# RED LIGHT, GREEN LIGHT: DYNAMIC TRAFFIC LIGHTS AT UB USING RL

Daniel Viola and Erin Gregoire

CSE 546: Reinforcement Learning with  Dr. Alina Vereshchaka

## Introduction

We are seeking to improve the efficiency of traffic surrounding UB's Medical Campus by optimizing traffic light patterns using RL. This is a busy area in downtown Buffalo, dense with traffic, especially during rush hour. Currently in this area, traffic lights are on a fixed pattern that does not allow for the complex reality of traffic flow. To enhance the experience for drivers, we put four RL agents to the test.

## Methods

These RL agents were trained on the Single Intersection and the best performers were applied to the UB Medical Campus Intersections. All agents were built from scratch, except PPO which was optimized from Stable Baselines3.

### Q-Learning

Learns via a Q-value and chooses an action based on the highest Q-value in a given state.

$$Q(s_t,a_t)=Q(s_t,a_t)+\alpha\left[r+\gamma \max_a Q(s_{t+1},a)-Q(s_t,a_t)\right]$$

### SARSA (State-Action-Reward-State-Action)

Learns via a Q-value and chooses an action based on the policy in a given state-action pair.

$$Q(s_t,a_t)\leftarrow Q(s_t,a_t)+\alpha\left[r_t+\gamma Q(s_{t+1},a_{t+1})-Q(s_t,a_t)\right]$$

### Double Deep Q-Network (DDQN)

Learns via a Q-value via a neural network and chooses an action based on the highest Q-value in a given state.

$$Q^A(s_t,a_t)\leftarrow Q^A(s_t,a_t)+\alpha\left[r_{t+1}+\gamma Q^B(s_{t+1},\arg\max_{a'} Q^A(s_{t+1},a')-Q^A(s_t,a_t)\right]$$
$$Q^B(s_t,a_t)\leftarrow Q^B(s_t,a_t)+\alpha\left[r_{t+1}+\gamma Q^A(s_{t+1},\arg\max Q^B(s_{t+1},a'))-Q^B(s_t,a_t)\right]$$

### Proximal Policy Optimization (PPO)

An actor-critic method that learns an advantage value via a neural network and chooses actions based on the highest probability.

$$J^{\text{CLIP}}(\theta)=\hat{\mathbb{E}}_t\left[\min\left(r(\theta)\hat{A}_{\theta_{\text{old}}}(s,a),\text{clip}(r(\theta),1-\epsilon,1+\epsilon)\hat{A}_{\theta_{\text{old}}}(s,a)\right)\right.$$

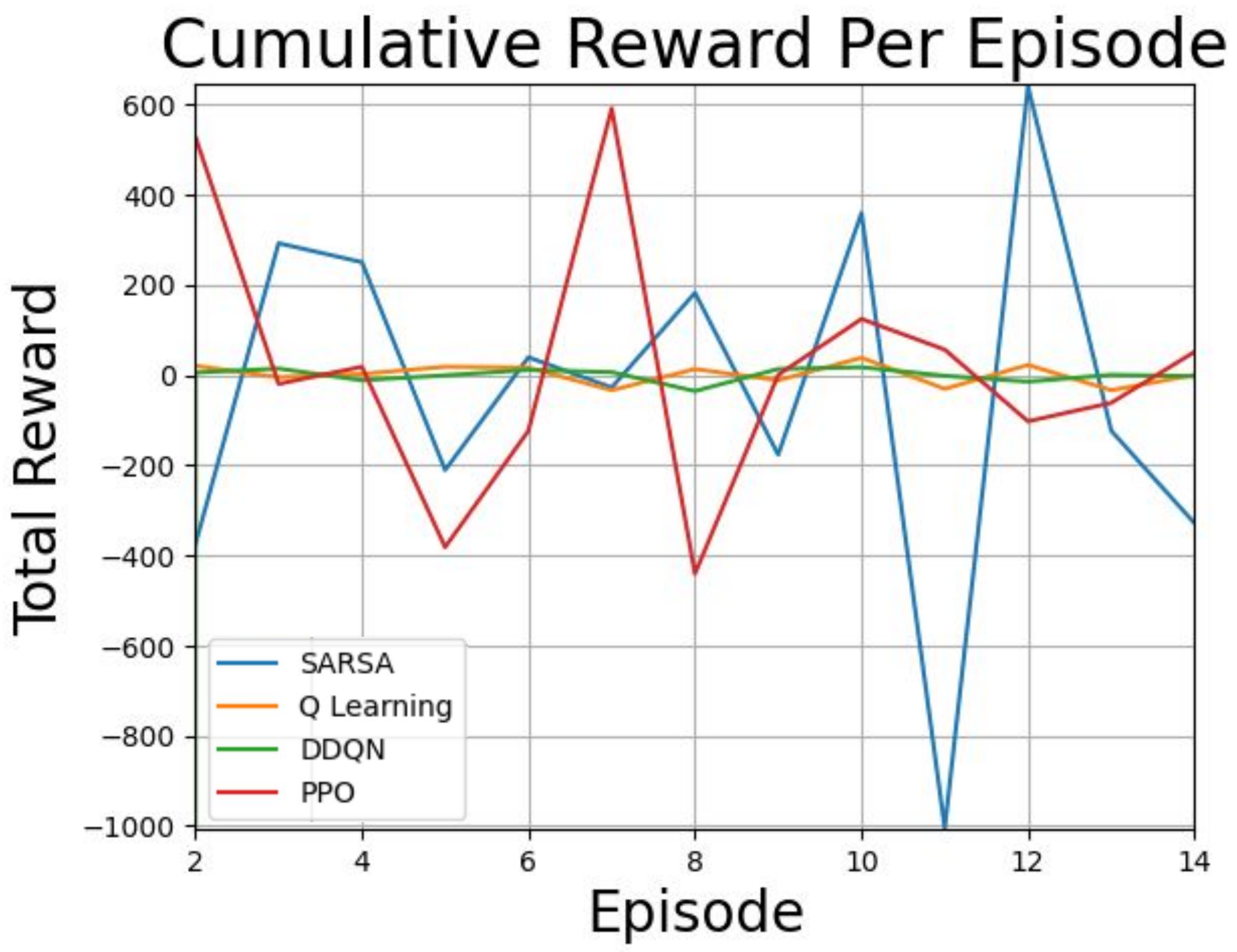
Metrics used to evaluate the agent's performance include the average wait time of per vehicle at a red light and cumulative reward of the agent earned per episode. Reward was earned by subtracting the total vehicle waiting time from the previous timestep's total vehicle waiting time.

## Single Intersection Simulation

The RL models were tested on a simple, single intersection to understand their functionality with dynamic traffic lights.

- Agents: Q-Learning, SARSA, DDQN, PPO
- No. of Intersections: 1

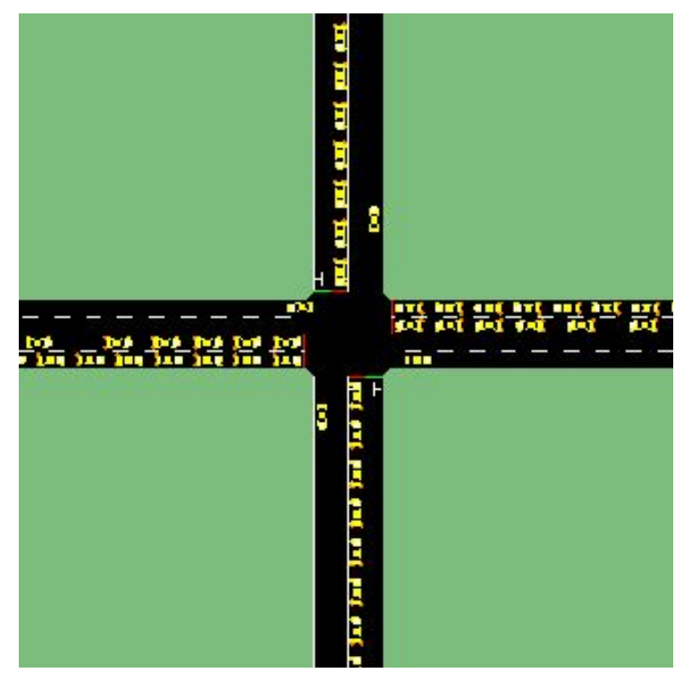
- Environment Type:
  - Single Agent
  - Uniform
  - Simulated



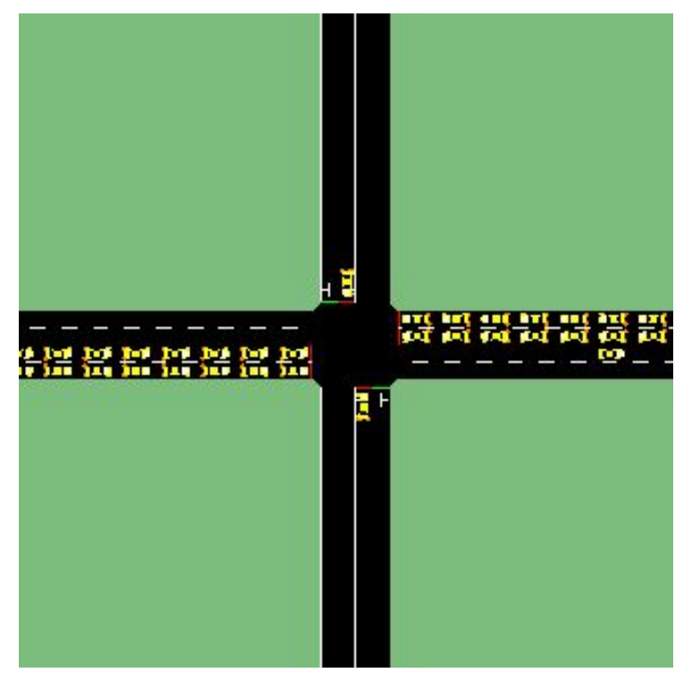
## Single Intersection Results

All four agents were able to learn the best policy to reduce vehicle waiting time and maximize reward during training. Based on the cumulative reward earned per episode during evaluation, as seen in the graph above, the Q-Learning and DDQN agents received the best stability. Due to this, these two agents will be trained and evaluated on the more complex UB Medical Campus traffic intersections.

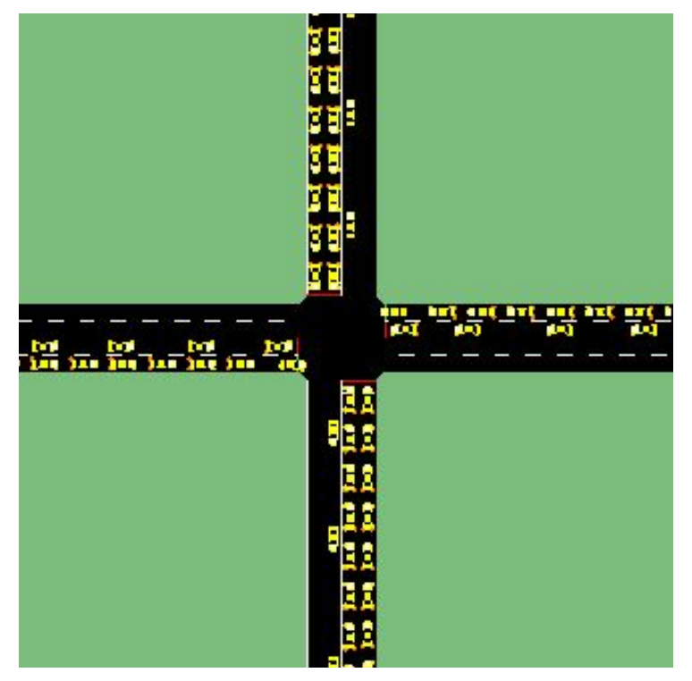
### Q-Learning vs. Random Agent



Q-Learning: Training



Q-Learning: Evaluation



Random Agent

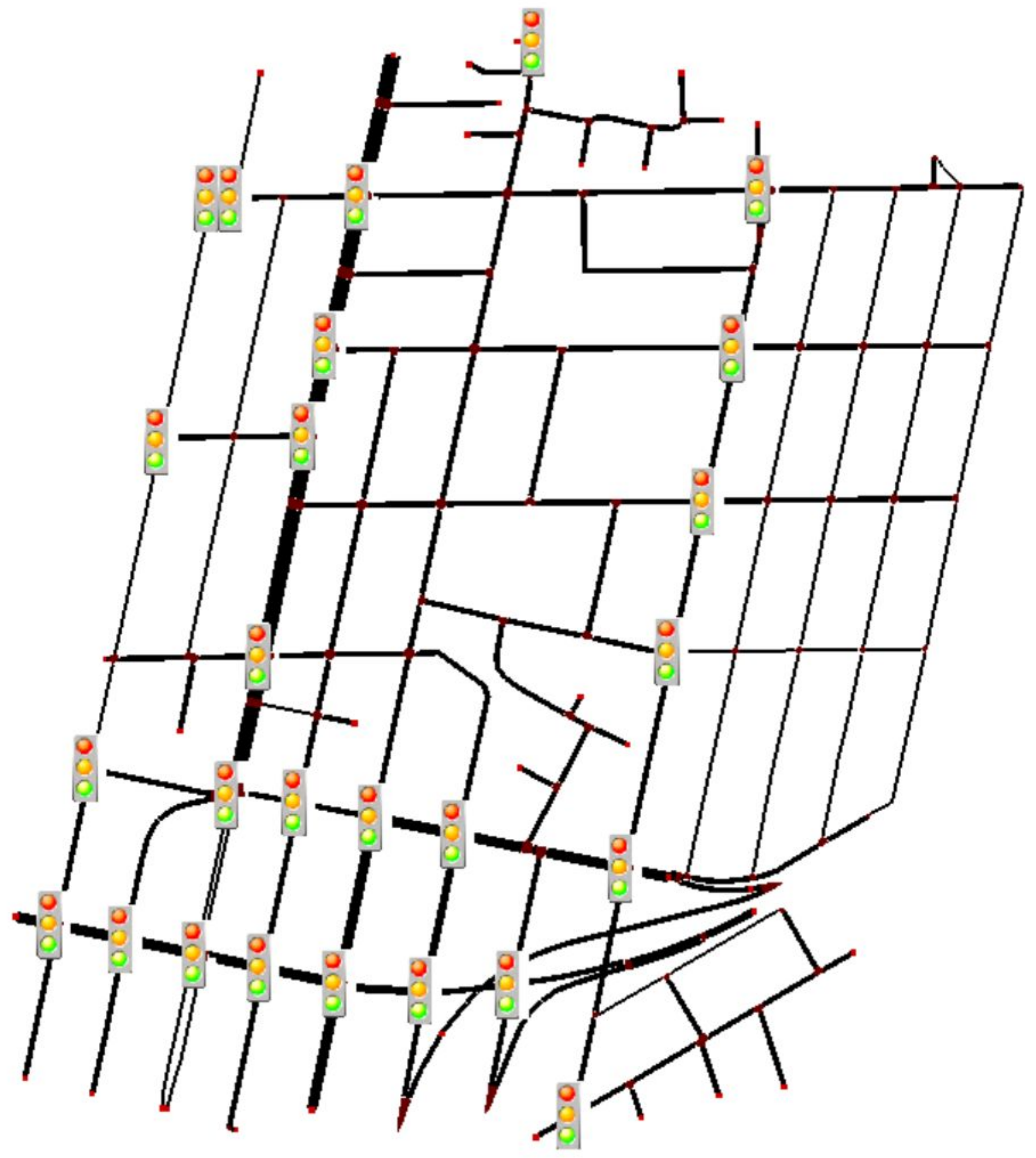
High density traffic is presented at the horizontal lanes to increase challenge. The Q-Learning agent is able to learn the best policy and greatly exceeds the Random agent.

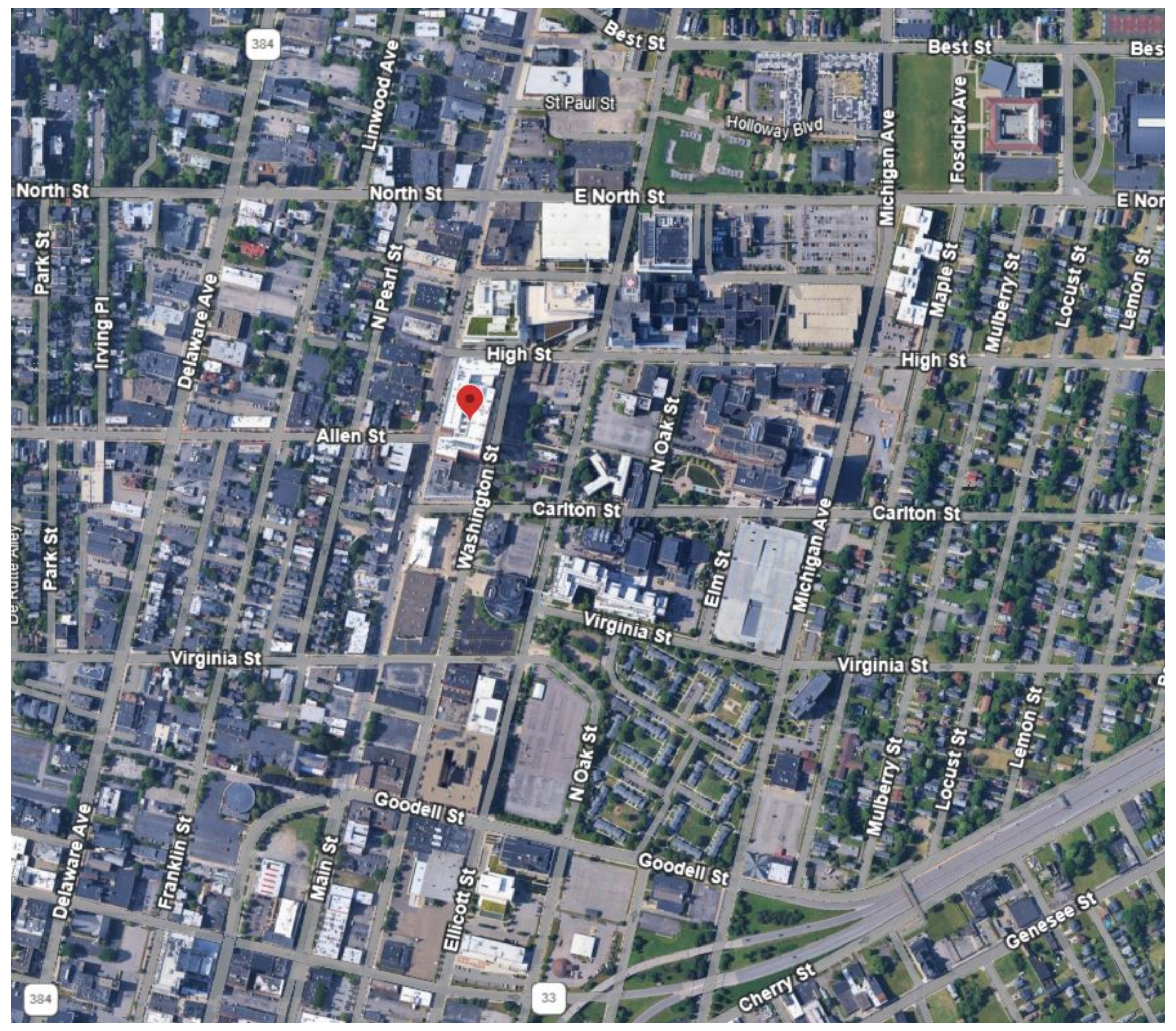
## UB Medical Campus Simulation

We built a custom traffic intersection of UB's Medical Campus using Netedit and manual construction. The model environment features the intersection surrounding Jacobs School of Medicine and realistic traffic flows based on City of Buffalo road data. Both DDQN and Q-Learning agents were trained and evaluated on this environment, but Q-Learning far exceeded performance. Thus, results will focus on Q-Learning.

- Agents: Q-Learning, DDQN
- No. of Intersections: 26

- Environment Type:
  - Multi-Agent
  - Dynamic
  - Real-world



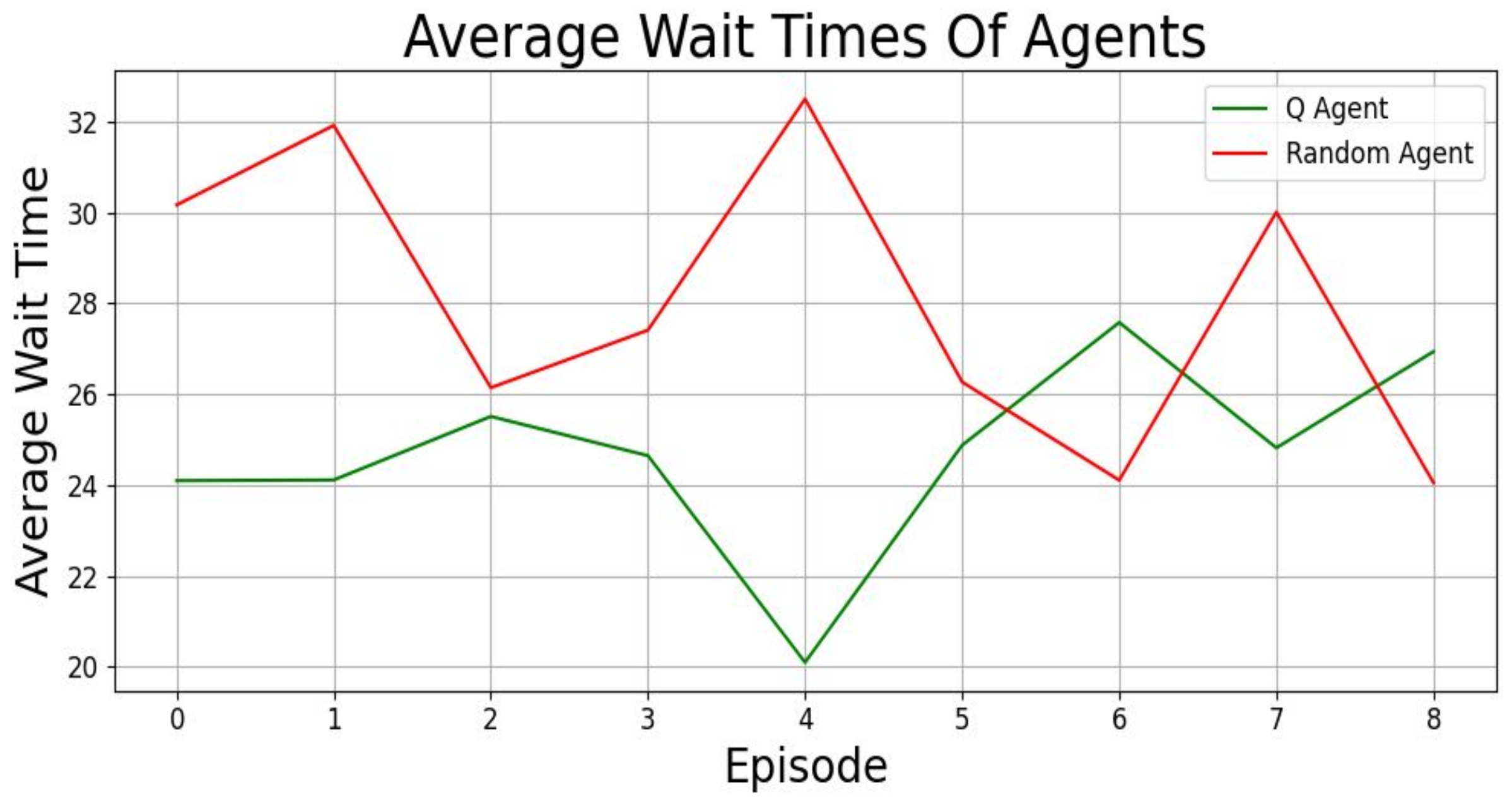


## Our Contributions

Our Q-Learning agent learned the best policy for optimizing traffic lights dynamically at UB's Medical Campus. These findings, if applied, could drastically reduce the average vehicle waiting time at these intersections in downtown Buffalo.

To compare our results to a baseline, we used a random agent that works on fixed time patterns, representing a traffic light process similar to the medical campus area. The average wait time, measured in seconds, for any given car to reach its destination through the map is seen below.

### Q-Learning Agent vs. Random Agent



These results show that the Q-Learning agent outperformed the random agent in nearly every trial by a substantial margin.

## Conclusion

Our RL approach to solving the dynamic traffic light problem was able to provide significant benefits over the simulated baseline of fixed traffic patterns. Q-Learning proved to be an effective solution to reduce vehicle waiting times.

Further development may be possible with continued training and more complex RL methods. One area for research is working to improve the response speed of the agent so that the traffic light patterns change with even less dead time than it is currently, even though it has already substantially improved this over baseline functionality.

We are proud of our strong results and contributions for the implementation of our Q-Learning agent into traffic lights at UB's Medical Campus and look forward to how we can bridge the gap from this project to real world implementation.

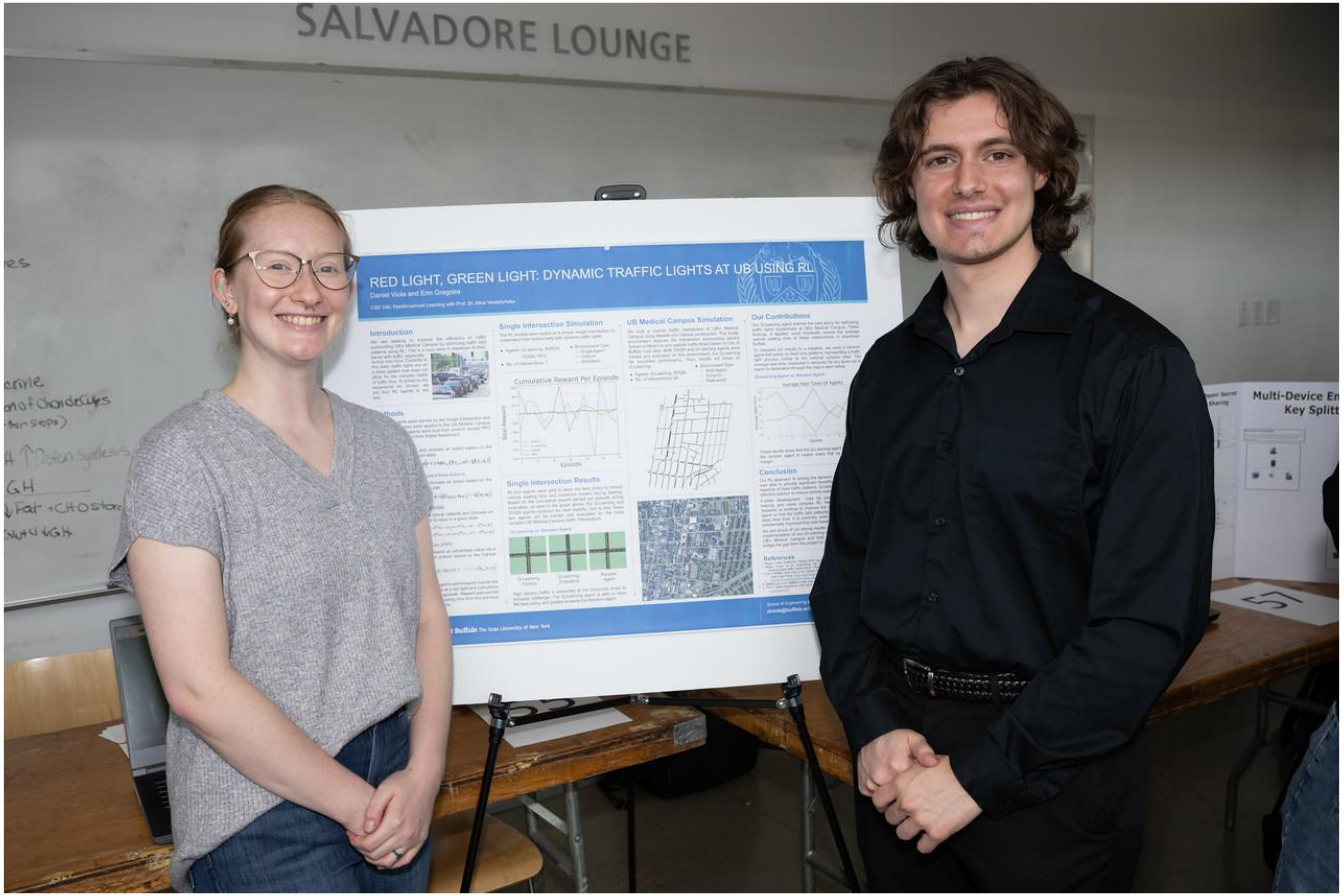
## References

1. Alegre, Lucas. "SUMO-RL." *Github,* 2019. https://github.com/LucasAlegre/sumo-rl.
2. Alegre, Lucas, et al. "Quantifying the impact of non-stationarity in reinforcement learning-based traffic signal control." *PeerJ Computer Science.* 27, May 2021.
3. City of Buffalo Open Data. (n.d.). *Annual Average Daily Traffic Volume Counts*. Buffalo Open Data Portal. https://data.buffalony.gov/Transportation/Annual-Average-Daily-Traffic Volume-Counts/y93c-u65y/about_data
4. BBBike.org. (n.d.). *BBBike: A bike route planner for many cities*. https://www.bbbike.org/
5. "PPO." *Stable Baselines3, 2021*-2025. https://stable-baselines3.readthedocs.io/en/master/modules/ppo.html

University at Buffalo The State University of New York

School of Engineering and Applied Sciences

**dviola@buffalo.edu & ergregoi@buffalo.edu**


SALVADORE LOUNGE
RED LIGHT, GREEN LIGHT: DYNAMIC TRAFFIC LIGHTS AT UB USING RL
Daniel Viola and Erin Gregoire
Introduction
Single Intersection Simulation
UB Medical Campus Simulation
Our Contributions
Cumulative Reward Per Episode
Single Intersection Results
Conclusion
References
Multi-Device En
Key Splitt

Erin Gregoire and Daniel Viola Presenting at CSE Demo Days