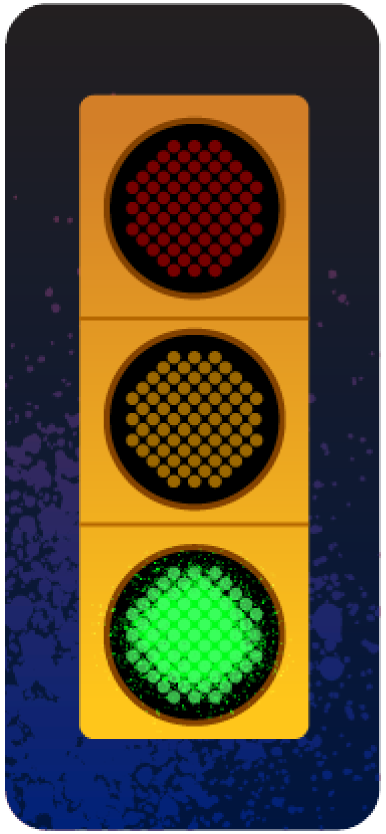
TRAFFIC SIGNAL OPTIMIZATION



Francisco Cavazos, Emmanuel Saenz, Yahir Duran

PROJECT INTRODUCTION

Problem

- Idling vehicles result in over 3 billion gallons of fuel burned each year in the United States alongside 30 million tons of CO2 emissions[1]. Our objective is to reduce traffic congestion by optimizing traffic signals.

Solution?

- By utilizing Reinforcement Learning, we hope in optimizing traffic flow in intersections, helping in reducing wasted fuel, CO2 emission, and **traffic delay for drivers**.

BRIEF LESSON REINFORCEMENT LEARNING

What is RL?

- Reinforcement Learning (RL) is a type of machine learning where there is an agent that perform actions based on an environment by receiving rewards or penalties based on those action.

Key Components:

Agent
The decision-maker

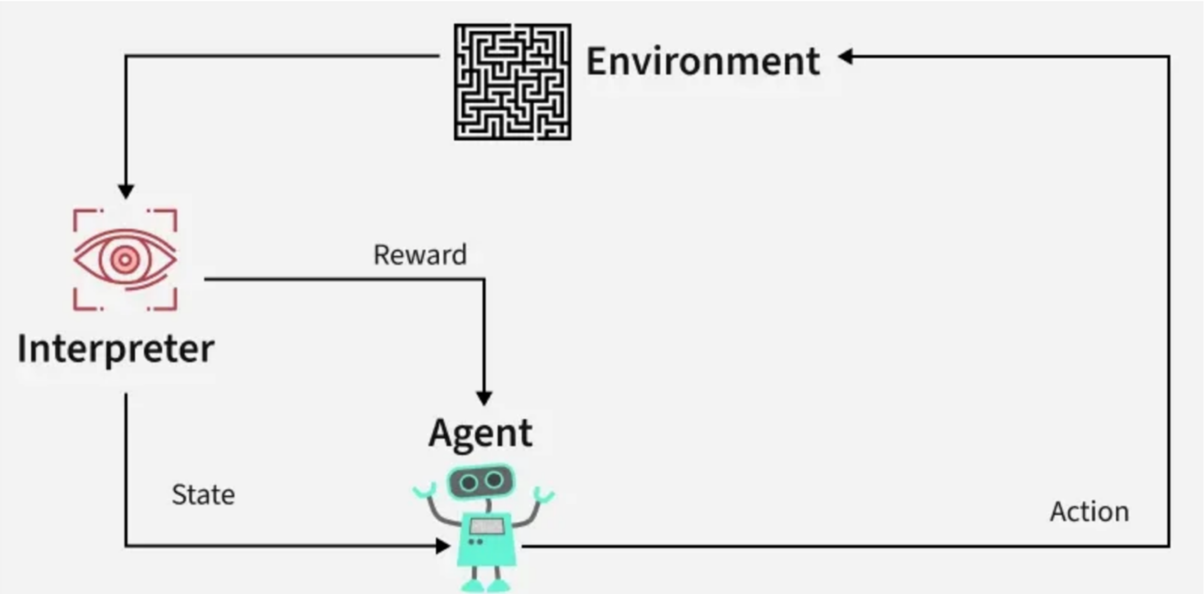
State
The current situation

Environment
The system the agent interacts with

Action
The possible choices the agent can make

Reward
Feedback signal that tells the agent how good its action was

RL APPLIED TO TRAFFIC SIGNALS

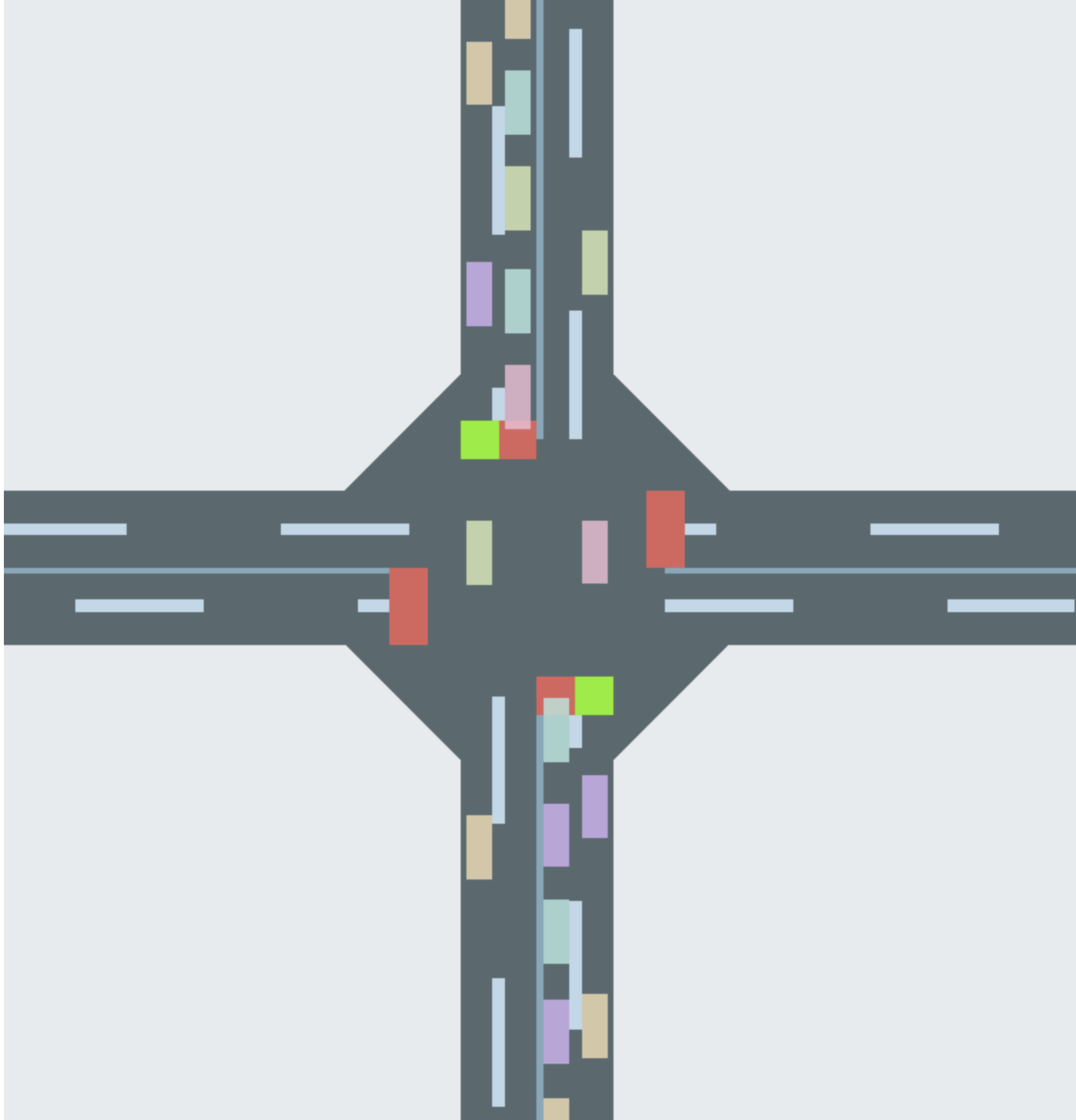


The RL agent learns by trial and error, and adjusts the signals to maximize the reward

Environment
The traffic intersections with cars arriving and waiting.
Agent
The system controlling the traffic light
Action (Prediction)
Changing the signal phases (Green or Red)
State (Features)
Current traffic conditions such as number of waiting cars, waiting times, etc.
Rewards
3 rewards utilized related to cars or time.

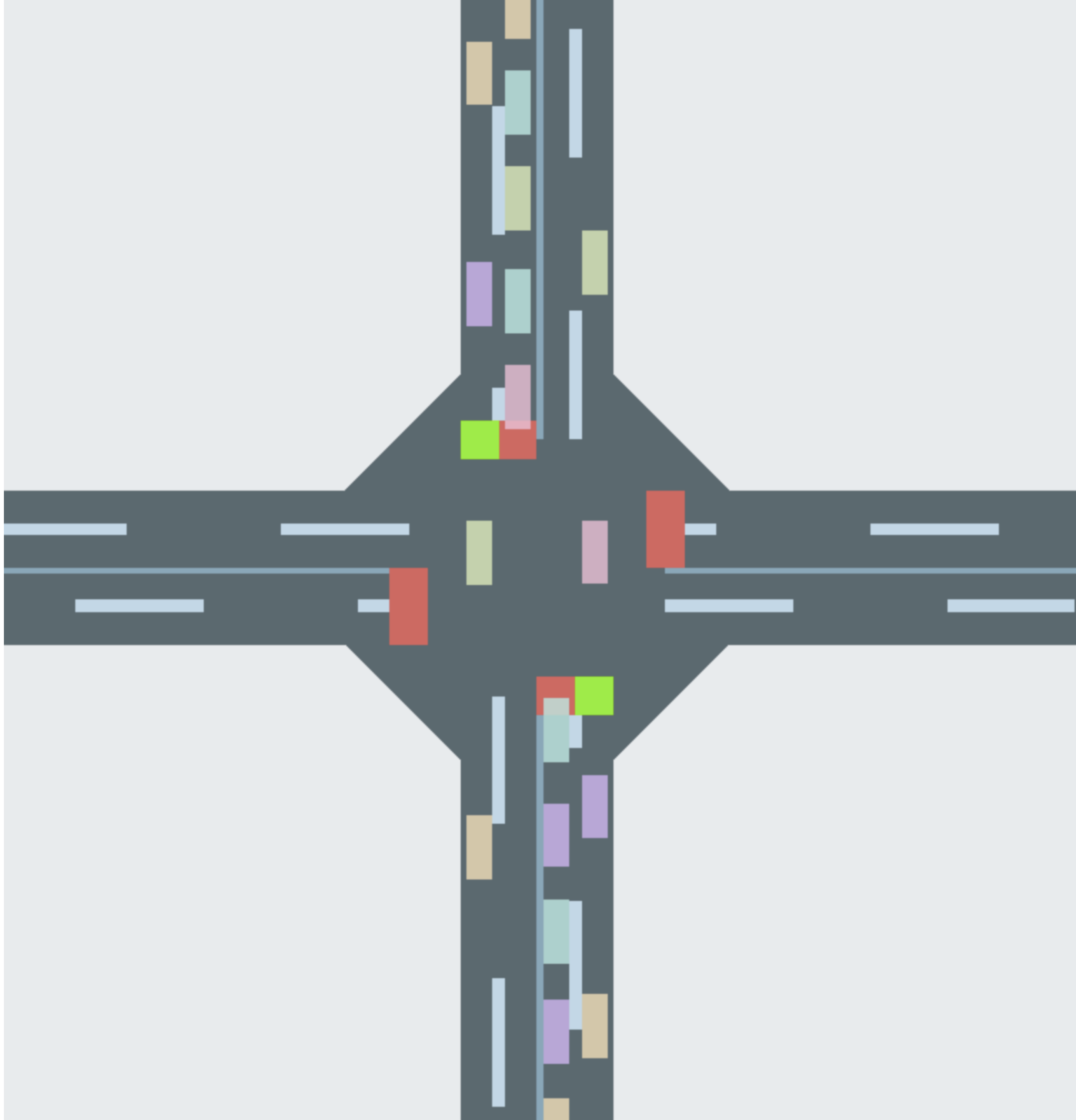
ENVIRONMENT

To tackle this problem, we created a 4-way intersection using a traffic engine[2] where the agent can modify the traffic signals.



ENVIRONMENT

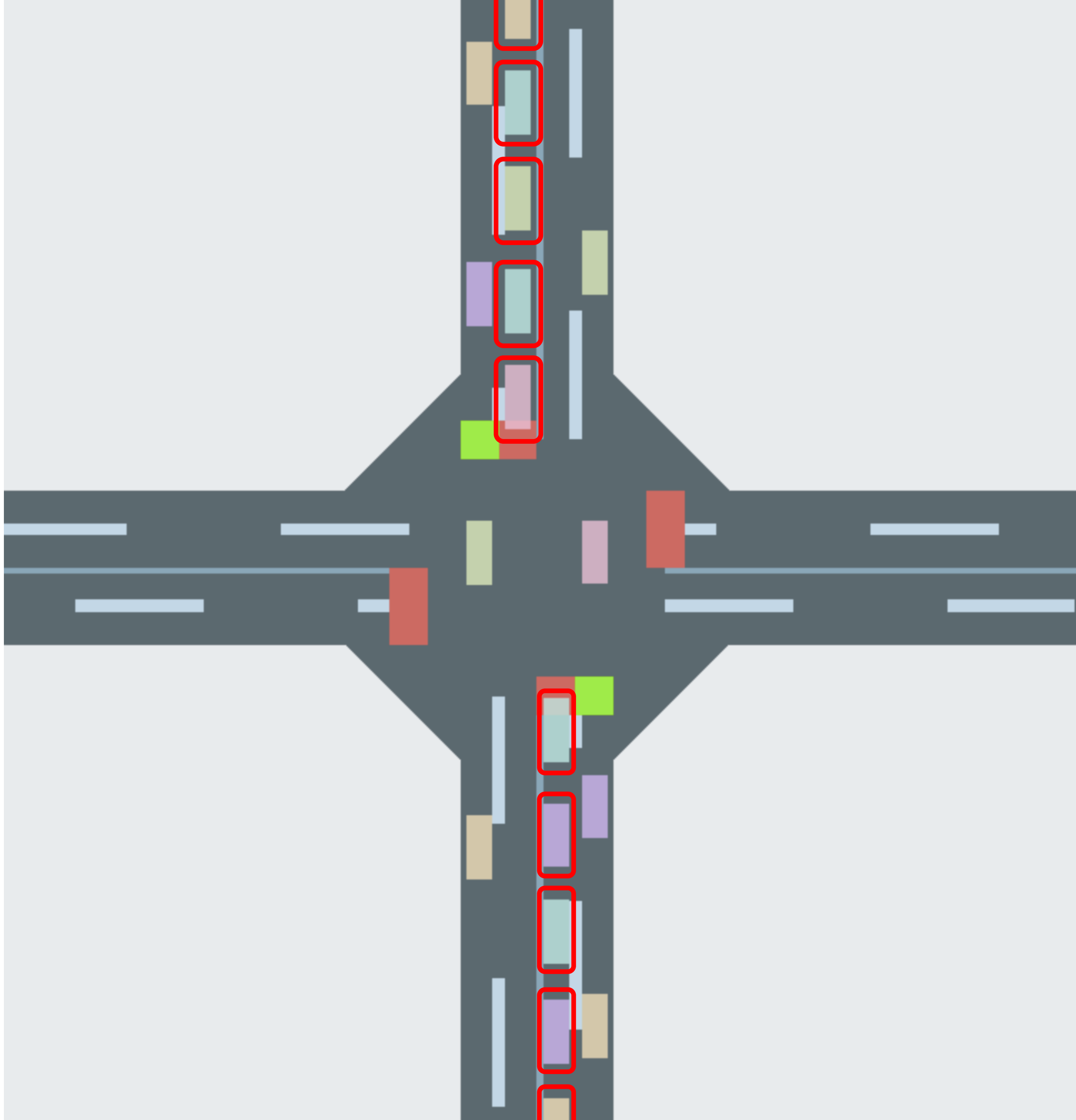
The agent has access to the following features:



ENVIRONMENT

The agent has access to the following features:

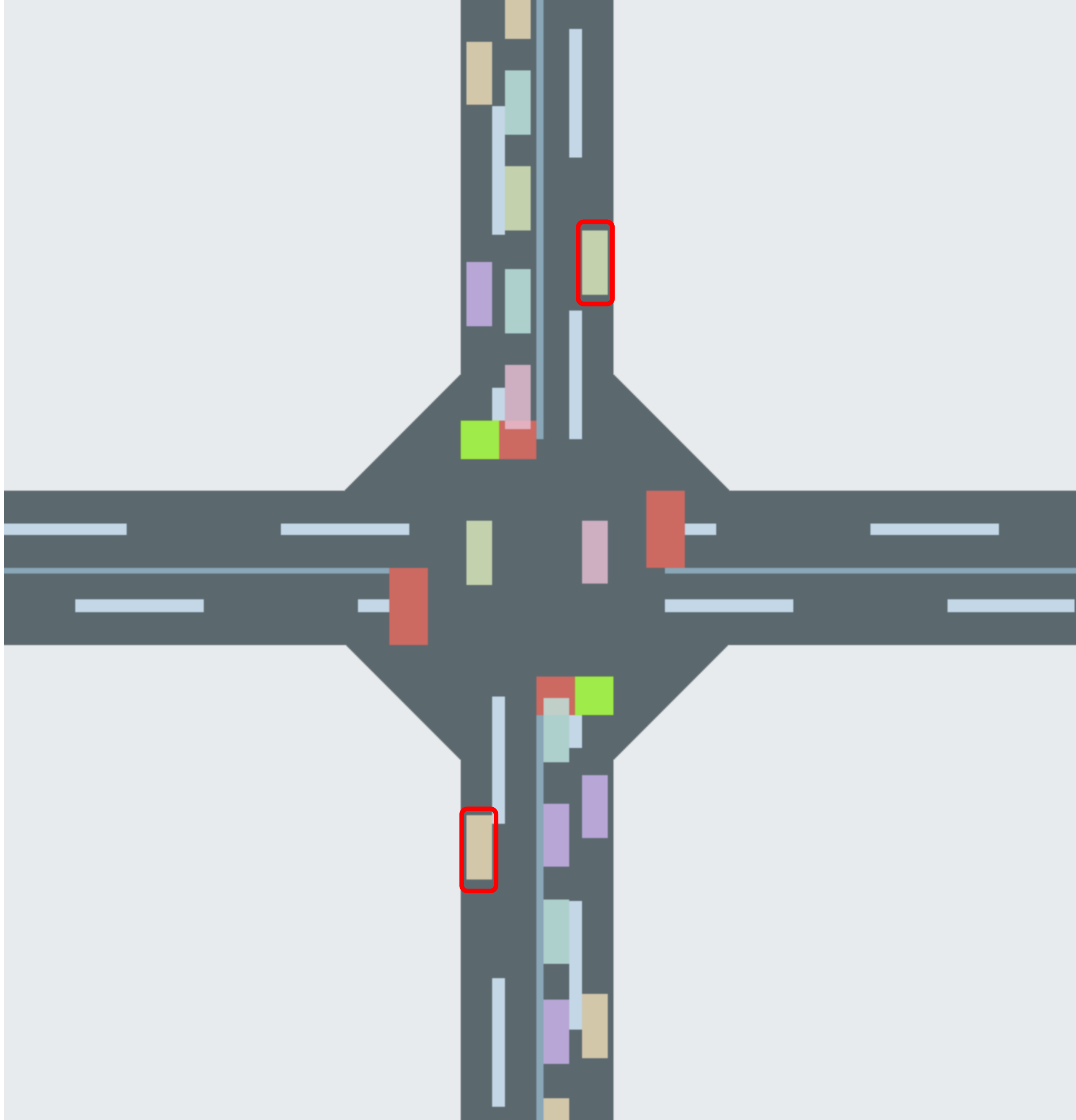
1. Number of waiting cars per lane



ENVIRONMENT

The agent has access to the following features:

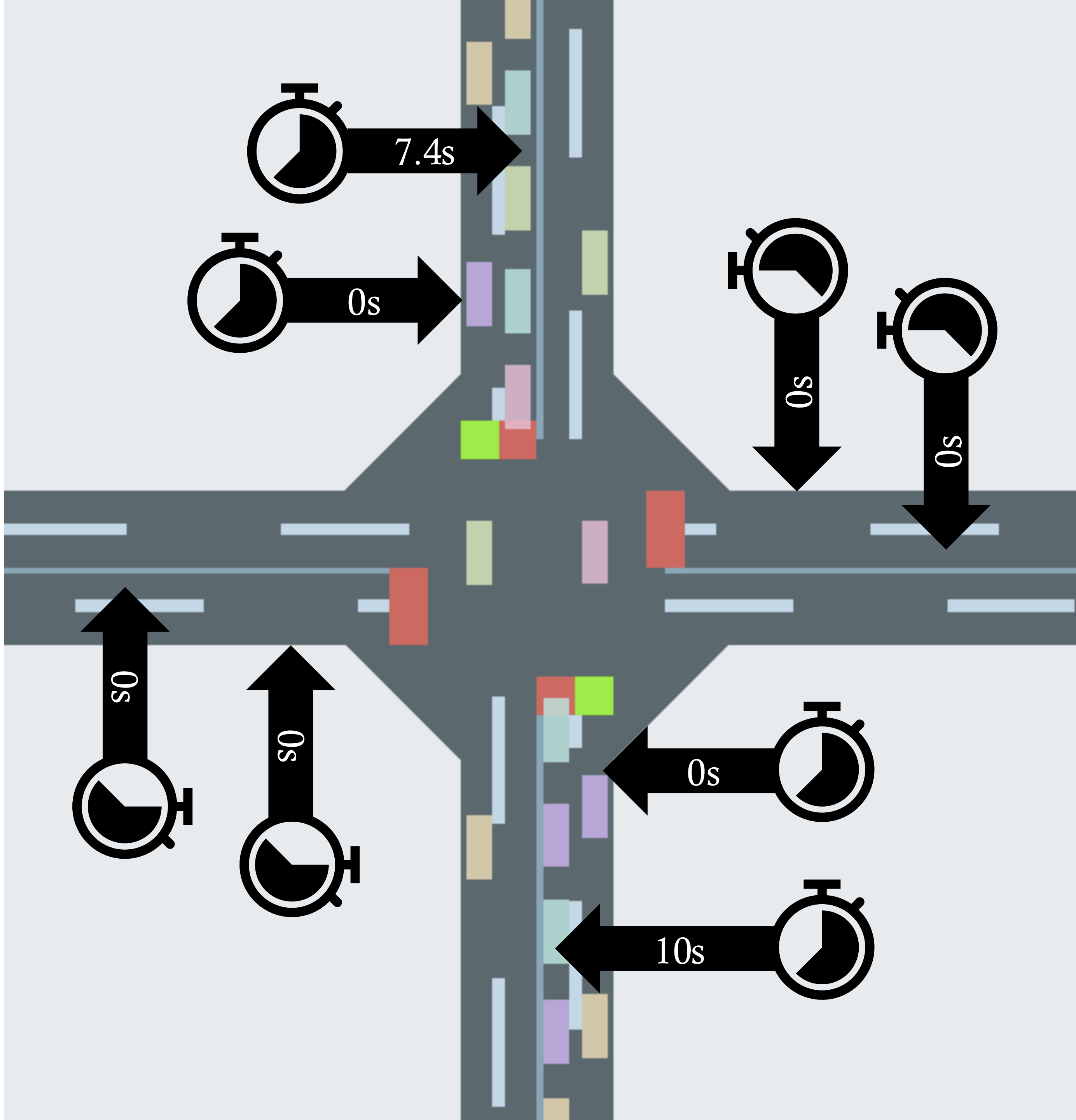
1. Number of waiting cars per lane
2. Number of cars in end lanes



ENVIRONMENT

The agent has access to the following features:

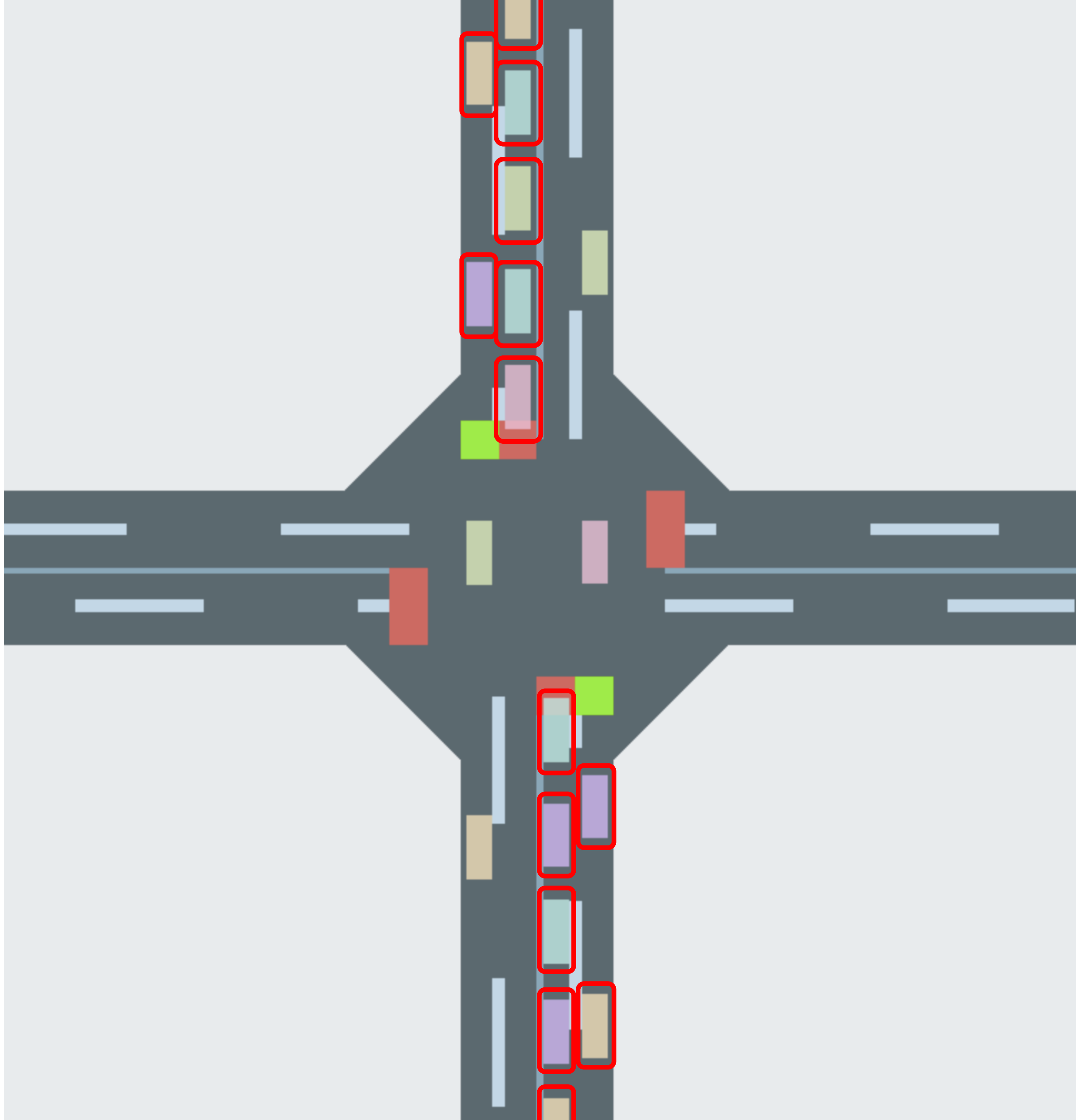
1. Number of waiting cars per lane
2. Number of cars in end lanes
3. Average waiting time per lane



ENVIRONMENT

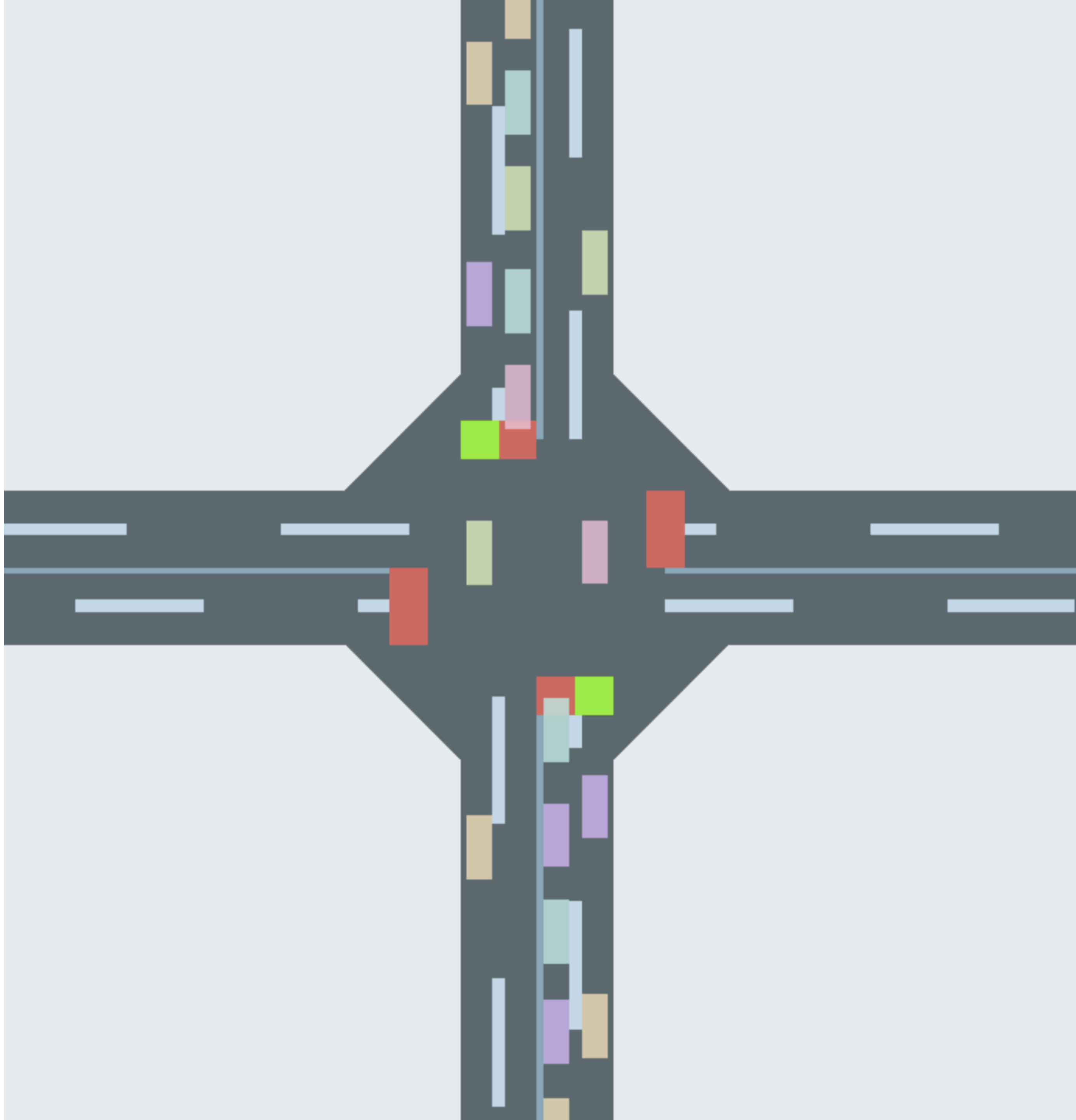
The agent has access to the following features:

1. Number of waiting cars per lane
2. Number of cars in end lanes
3. Average waiting time per lane
4. Number of cars in start lanes



ENVIRONMENT

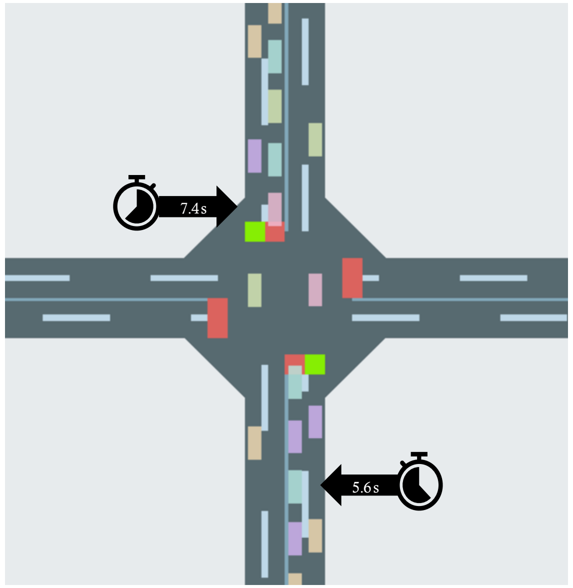
We will aim to find the most relevant features.



REWARDS

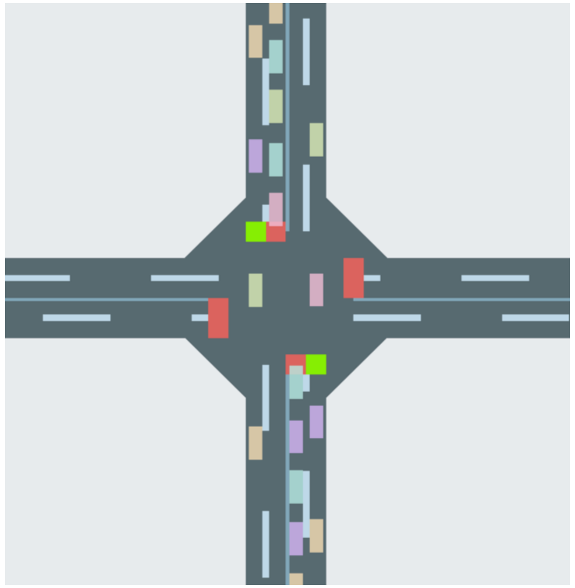
Negative sum of the average waiting time of vehicles per lane

-7.4s, -5.6s



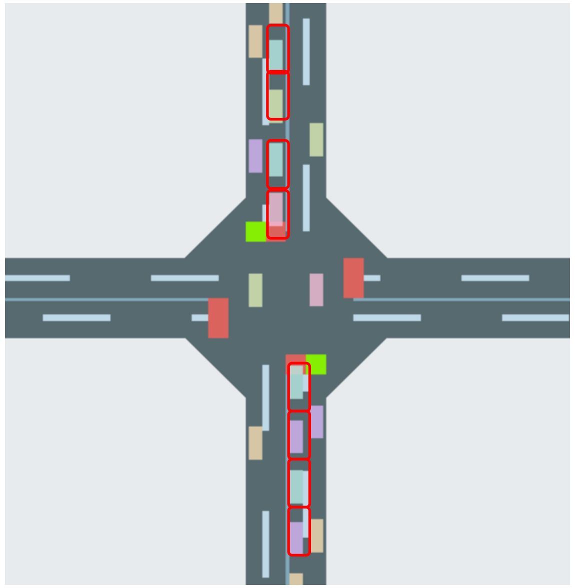
Negative sum of the average travel time

-20.5s



Negative sum of waiting vehicles

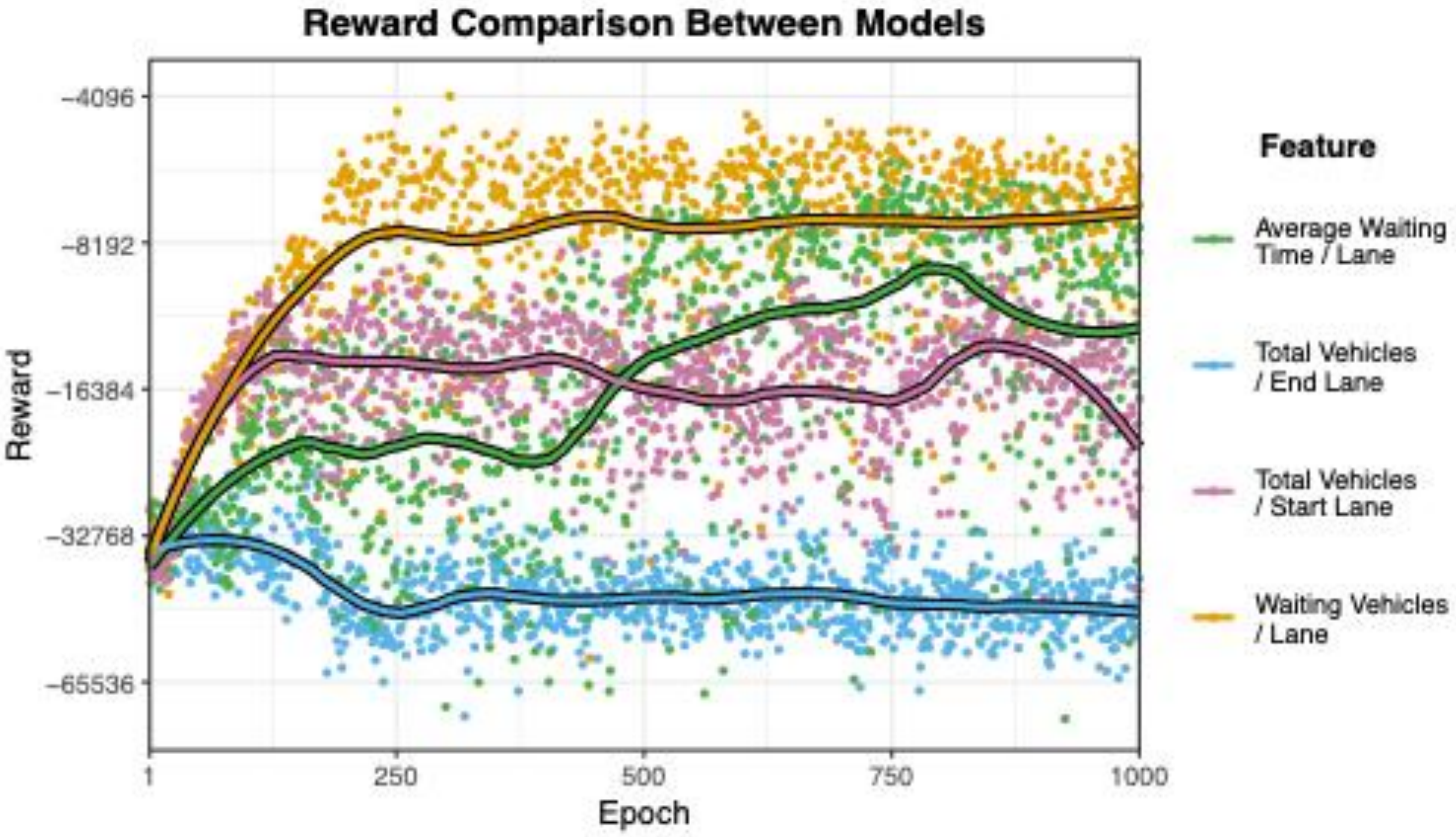
-8



DQN TRAINING

Training results using **reward 1**: negative sum of waiting vehicles

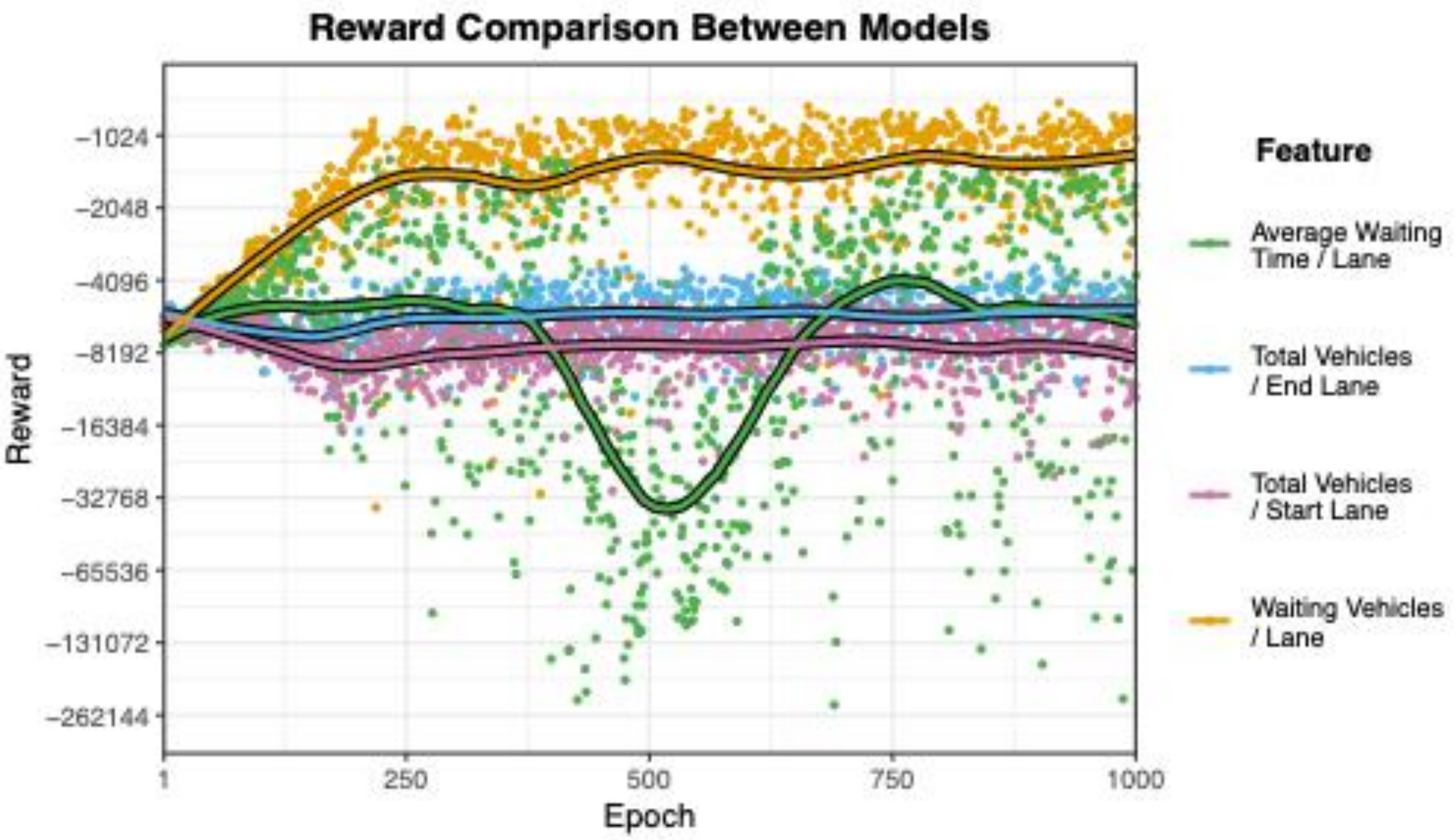
- Feature 1 performed best
- Feature 2 did not improve the reward
- Features 3 & 4 saw improvements, but slowly



DQN TRAINING

Training results using **reward 2**: Negative sum of the average waiting time of vehicles per lane

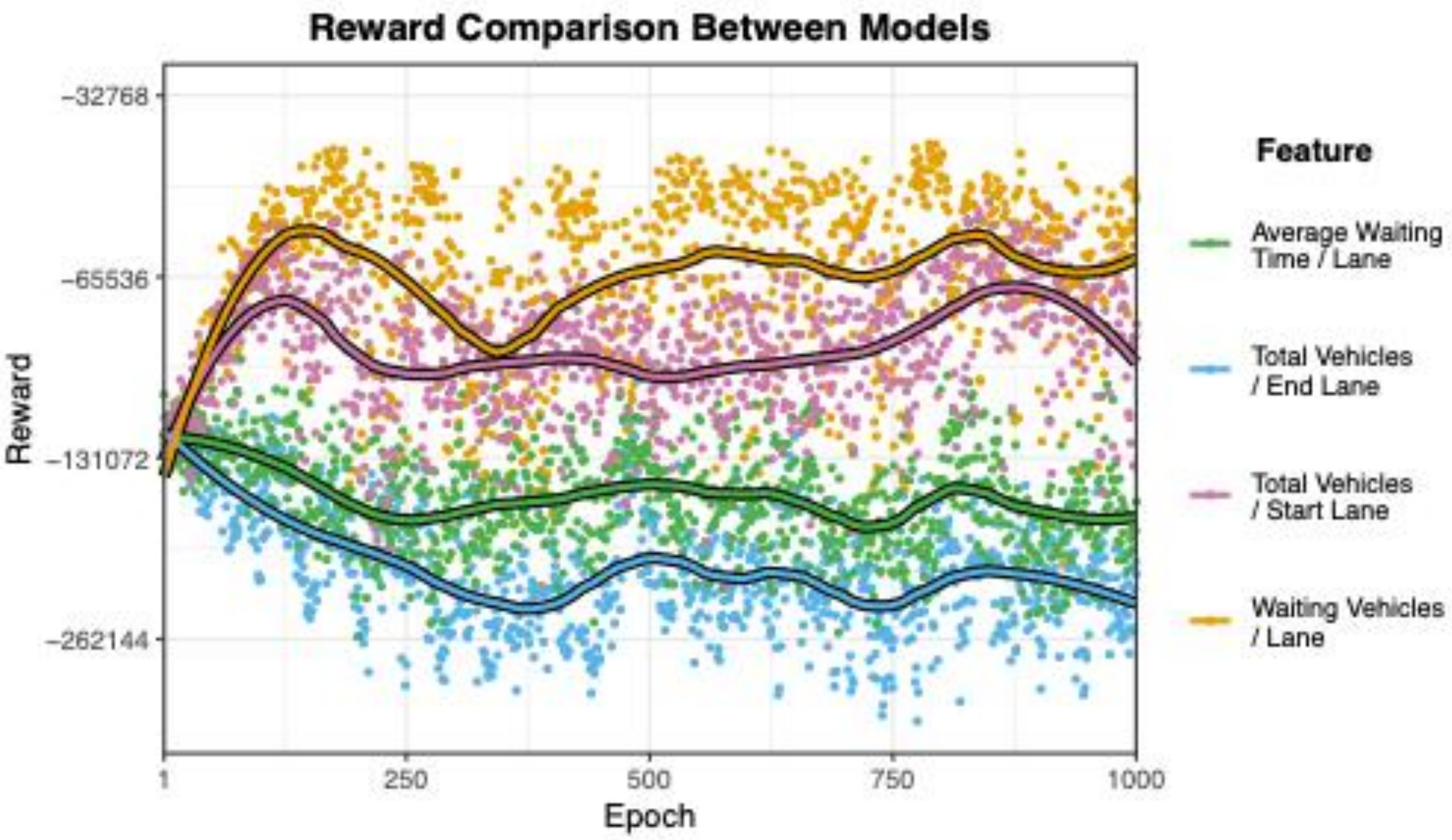
- Feature 1 performed best (again)
- Feature 2-4 saw little to no improvement



DQN TRAINING

Training results using **reward 3**: Negative average travel time of vehicles

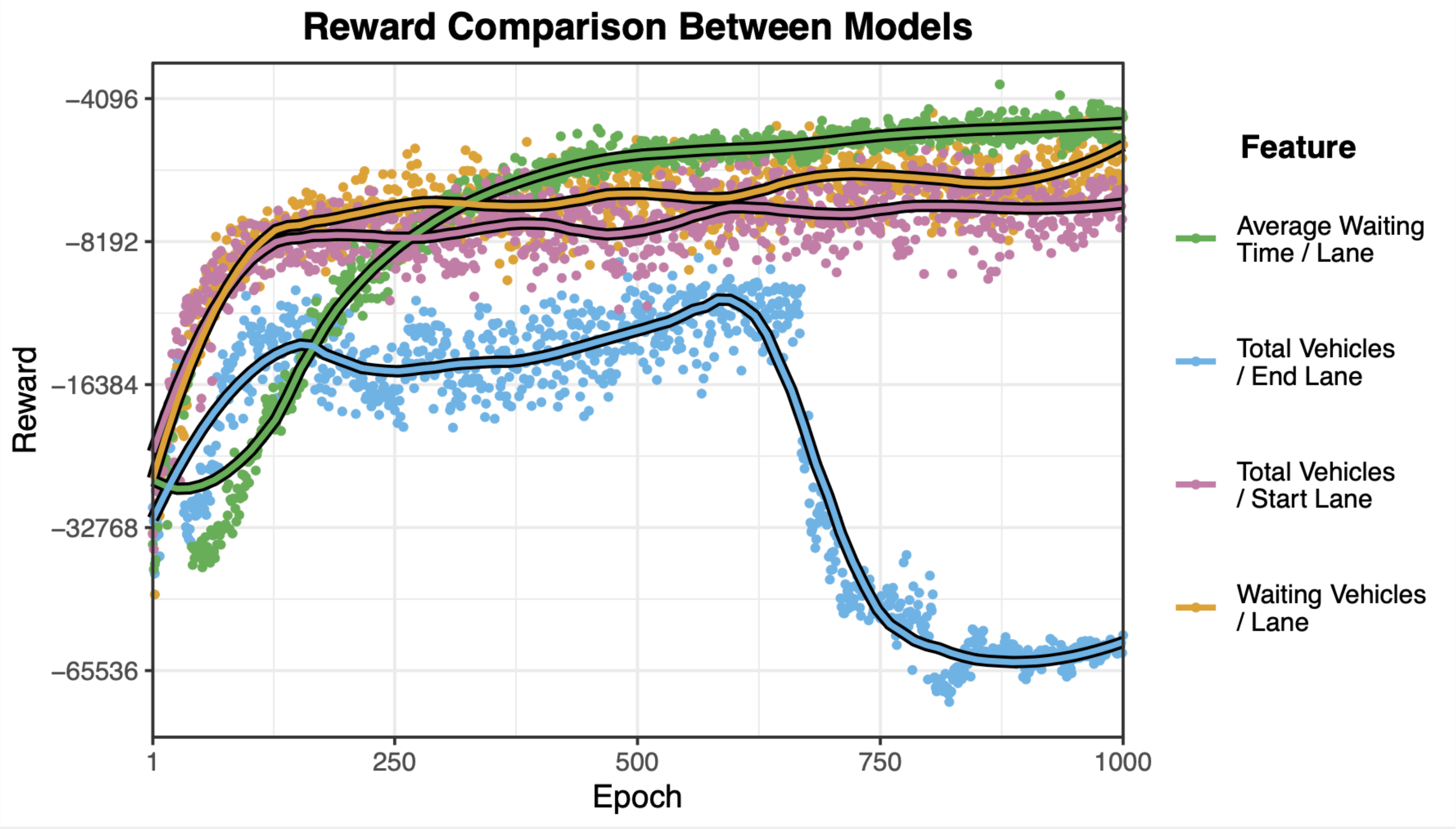
- Feature 3 & 4 performed best (once again)
- Feature 3 saw improvements
- Feature 2 & 4 got worse



PPO TRAINING

Training results using **reward 1**: negative sum of waiting vehicles

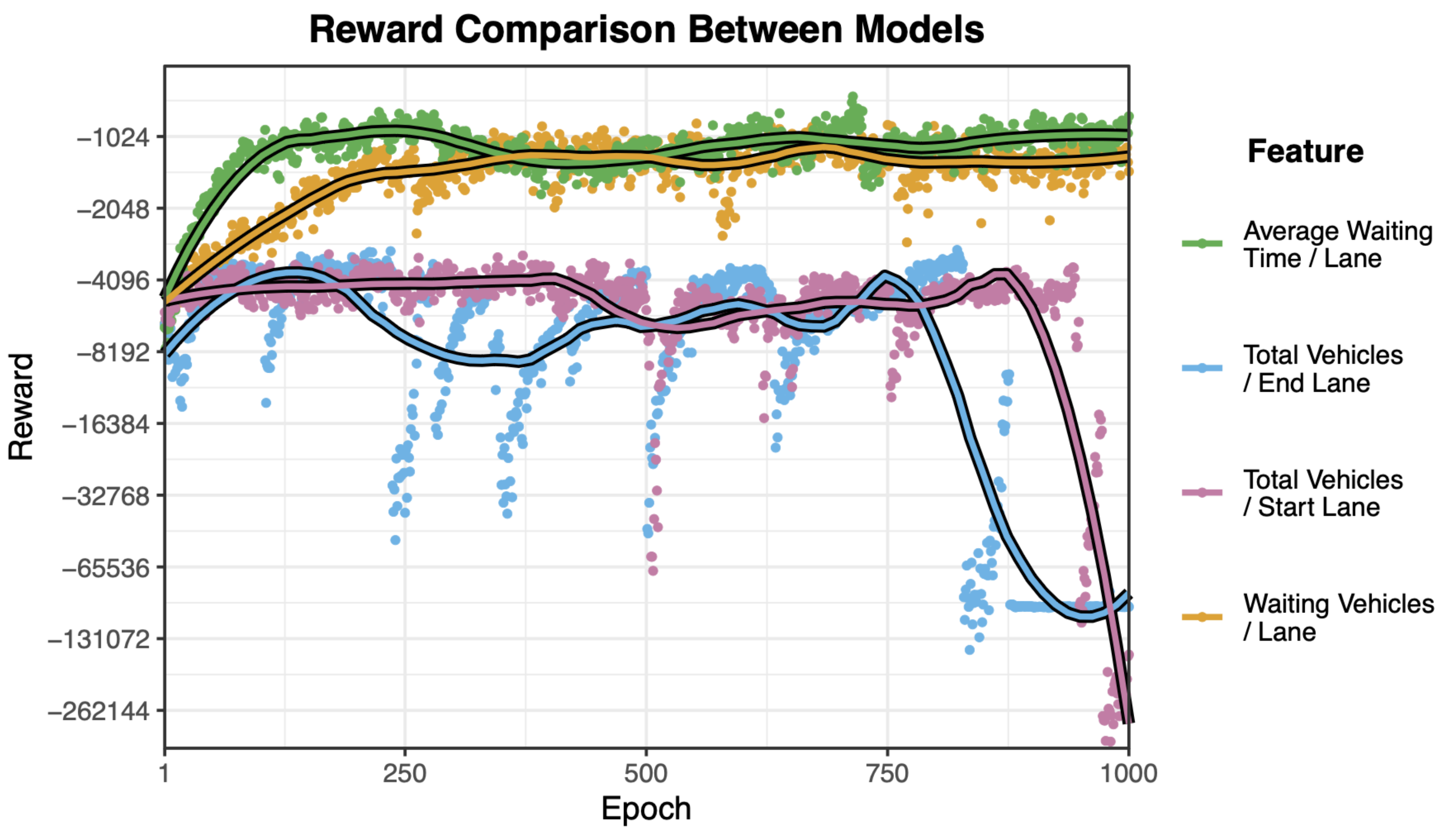
- Feature 1 (Average waiting) performed best.
- Feature 3 & 4 performed moderately well.
- Features 2 initially performed well but then dropped.



PPO TRAINING

Training results using **reward 2**: Negative sum of the average waiting time of vehicles per lane

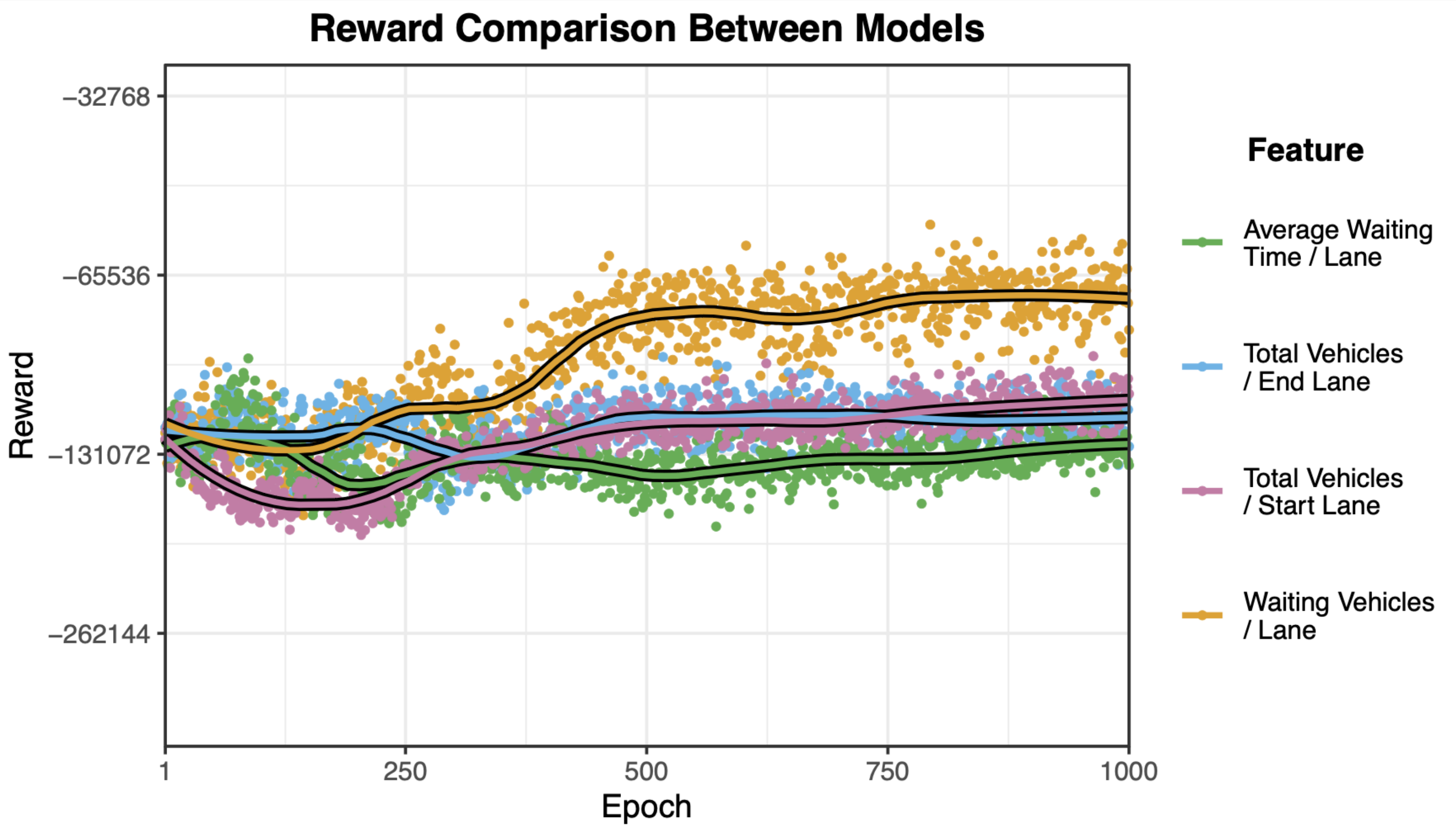
- Feature 1 & 4 performed moderately well
- Feature 2 & 3 saw no improvement



PPO TRAINING

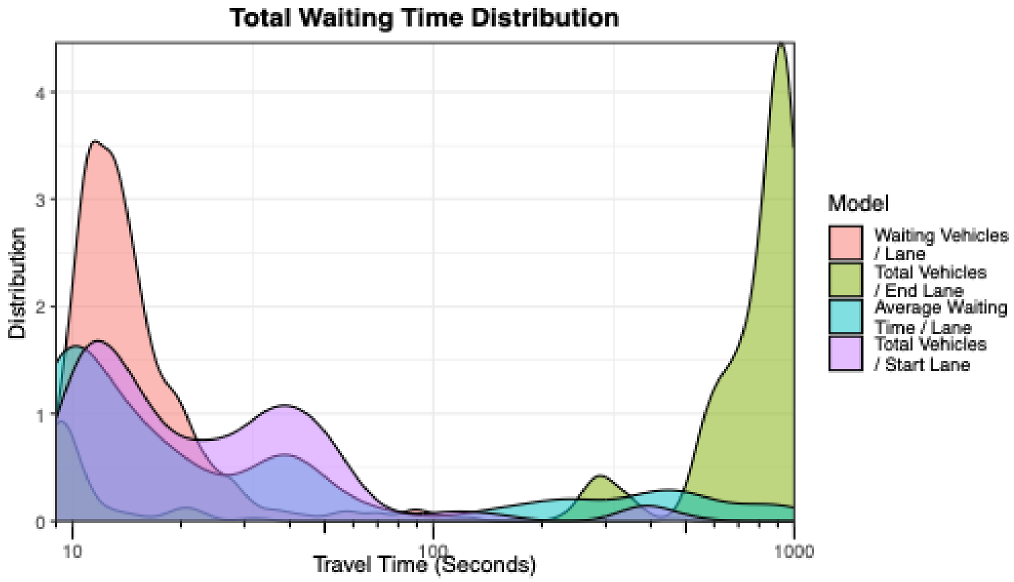
Training results using **reward 3**: Negative average travel time of vehicles

- Feature 4 performed best
- Feature 1, 2, & 3 saw minimal to no improvement

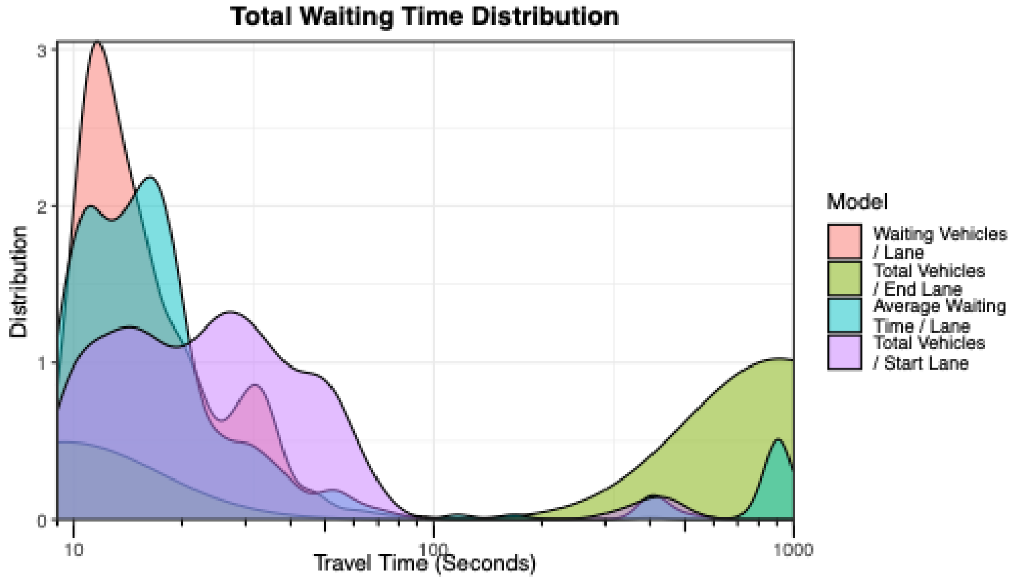


DQN EVALUATION

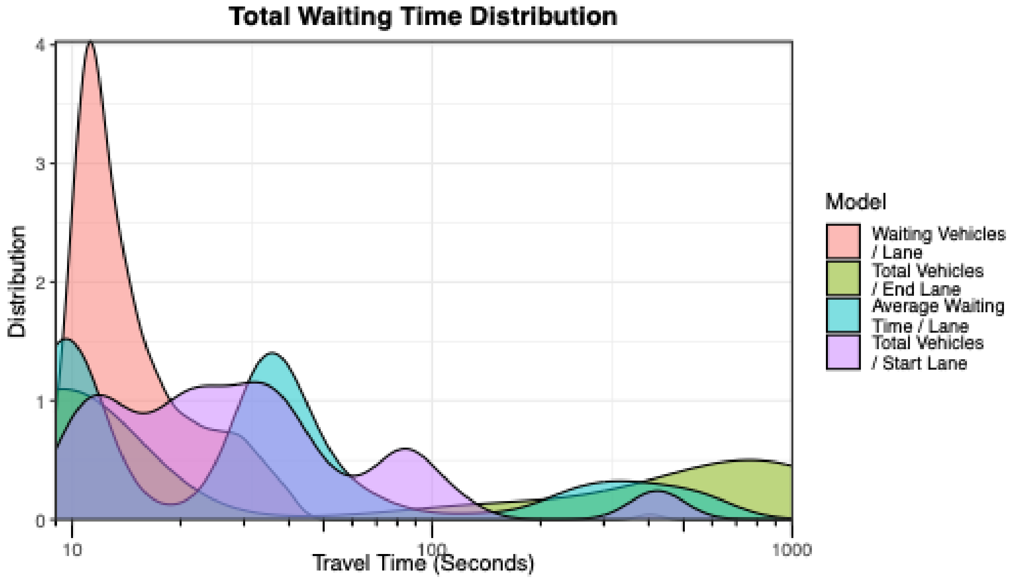
- Feature 1 (waiting vehicles / lane) consistently performed the best
- Other features fell behind



Reward 3 – average travel time

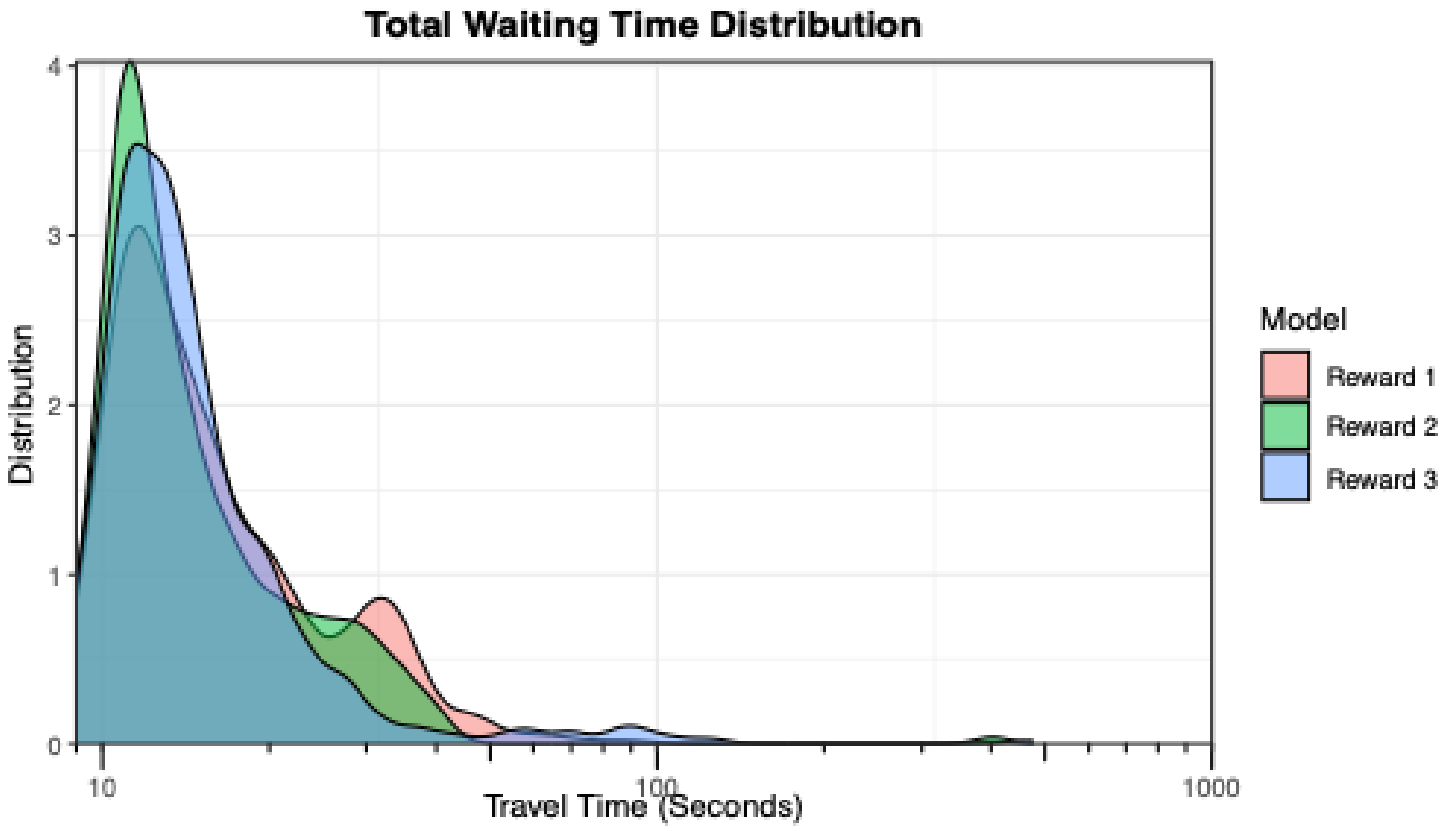


Reward 1 – # of waiting vehicles



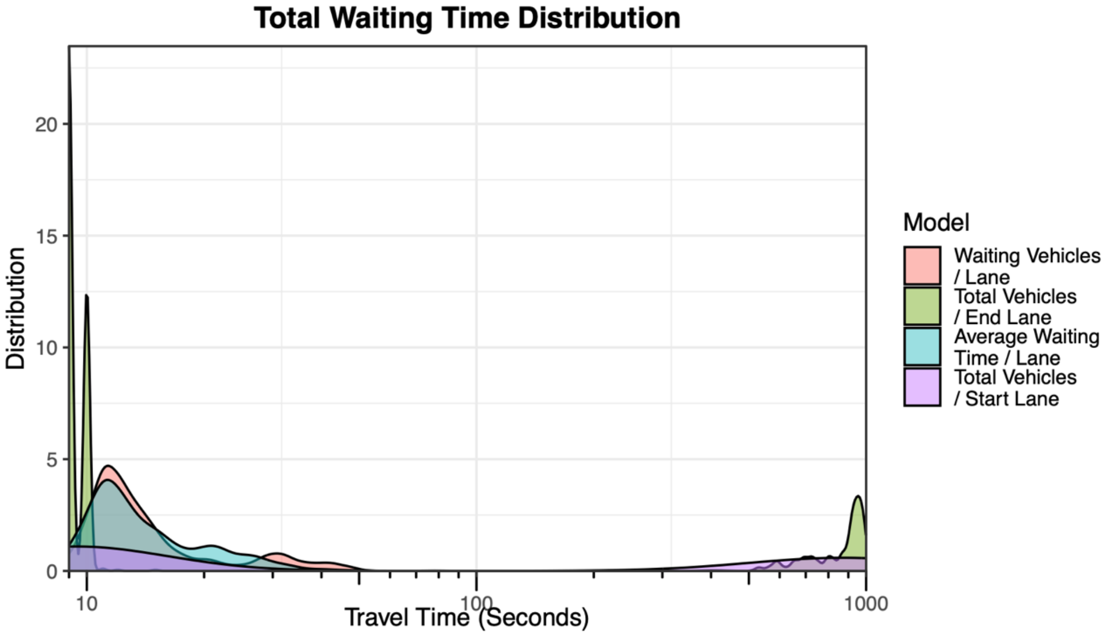
Reward 2 – mean of waiting time

DQN EVALUATION

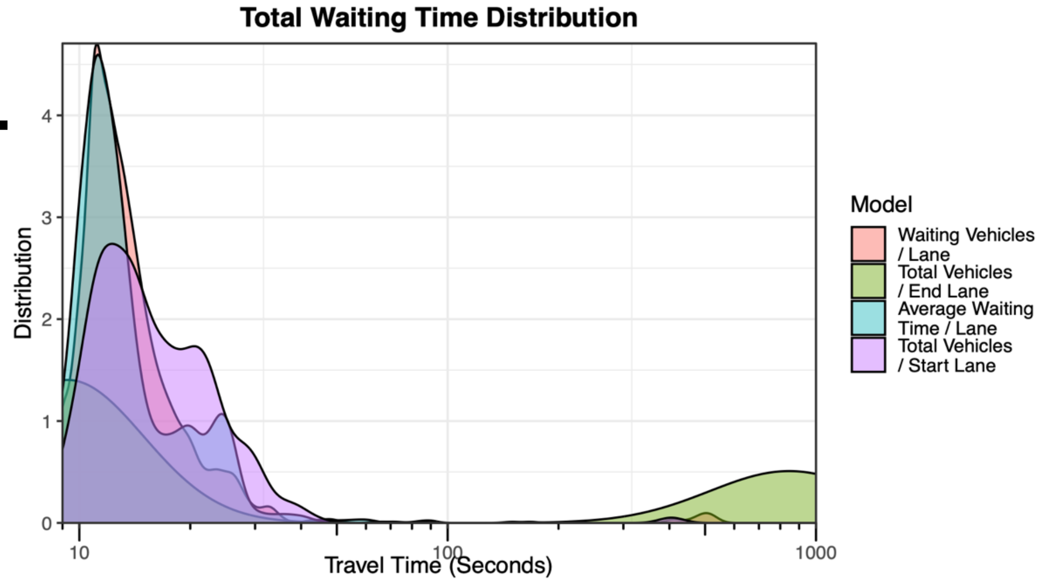


Comparing the best models (feature 1) across the different rewards

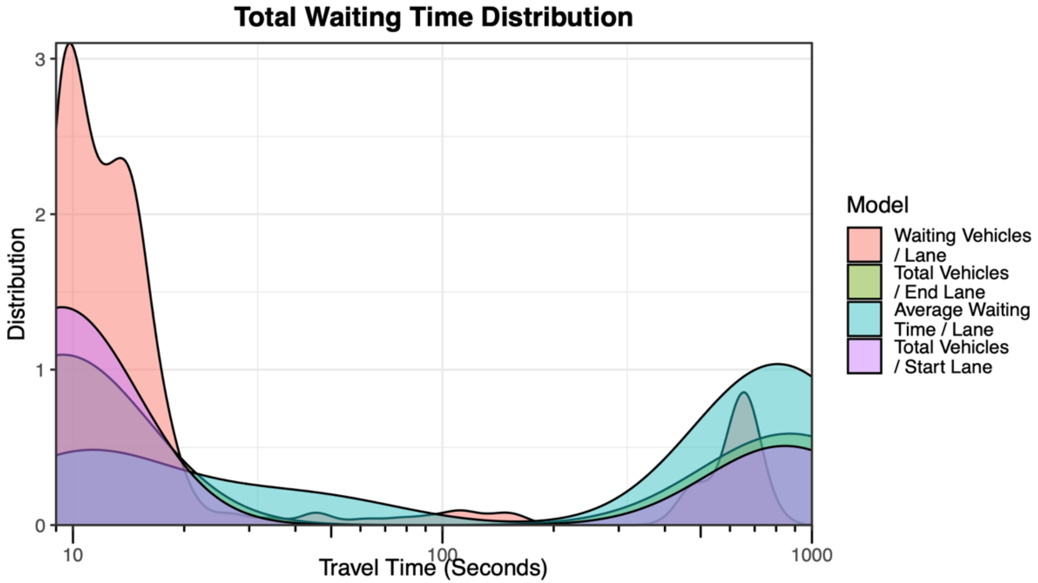
PPO EVALUATION



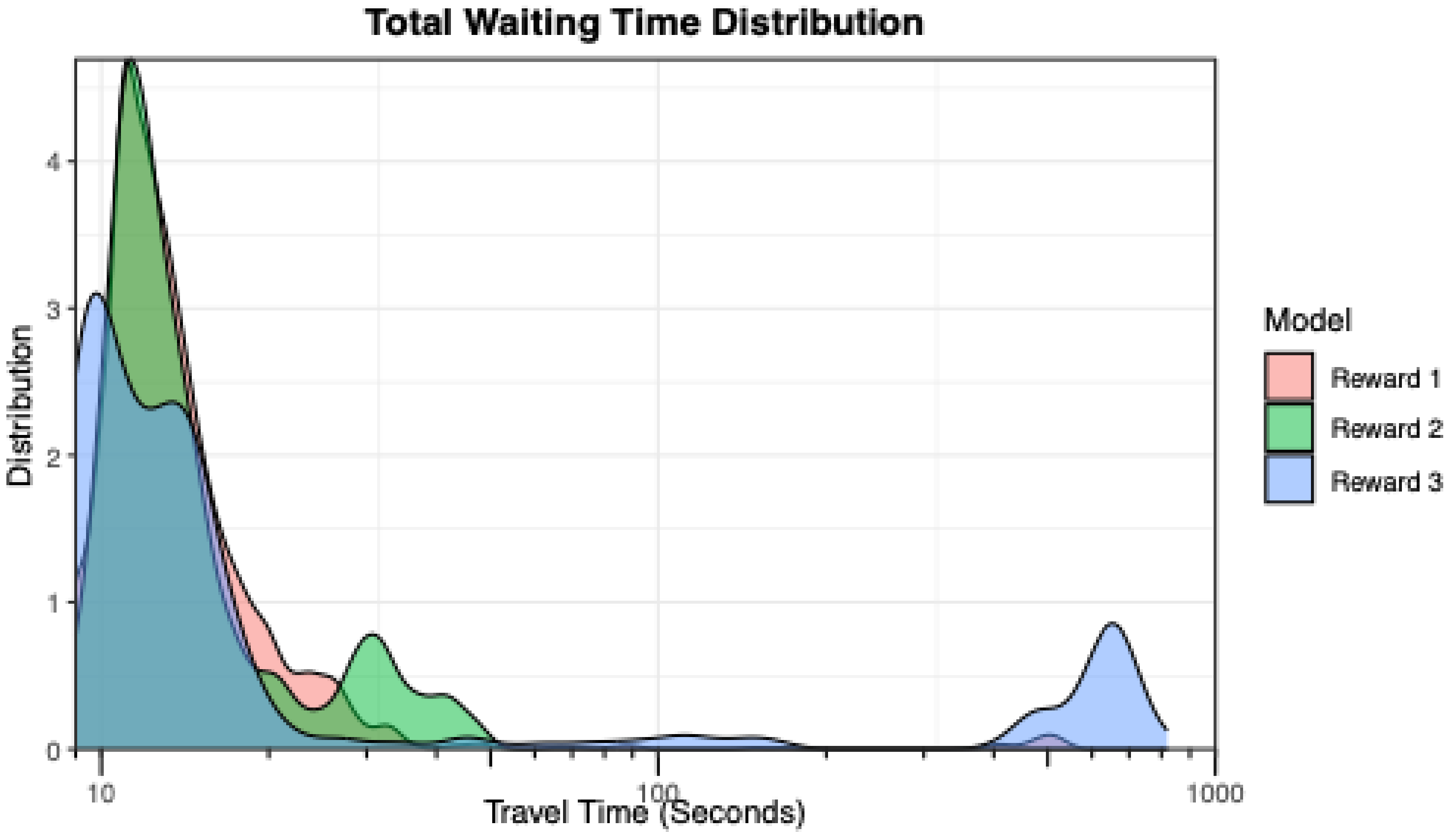
Reward 2 – mean of waiting time



Reward 1 – # of waiting vehicles



Reward 3 – average travel time



PPO EVALUATION

Comparing rewards for best model (feature 2)

Reward 2 performed the best

While reward 1 seems very similar, and has a smaller second bump, it has a decently large bump at the 400-600 seconds, where ~100 vehicles couldn't make it through.

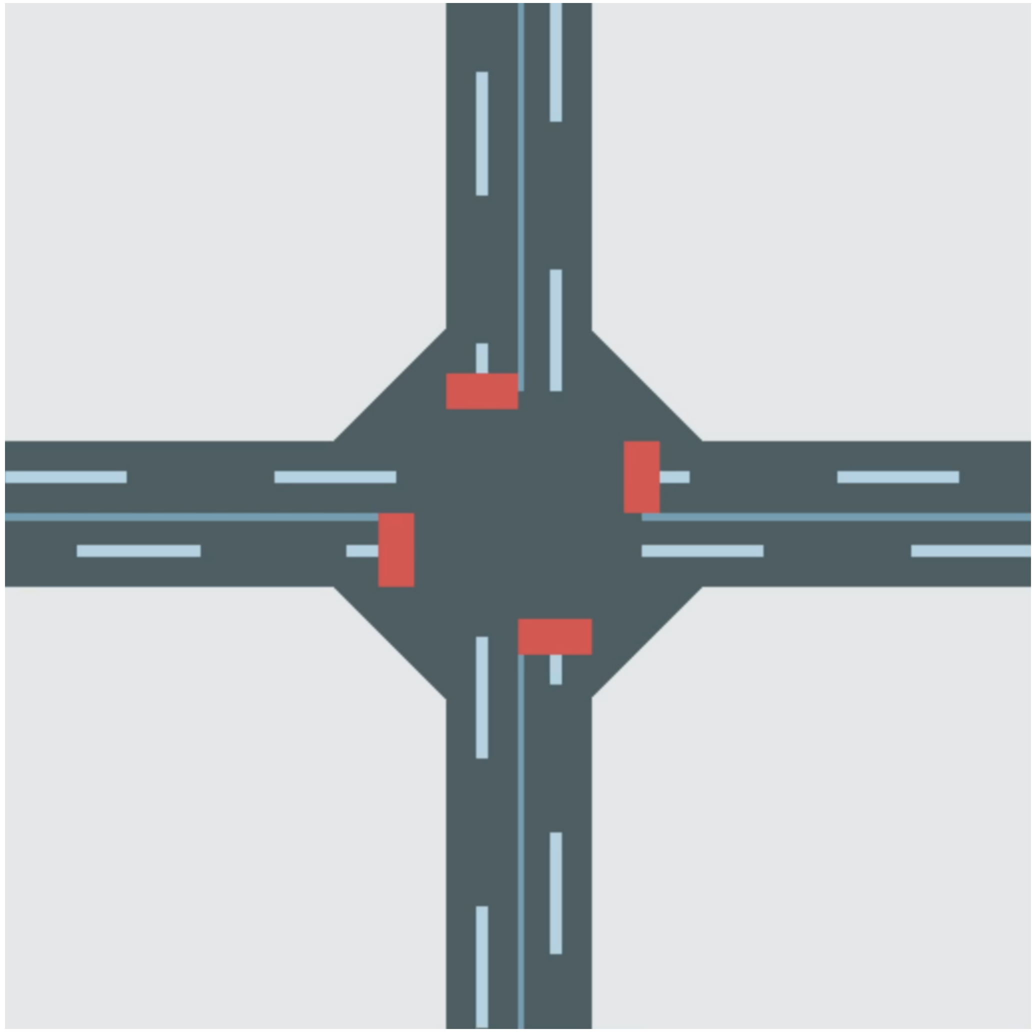
REFERENCES

[1] A. N. Laboratory, “Idling reduction for personal vehicles,” 05 2015.

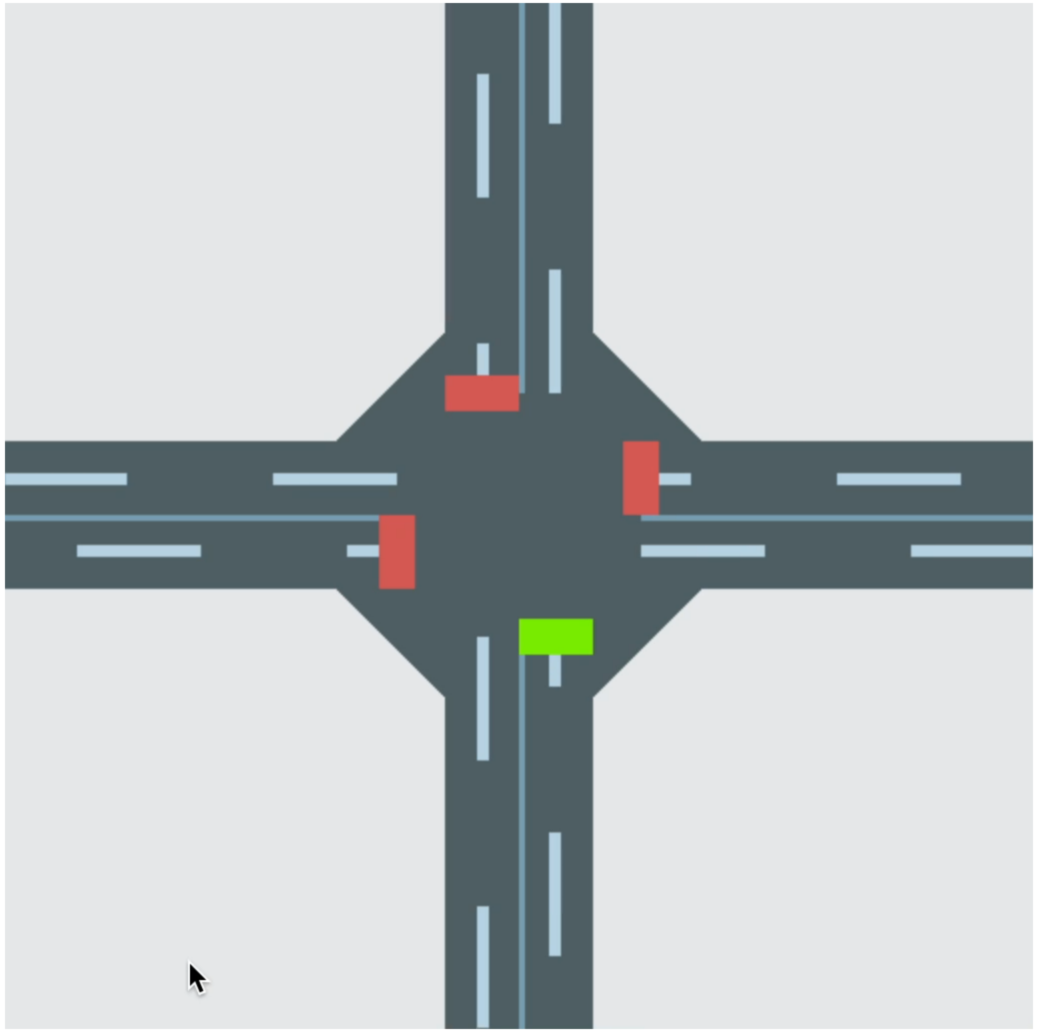
[2] Huichu Zhang, CityFlow, (2020), GitHub repository, https://github.com/cityflow-project/CityFlow

[3] Standford University, "Reinforcement Learning: An Introduction", 2014

SIMULATION!



DQN Reward 1 Feature 1



PPO Reward 2 Feature 2