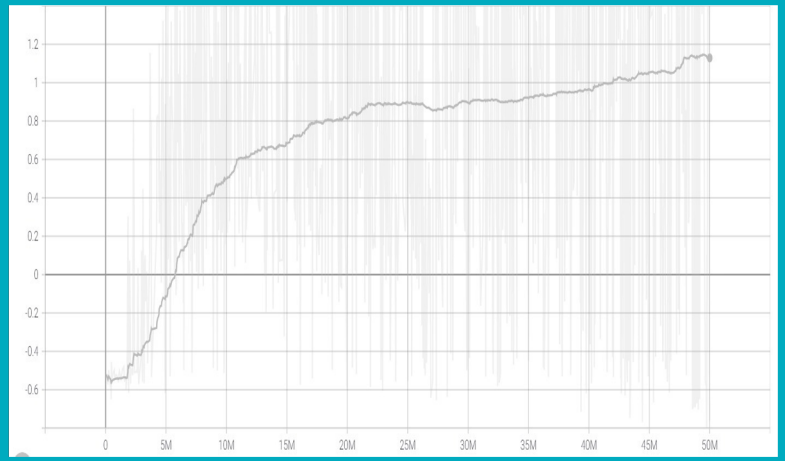
Hyperparameters - Level 2

- Performed on 2 different hyperparameters*
- Smaller batch size but higher buffer size for stability
- Trained for 50M steps (6 days)!
- 64 neurons in hidden layers work best!

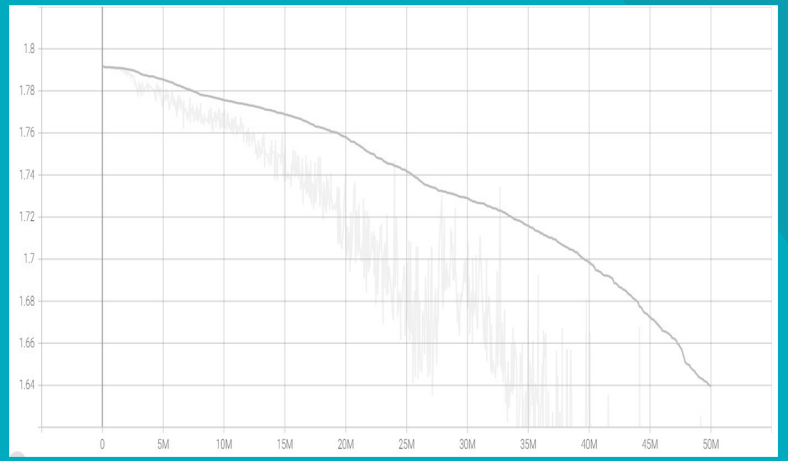
PPO + LSTM + GAIL

- batch size = 256
- buffer size = 20480
- beta = 0.03
- epsilon = 0.1
- hidden units= 64
- Number of layers = 2
- Normalize = False
- lambd=0.92
- **Gail strength = 0.7**

Rewards and Entropy - Level 2



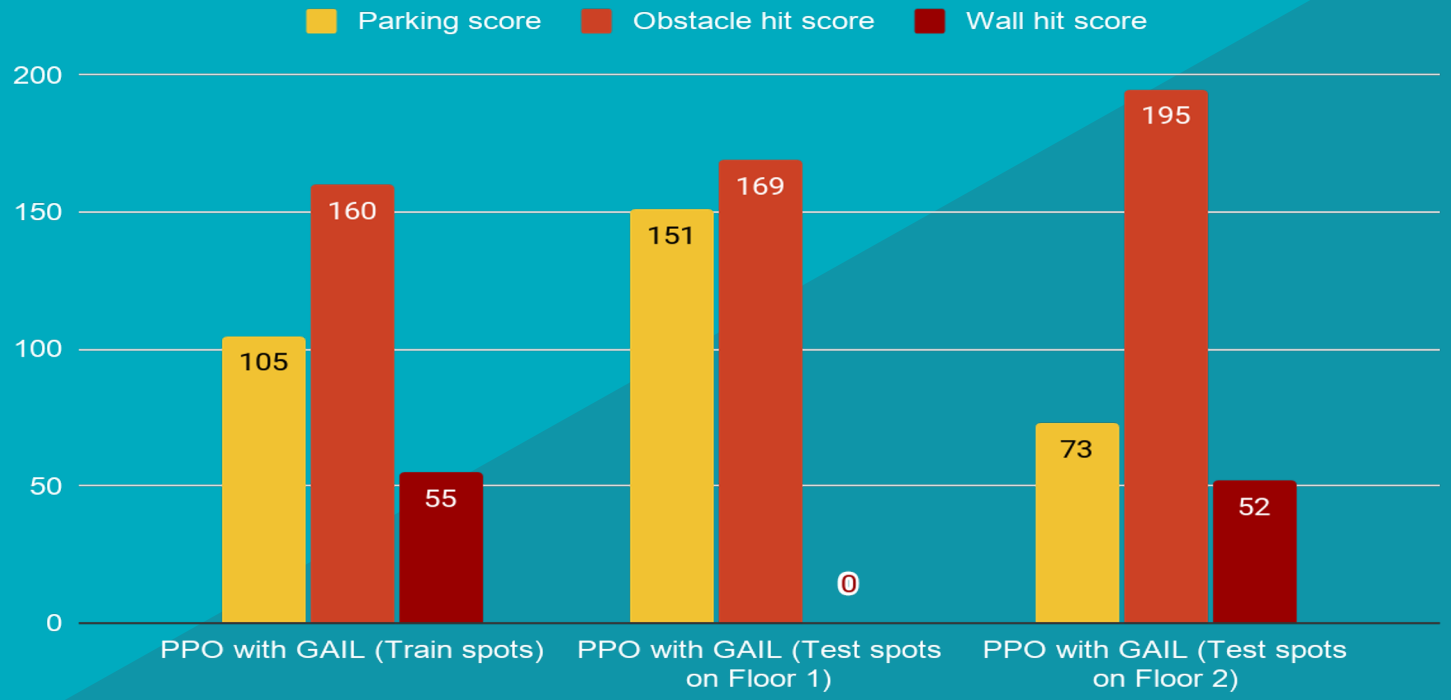
Cumulative Rewards (PPO with GAIL)



Entropy (PPO with GAIL)

- Cumulative rewards keep on increasing with the number of steps
- Entropy decreases continuously

Inference Statistics - Level 2



Demo

Work Division

KEERTHANA

- Algorithm Research
- Project setup and training
- PPO with GAIL
- SAC
- Hyperparameter Tuning
- Training with Ray Perceptron 3D

KRISHNA

- Game modifications
- PPO, PPO with GAIL
- Hyperparameter Research & Design
- Positive reward system design
- Ray Perceptron 3D to read observation space

NIDHI

- Algorithm Research
- PPO, RDN, PPO with GAIL
- Hyperparameter Research & Design
- Negative reward system design
- Observations collection and changes

SUMANTH

- Curiosity Learning
- Hyperparameter Tuning
- Object detection
- Architecture Design
- Observations collection and changes


THANK YOU

APPENDIX

Hyperparameters-Level 1

learning rate = 1e-05
Lambd = 0.92
No normalization

SL.No	Parameters	Steps	Result
1	PPO, batch size = 256 , buffer size = 10240, beta = 0.01, epsilon = 0.3, layers = 2, hidden units = 128, time horizon = 256	5M	
2	PPO, batch size = 32, buffer size = 2048, beta = 0.01, epsilon = 0.3, layers = 2, hidden units = 64, time horizon = 128	1M	
3	PPO, batch size = 32, buffer size = 3028, beta = 0.03, epsilon = 0.1, layers = 2, hidden units = 64, time horizon = 256	1M	
4	PPO, batch size = 256, buffer size = 20480, beta = 0.03, epsilon = 0.1, layers = 2, hidden units = 64, time horizon = 256	1M	
5	PPO, batch size = 256, buffer size = 20480, beta = 0.03, epsilon = 0.1, layers = 3, hidden units = 128, time horizon = 256	5M	

Hyperparameters-Level 1

learning rate = 1e-05
Lambd = 0.92

SL.No	Parameters	Steps	Result
6	PPO with RND, gamma: 0.99, strength: 0.01, encoding_size: 64, learning_rate: 0.0001, batch size = 512, buffer size = 10240, beta = 0.001, epsilon = 0.3, normalize = True, layers = 2, hidden units = 64, time horizon = 128	4M	
7	PPO with Curiosity, gamma: 0.99, strength: 0.2, encoding_size: 128, learning_rate: 0.0001, batch size = 512, buffer size = 10240, beta = 0.001, epsilon = 0.3, normalize = True, layers = 2, hidden units = 64, time horizon = 128	1M	
8	**PPO, LSTM, batch size = 512, buffer size = 10240, beta = 0.001, epsilon = 0.3, hidden units= 64 number of layers = 2, normalize = True**	5M	
9	**PPO with gail, LSTM, batch size = 256, buffer size = 20480, beta = 0.03, epsilon = 0.1, hidden units = 64, number of layers = 2, gail strength = 0.7, normalize = False**	5M	

Hyperparameters-Level 2

learning rate = 1e-05
Lambd = 0.92
No normalization

SL.No	Parameters	Steps	Result
1	PPO with GAIL, batch size = 64 , buffer size = 409600, beta = 0.03, epsilon = 0.3, layers = 2, hidden units = 64, time horizon = 256	8M	
2	PPO, batch size = 256, buffer size = 204800, beta = 0.03, epsilon = 0.1, layers = 2, hidden units = 64, time horizon = 256	50M