Assessment of Reward Functions for Reinforcement Learning Traffic Signal Control under Real-World Limitations

Alvaro Cabrejas-Egea, Shaun Howell, Maksis Knutins & Colm Connaughton

WARWICK
THE UNIVERSITY OF WARWICK

centre for complexity science

EPSRC
Engineering and Physical Sciences Research Council

VIVACITY

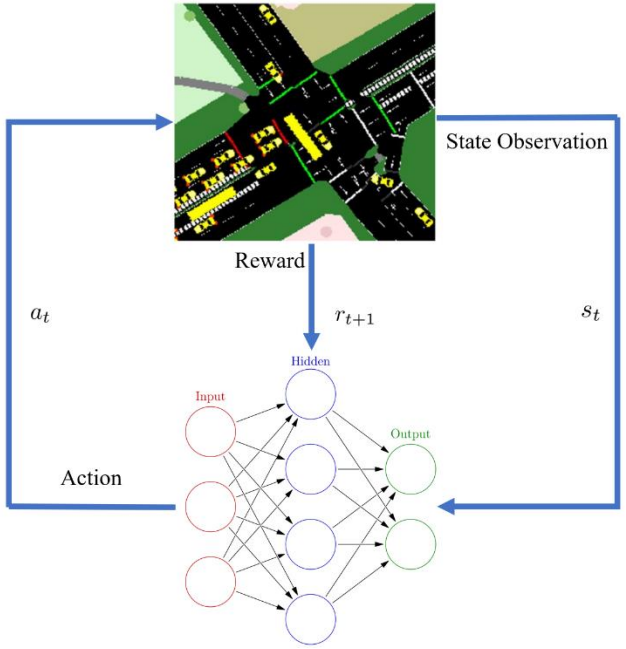
Evolution of Urban Traffic Control (UTC)

1. Fixed-Cycle
2. Adaptive
    - SCOOT
    - MOVA
3. Reinforcement Learning

Reinforcement Learning for UTC

Components:

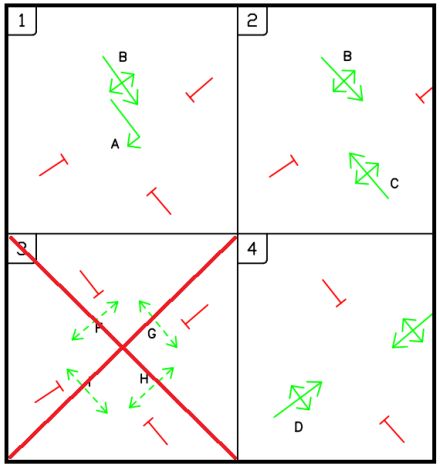
- State estimation (sensing)
- Choosing an action (decision)
- Reward Calculation (feedback)

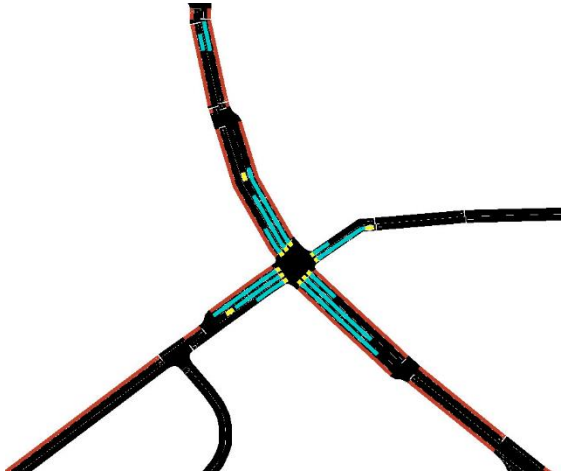


Sensing in the Real World


van
van
van
car
van
car
pedestrian
pedestrian
pedestrian
pedestrian
pedestrian
BUS
LANE

Deployment Junction





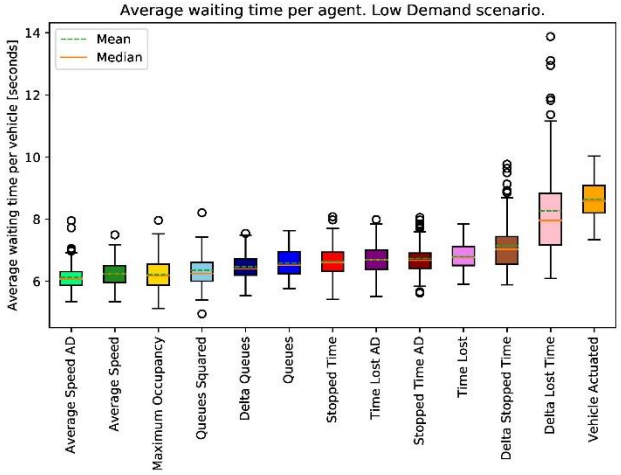
Reward Functions

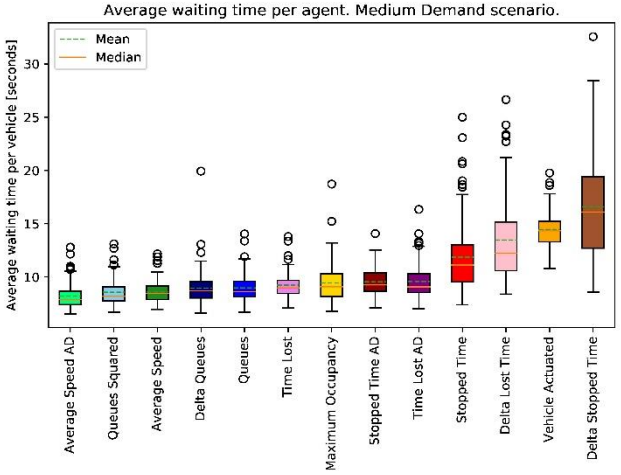
Queue Length	Time Lost (Delay)
Queue Length Squared	Δ Time Lost
Δ Queue	Time Lost Adjusted by Demand
Wait Time	Average Speed
Δ Wait Time	Avg. Speed Adjusted by Demand
Wait Time Adjusted by Demand	Throughput

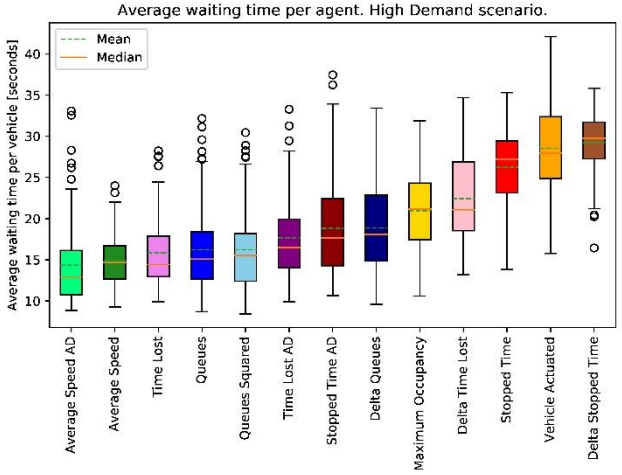
Agents, Training and Testing

- Standard DQN implementation
- State = Occupancy, 12 second buffer at δ=0.6s
- 2 Hidden layers (sizes 500, 1000), using ReLU.
- Optimized with ADAM, $\alpha = 10^{-5}$; $\gamma = 0.8$
- Trained 1500 episodes of 30 minutes (10 runs)
- Testing 100 copies of 3 demand levels

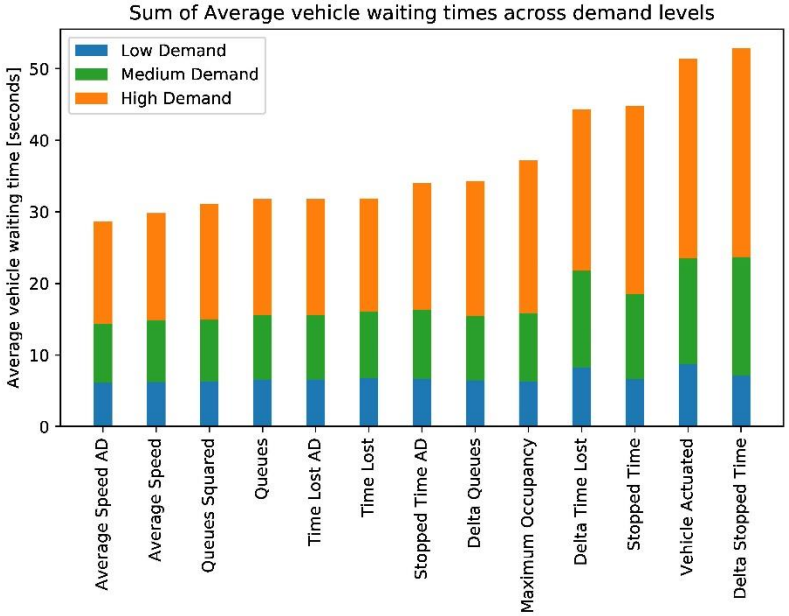
Simulation Results I







Simulation Results II

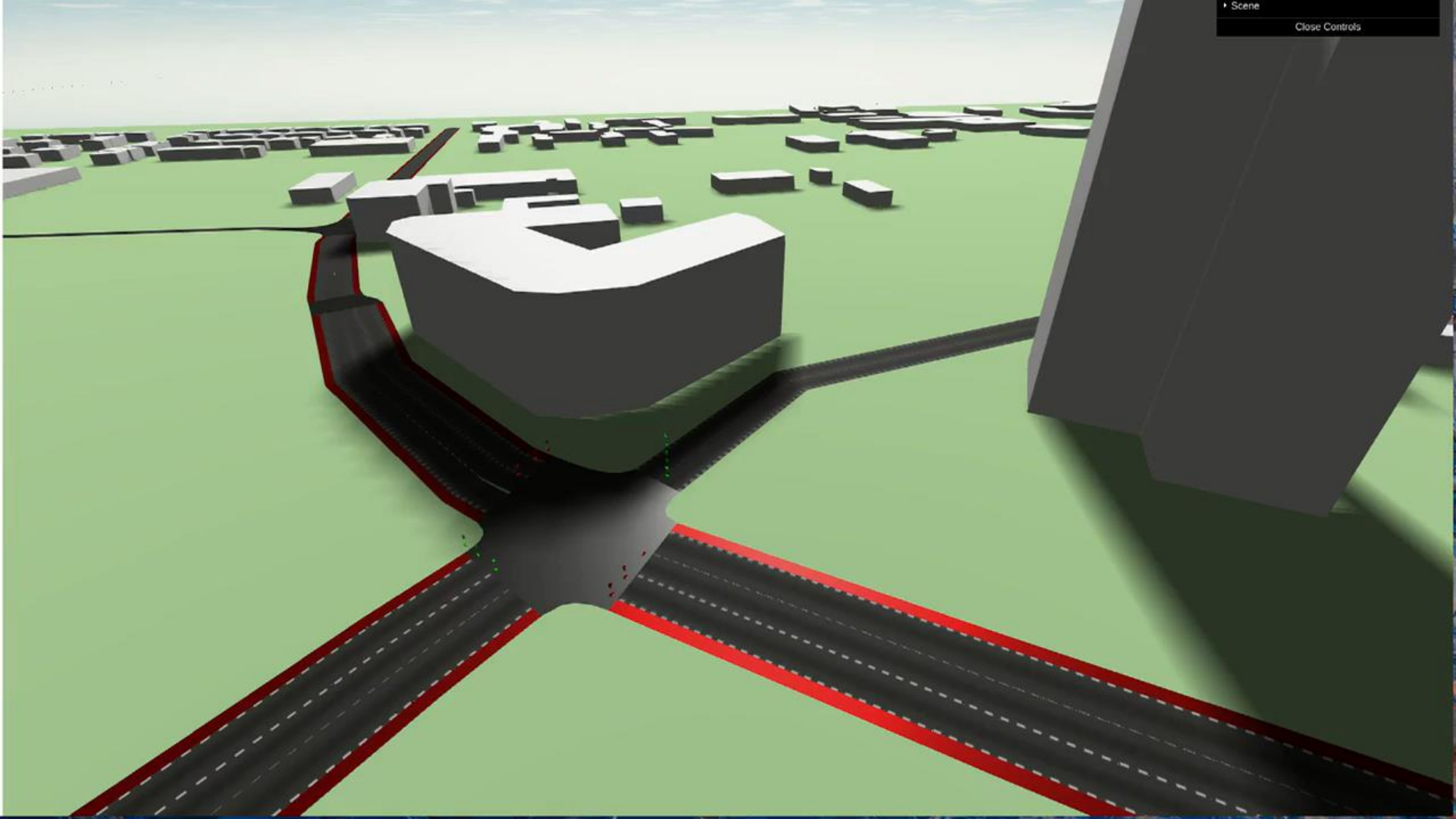


AVERAGE WAITING TIME IN SECONDS ACROSS DEMAND SCENARIOS

Reward Function	Low	Medium	High
Queues	6.59±0.46	8.97±1.27	16.21±5.07
Queues Squared	6.35±0.53	8.56±1.26	16.22±5.08
Delta Queues	6.47±0.41	8.96±1.59	18.87±5.07
Stopped Time	6.64±0.52	11.88±3.39	26.27±4.61
Stopped Time AD	6.70±0.46	9.60±1.66	17.68±4.95
Delta Stopped Time	7.15±0.81	16.59±4.95	29.21±3.67
Time Lost	6.79±0.42	9.23±1.15	15.84±4.36
Time Lost AD	6.59±0.46	8.97±1.27	16.21±5.07
Delta Time Lost	8.27±1.48	13.48±4.04	22.54±5.54
Average Speed	6.24±0.39	8.61±1.07	14.95±3.40
Average Speed AD	**6.13±0.44**	**8.22±1.24**	**14.33±4.97**
Throughput	28.02±9.36	51.16±7.23	55.72±7.02
Vehicle Actuated	8.70±0.62	14.76±1.69	27.9±6.05
Maximum Occupancy	6.32±0.51	9.51±1.87	21.33±5.77

Future Work

- Extension to pedestrians (stay tuned…)
- Multimodal optimisation
- Area controllers


Scene
Close Controls