Interview

Adaptive Learning for network-wide traffic management

Zhizhou Fang 16 Sep 2022

Assignment

- Goal: to maximise the traffic flow by control the density on the main road.
- According to Greenshield's model, the traffic flow reaches the maximum when density equals the critical density. The critical density ρ_c is set to 35.
- Rewards: the rewards is computed based on the estimated flow.

Assignment

Method: Deep Q-Learning with MLP

- State: current density on the main road & previous target flow. The density is converted to binary representation.
- Action Space: The action space is discretised into 19 actions: a_0 to a_{18}. The action a_i denotes setting the target flow value to $100 \times i$. The final target flow is clipped into the range $[60, 1800]$.
- An MLP with skip connections learns to map the state to the Q-values of the 19 actions. During the evaluation, the action with the highest Q-value will be selected.

Assignment

Method: Deep Q-Learning with temporal convolutional network (TCN)

- State: a stack of the densities and the target flows of the last 4 time steps. Inspired by the trick of Atari breakout game.

- Action Space: The action space is discretised into 19 actions: a_0 to a_{18}. The action a_i denotes setting the target flow value to $100 \times i$. The final target flow is clipped into the range $[60, 1800]$.

- A TCN block with the gated activation and skip connections (proposed by Wavenet [3]) learns to map the state to the Q-values of the 19 actions.

Assignment

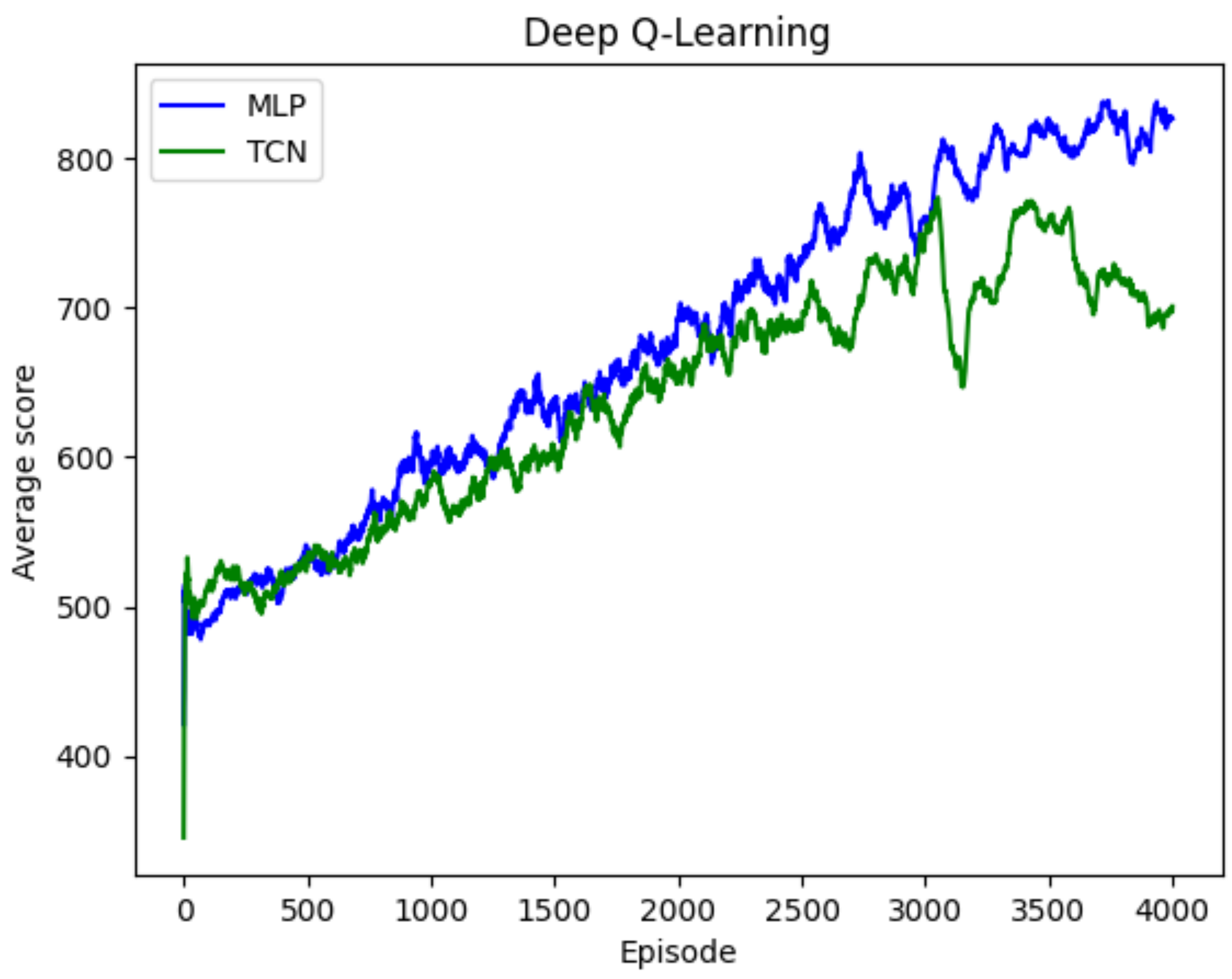
Method: Deep Deterministic Policy Gradient (DDPG)

- State: the current density on the main road with the previous target flow.
- Continuous action space: the action value ranges from $[-1,1]$, which will be mapped to the corresponding target flow ranging from $[60,1800]$.
- An MLP (actor) learns to map the state to the continuous action space. Another MLP (critic) learns the Q-values of the state and the corresponding action output by the actor. The actor is trained to maximise the Q-values of the critic.

Assignment

Training: Deep Q-Learning

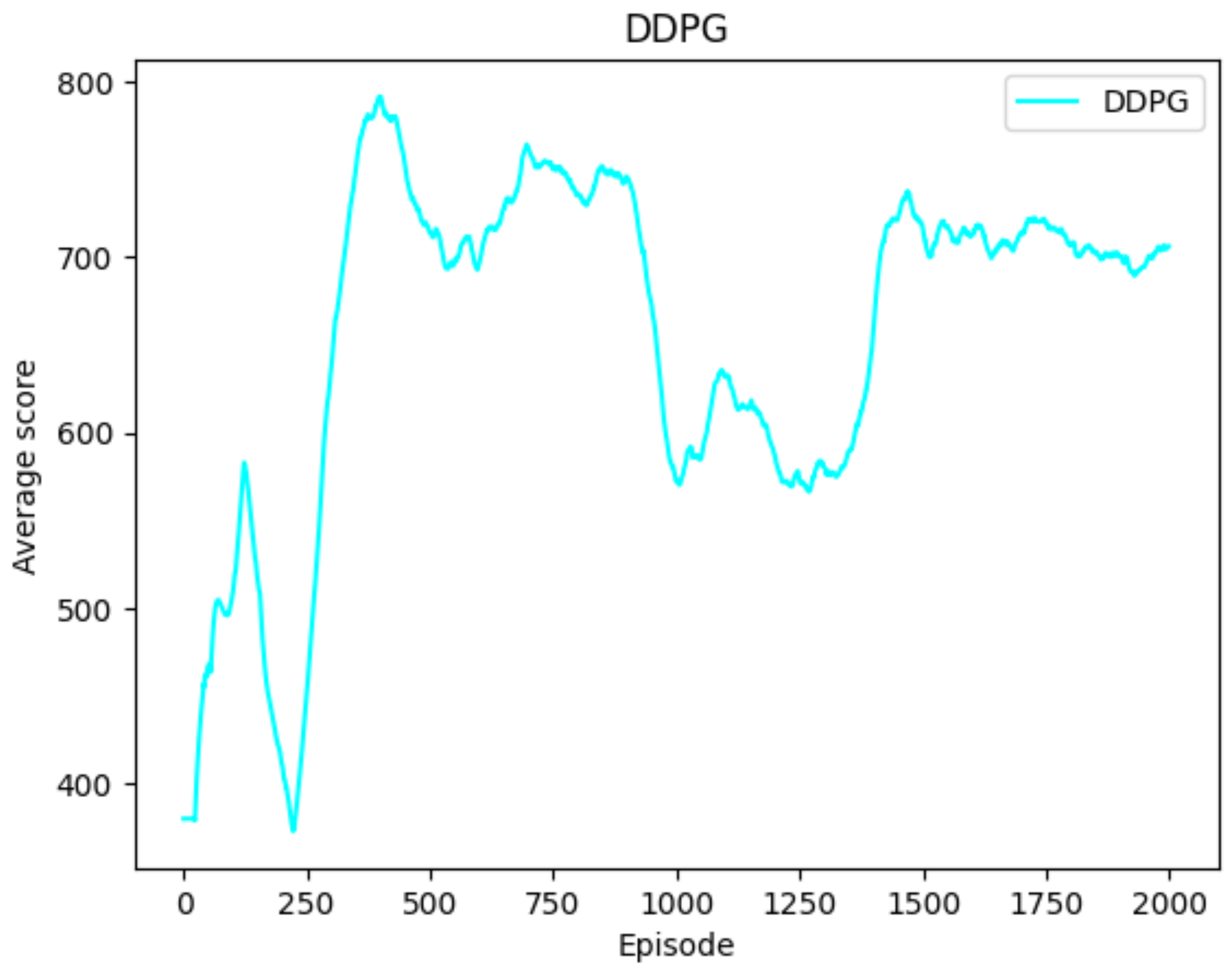
The two DQN models are trained for 4000 episodes.



Assignment

Training: DDPG

The DDPG model is trained for 2000 episodes.



Assignment

Baselines

- No-Control: The target flow is always set to the capacity (1800).
- On-Off: When the density equals to the critical density (35), set the target flow to 0. Otherwise the target flow is set to the capacity.

Assignment

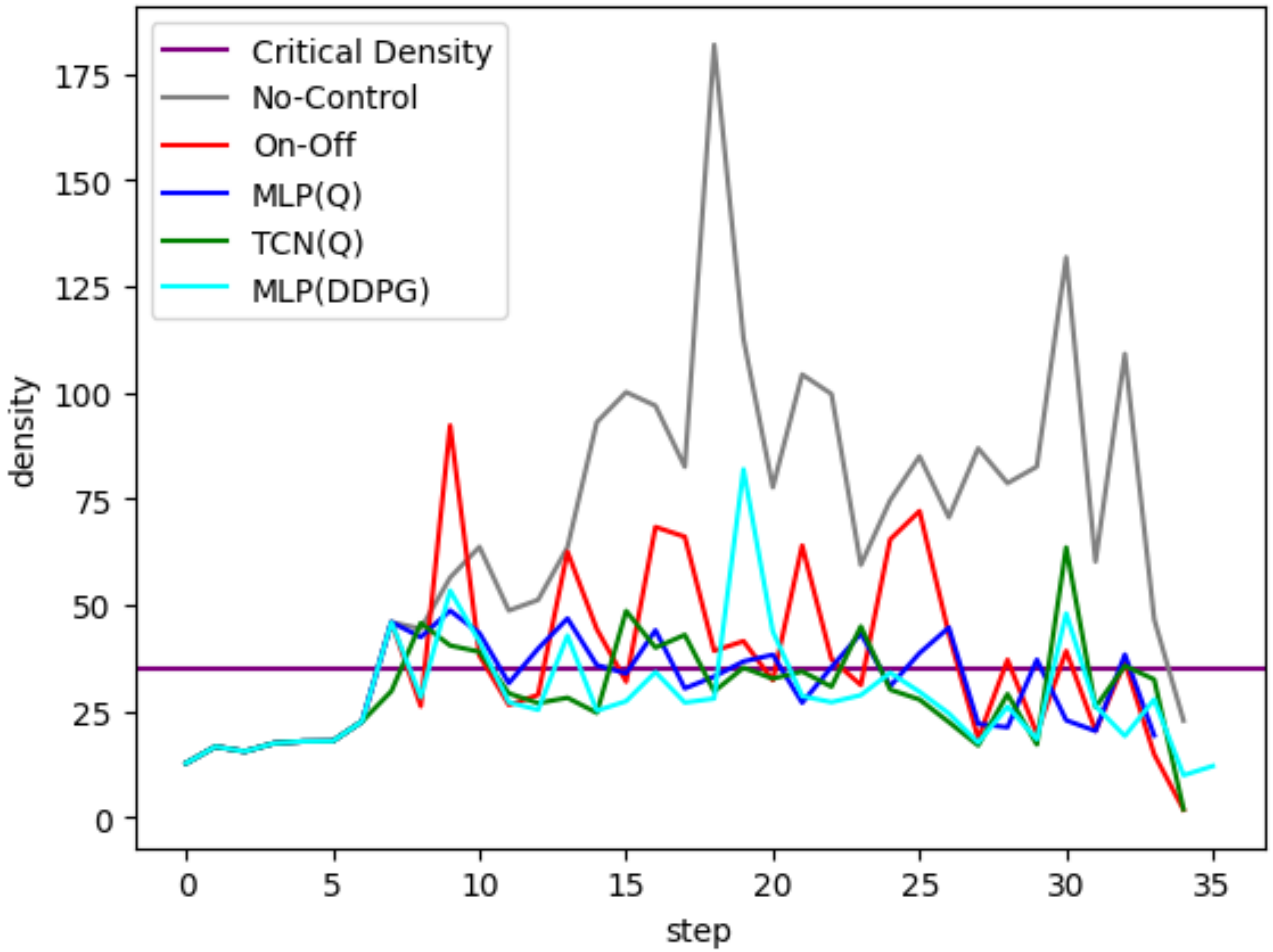
Results: densities over time

Average error: $\bar{e} = \sqrt{\dfrac{\sum_t (\rho_t - \rho_c)^2}{T}}$

- No-Control: 49.460
- On-Off: 20.218
- MLP(Q): **11.135**
- TCN(Q): 13.117
- MLP(DDPG): 15.085

Assignment

Results: densities over time



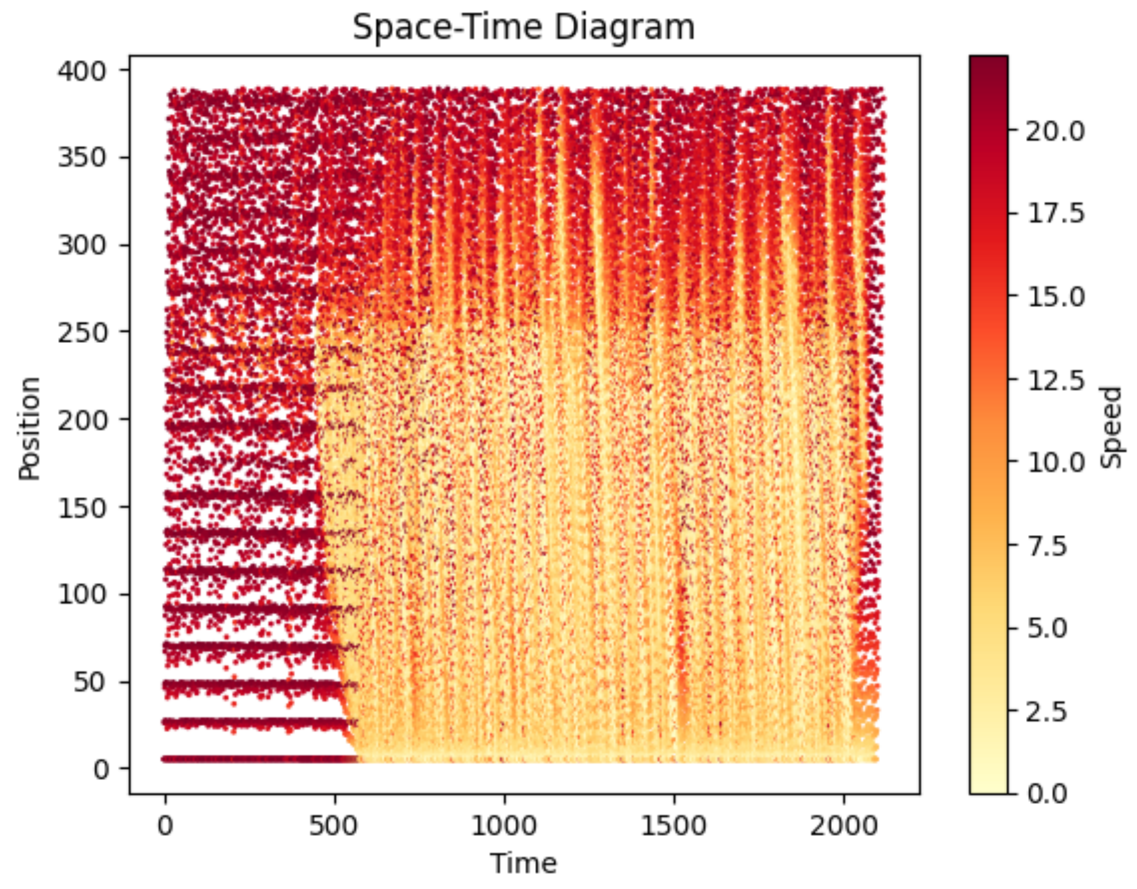
Assignment

Results: Estimated Traffic Flow & Velocity

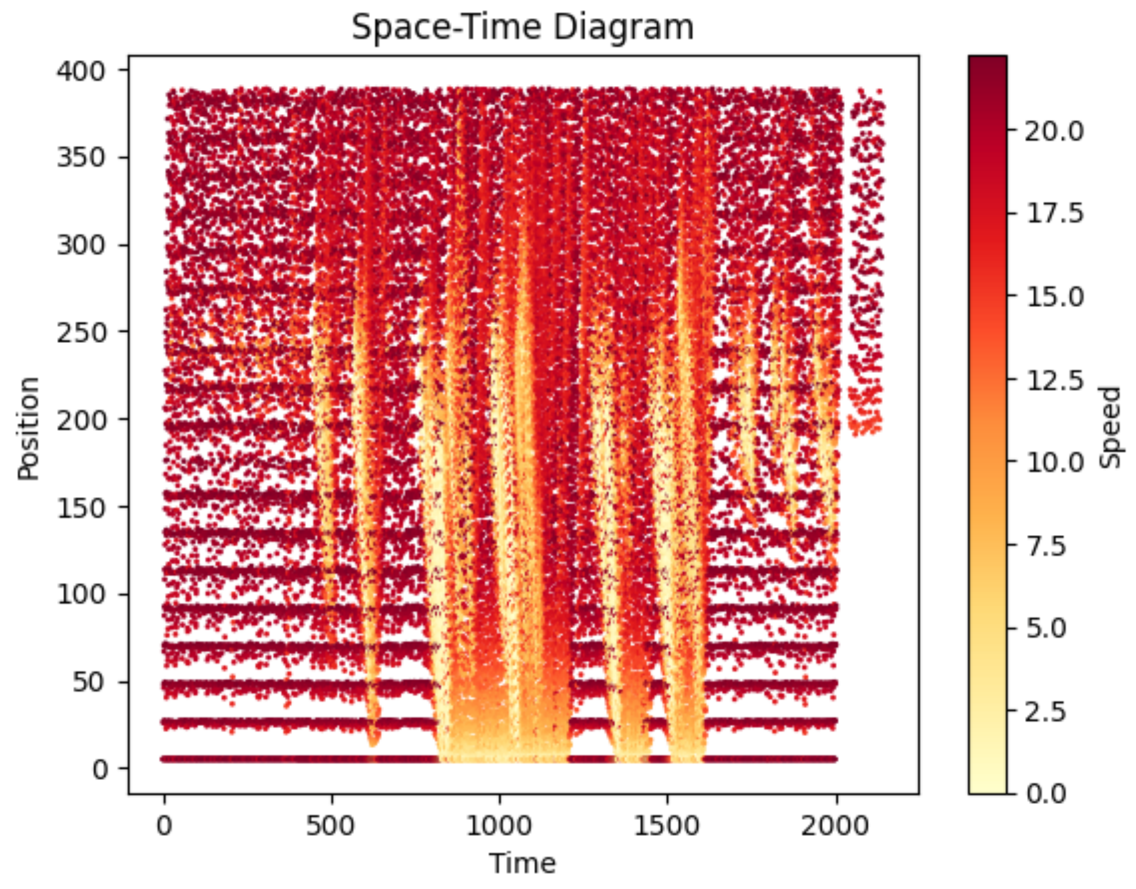
Method	Estimated Traffic Flow	Velocity
No-Control	380.272	8.822±6.767
On-Off	753.516	14.470±6.852
MLP (Q)	**943.716**	15.953±6.049
TCN (Q)	902.529	**17.688±4.766**
MLP (DDPG)	878.206	17.532±5.005

Assignment

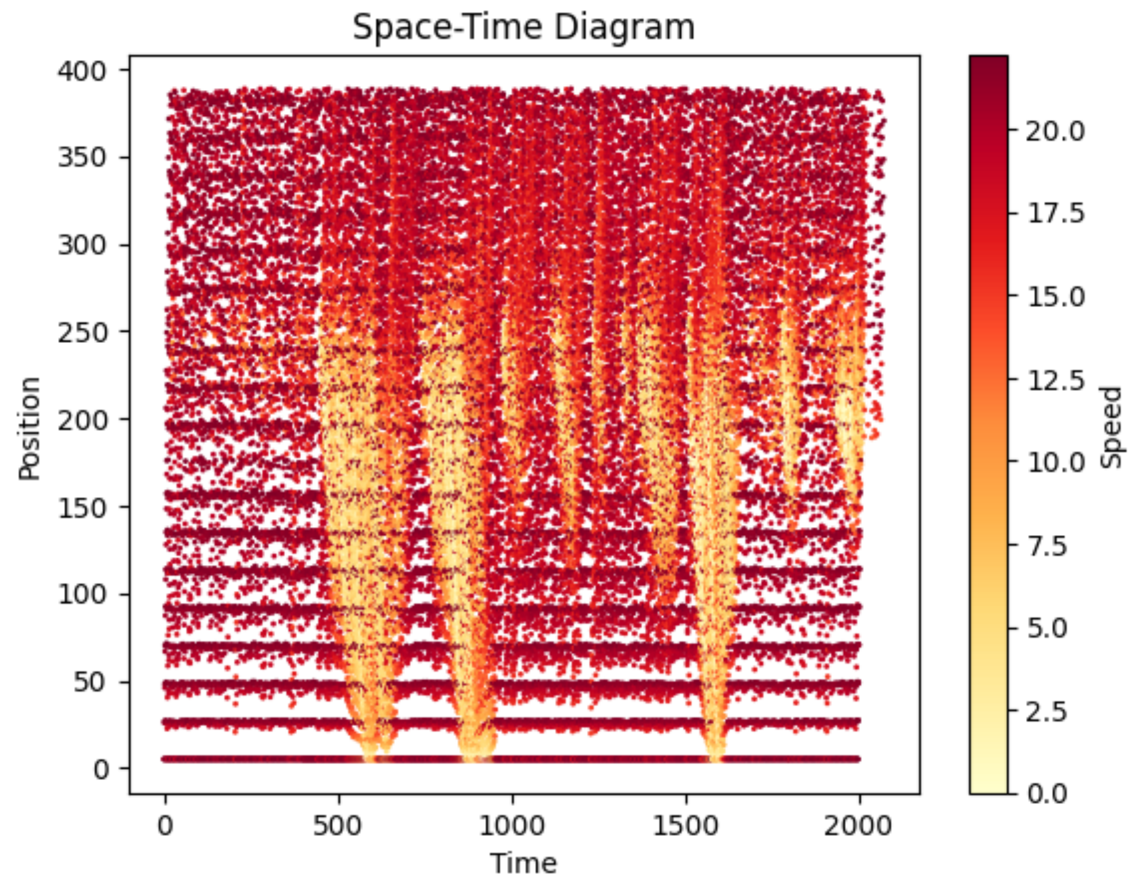
Results: Congestion



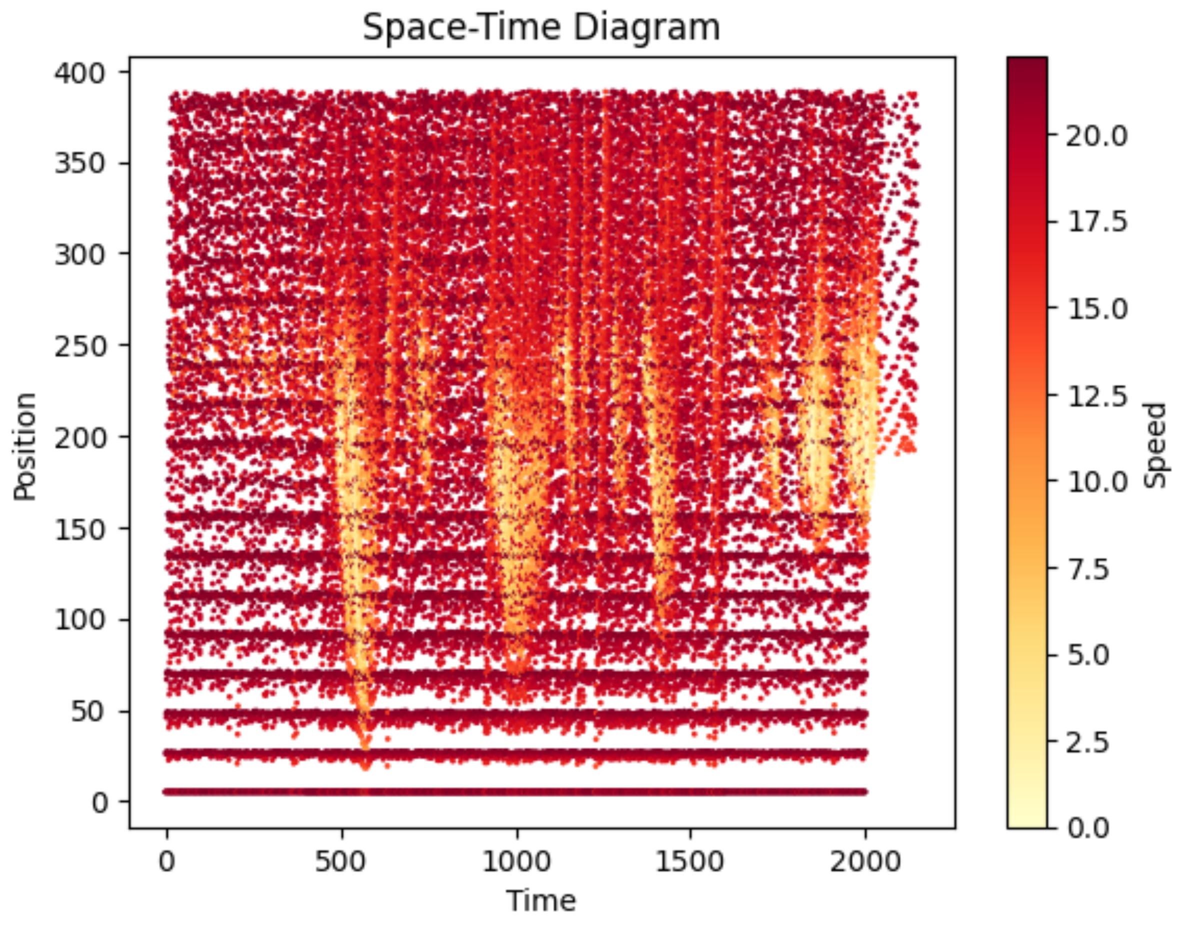
No Control



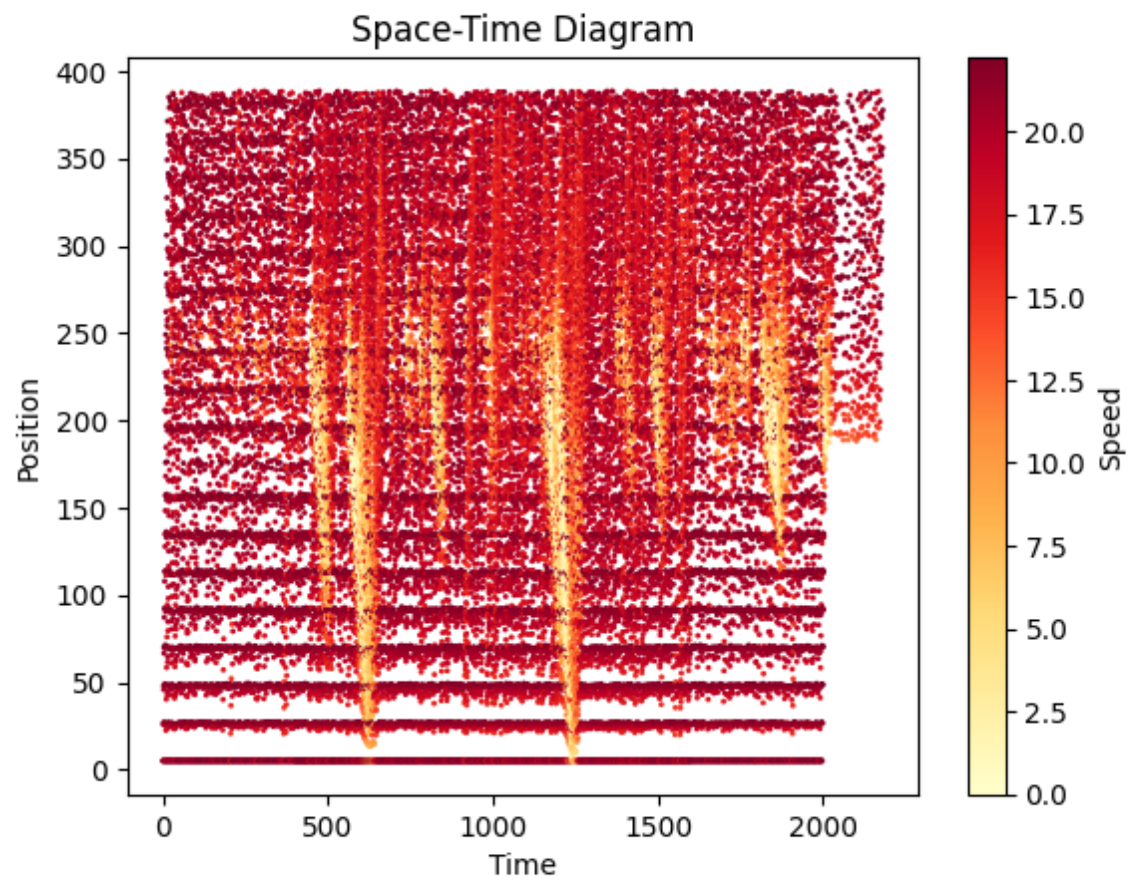
On-Off



MLP (Q)



TCN (Q)



MLP (DDPG)

Assignment

Future work

- Since the roads can be modelled as a network, it is worth studying how to combine Graph Neural Networks (GNN) with Reinforcement Learning (RL) to this problem.
- Design a new reward mechanism that can reflect the traffic flow, congestion and other measurements.

Thank you for your listening