A Deep Reinforcement Learning Approach for Dynamic Traffic Light Control with Transit Signal Priority

Tobias Nousch[1], Runhao Zhou[1], Django Adam[1], Angelika Hirrle[1], Meng Wang[1]
[1] Chair of Traffic Process Automation, Technische Universität Dresden, Germany

Abstract

Traffic light control (TLC) with transit signal priority (TSP) is an effective way to deal with urban congestion and travel delay. The growing amount of available connected vehicle data offers opportunities for signal control with transit priority, but the conventional control algorithms fall short in fully exploiting those datasets. This paper proposes a novel approach for dynamic TLC with TSP at an urban intersection. We propose a deep reinforcement learning based framework JenaRL to deal with the complex real-world intersections. The optimisation focuses on TSP while balancing the delay of all vehicles. A two-layer state space is defined to capture the real-time traffic information, i.e. vehicle position, type and incoming lane. The discrete action space includes the optimal phase and phase duration based on the real-time traffic situation. An intersection in the inner city of Jena is constructed in an open-source microscopic traffic simulator SUMO. A time-varying traffic demand of motorised individual traffic (MIT), the current TLC controller of the city, as well as the original timetables of the public transport (PT) are implemented in simulation to construct a realistic traffic environment. The results of the simulation with the proposed framework indicate a significant enhancement in the performance of traffic light controller by reducing the delay of all vehicles, and especially minimising the loss time of PT.

Keywords: double deep Q-learning, traffic light control, transit signal priority, two-layer state space, reward

1 Introduction

Travel delay and traffic congestion are common problems that disturb the economic and sustainable development of our society. There is also an urgency to reduce CO_2 emissions and fuel consumption. At the supply side, traffic light control (TLC) is one effective measure

to respond to the needs, which can minimise unnecessary stops, optimise the traffic flows, and reduce the motorised delay. At the demand side, promoting public transport (PT) is a strategic way to address urban traffic problems. From this viewpoint, it is indispensable to make the PT as attractive as possible. One approach to achieve this goal is transit signal priority (TSP) combined with a particular ameliorating traffic light control strategy. The great challenges are to minimise loss time for all road users, coordinated traffic flows and traffic lights throughout the whole road network, and to find an optimal prioritisation strategy for PT.

Today, many cities are still deploying traditional TLC or TSP strategies. Traditional TLC strategies can be categorised into three types: pre-timed, actuated and adaptive control [Koo08]. However, there are still shortcomings: 1) The traffic volume data is collected by the section-based sensors such as loop sensors and cameras. 2) These adaptive control methods are developed based on models and with assumptions about traffic dynamics. Traditional TSP has two types: Passive priority and active priority [Lin15]. Both types are currently facing two difficulties: Processing the conflict of multiple priority requests and reducing delay to motorised individual traffic (MIT).

To overcome aforementioned disadvantages of traditional control strategies, researchers have been utilising deep reinforcement learning (DRL) techniques. In the past decade, DRL based TLC or TSP has gained huge attention from both academia and industry. The neural network technology for function approximation has improved RL, enabling it to complete more challenging and complicated tasks. For the DRL based TLC, Wei et al. [Wei18] proposed a DRL model based on the Convolutional Neural Network (CNN) architecture and the value-based approach. Van der Pol et al. [Pol16] integrated transfer planning and max-plus coordination into the conventional Q-network. The CNN architecture and the value-based approach are the foundations of the DRL model. Both approaches allow one to analyse visual imagery while mapping each state-action pair to a state value and optimise the Q-values to resolve inappropriate traffic phase sequence. Liang et al. [Lia18] divided the whole intersection into small grids to quantify the complex traffic pattern as states. A CNN was designed to match the states and anticipated rewards. Guo et al. [Guo19] considered the spatial-temporal characteristics of urban traffic in DRL model. For the DRL based TSP, Long et al. [Lon21] proposed a DRL framework to solve the priority request conflict in connected vehicle environment. The action is discrete and traffic signal phases can be skipped. However, to the best of our knowledge, the existing researches do not represent the complex dynamics of urban traffic, mainly because the existing approaches are usually either trained in a fixed traffic demand or lack phase-skipping capabilities. Hence, the control strategies are trained with a non-realistic traffic environment setting, which cannot be applied to complex and realistic traffic demand, and traditional human driver environment.

In summary, some works simplify the real traffic environment. Secondly, the capability of neural network to extract features from high-dimensional data is not well-addressed in previous works. The state matrix of most previous works is two-dimensional and cannot capture all influencing factors of traffic. Finally, a significant research gap is that the limitation of ex-

isting DRL-based control strategies is that they are applied on TLC only. Hence, the research activities of signal control strategy for multi-modal transport have to be constructed.

In this paper, we redefine a DRL model inspired by [Has16] and propose a novel traffic light phase controller utilising Double Deep Q-Learning (DDQL) model. A two-layer high-dimensional state space is proposed to capture the influencing factors including the incoming lanes. And a reward function to minimise the loss times of all vehicles are formed. Meanwhile, TSP is integrated in our DDQL model. To train and validate our model, we set up a realistic traffic environment with varying traffic demand based on the road network of the city of Jena in Germany, which includes the original PT timetables. The phase controlling system presented can be readily superimposed to an existing local traffic light control, and minimise implementation costs.

2 DDQL-based Model

Because Double Deep Q-Learning is not plagued by the overestimation bias and has good performance on handling high-dimensional data, we redesign the neural network architecture of DDQL according to the complex traffic environment. In DDQL, during training, there are two Q-networks and three important elements S, A and R, where S is the state space, A is the action space, and R is the reward function. The system architecture is shown in Figure 1.

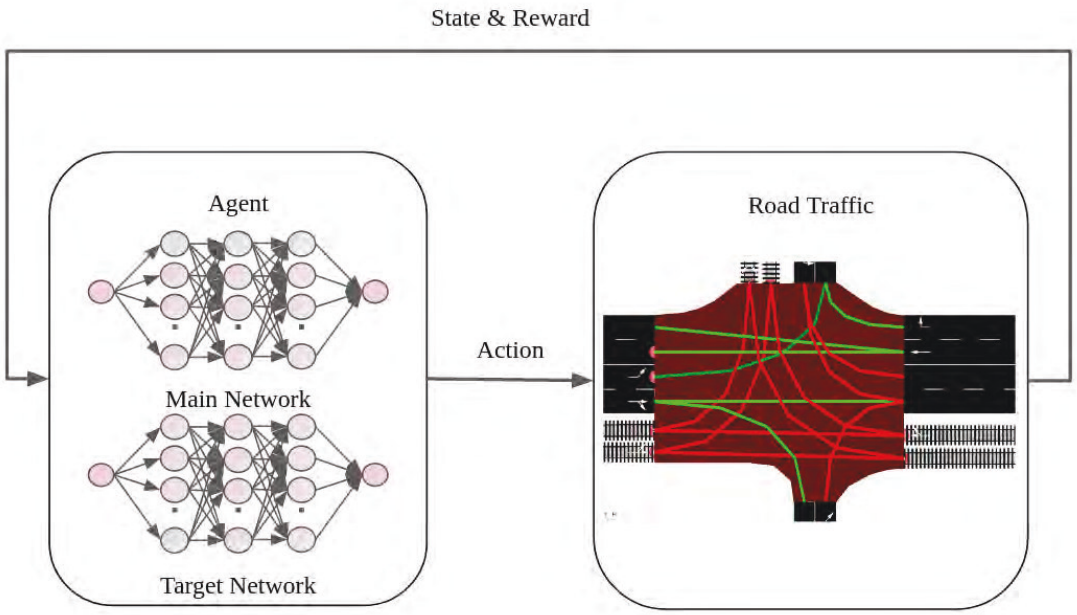


Figure 1: Double Deep Q-Learning Cycle. The agent receives the state space and reward, and performs actions in the road traffic environment.

2.1 Agent Design

The Neural Network Architecture of Double Deep Q-Learning

Our model is equipped with experience replay buffer and implemented with the python framework TensorFlow. The architecture of main network is shown in Figure 2. The target network has the same settings as the evaluation network and obtains update every 200 training steps. The replay buffer saves 10000 simulation steps and overwrites the memory from the beginning if exceeding the buffer size.

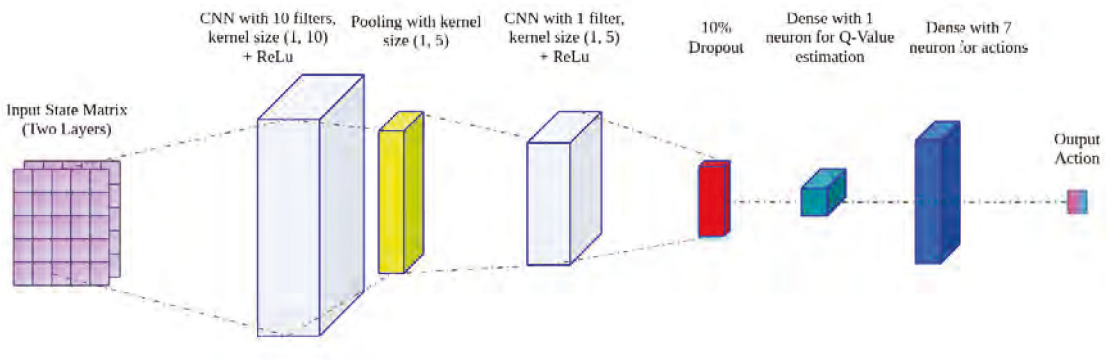


Figure 2: Main Q-Network architecture.

State Space

We propose a two-layer state space which is derived as follows: Vehicles are considered to be entering an intersection $500\,\mathrm{m}$ in front of the traffic light on one of the incoming lanes k of the intersection. We partition each lane into segments of length $l = 1\,\mathrm{m}$. Hence, the whole intersection is divided into small grids of equal size. We are now able to form a matrix $P_t \in \mathbb{R}^{500\times k}$, which is called position layer. Each element of the matrix is in $\{0, 1\}$, where it is non zero if there is a vehicle in the corresponding grid. In the same way, we construct a matrix T_t, called type layer. It is of same size as P_t but with entries, representing the type of vehicle, which is zero for MIT and one for PT. An example of the first layer of the state matrix is shown in Figure 3.

If a transition phase or the minimum green time has to be executed, the agent due to the legal constraints can only make a decision after it has been completed. It ensures that all design constraints are fulfilled in every time step according to the received local legal guideline.

Action Space

The action space contains all possible actions and corresponding duration for a given state. In our case study, the agent decides in an interval of one second whether to keep the cur-

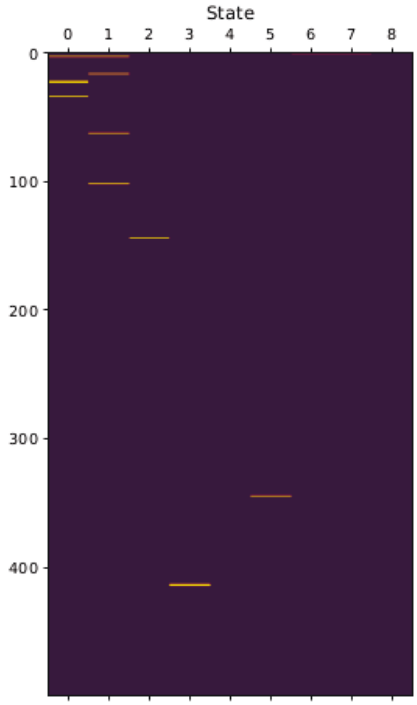


Figure 3: First layer of the state space, where each line represents the position of a vehicle.

rent signal phase or to switch to one of the other phases which controls the current traffic efficiently.

Reward Function

One of the great challenges in DRL is to setup a reward function that represents all the desired properties that the agent can learn and act upon. Our reward function r_t for the selected action in the time step t is formulated as follow.

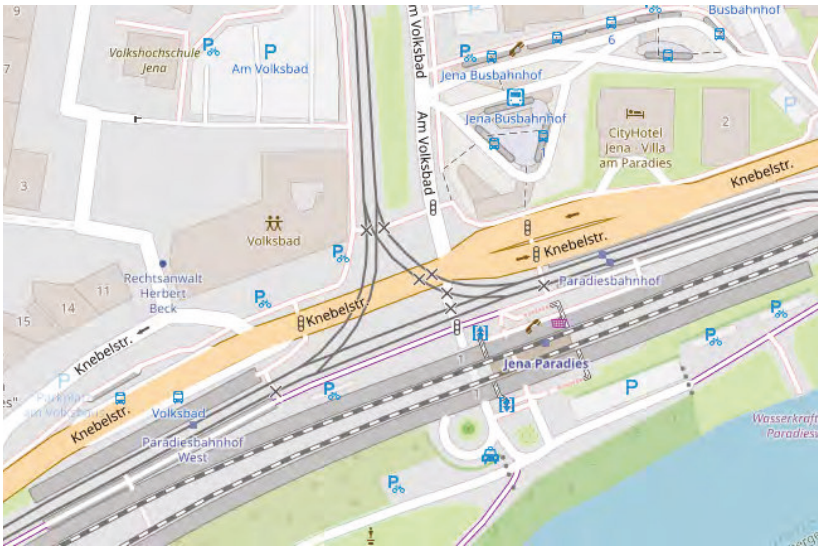
$$r_t = -\sum_{T}\sum_{v_T}\left[\eta_T\left(\frac{\tau_{v_T}}{C_T}\right)^{\rho_T}\right] - \vartheta\sum_{l} q_l. \tag{1}$$

In the first term of Equation (1), T indicates the type of a vehicle, v_T denotes the number of every vehicle type, the waiting time is τ_{v_T}, C_T represents the conventional waiting time according to the empirical experience, η_T and ρ_T are defined as specific computational parameters. This term indicates the waiting time of every vehicle in front of the intersection. In the second term of Equation (1), l indicates the index of every incoming lane, ϑ is a normalisation parameter, and q_l denotes the queue length of every lane. The second term denotes the total queue length of all lanes.

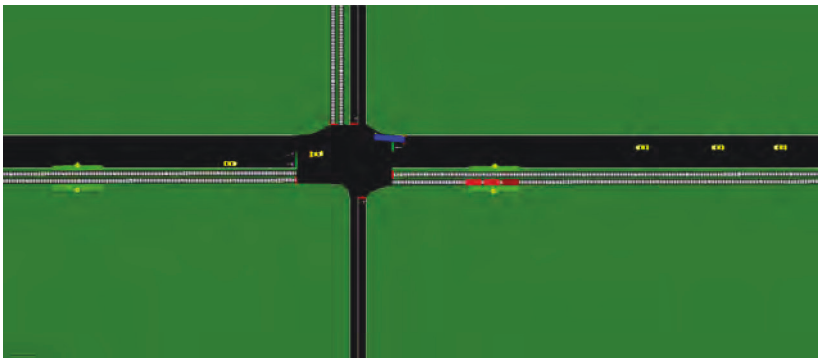
2.2 Simulation Setup

To implement a complex training environment and sufficient realistic network layout, we set up a simulation based on a real intersection of the city of Jena in Germany in the microscopic

software Simulation of Urban MObility (SUMO) (see Figure 4). The peculiarity of this intersection is the multi-modal traffic involved with trams in the extra track require prioritisation. In addition, the nearby tram stops “Paradiesbahnhof” and “Paradiesbahnhof West” generate a special challenge for the DDQL model to coordinate TLC and priority of public transit.



(a) OSM map of the intersection Knebelstr./Volksbad in Jena, Germany.



(b) SUMO simulation of a intersection Knebelstr./Volksbad in Jena, Germany.

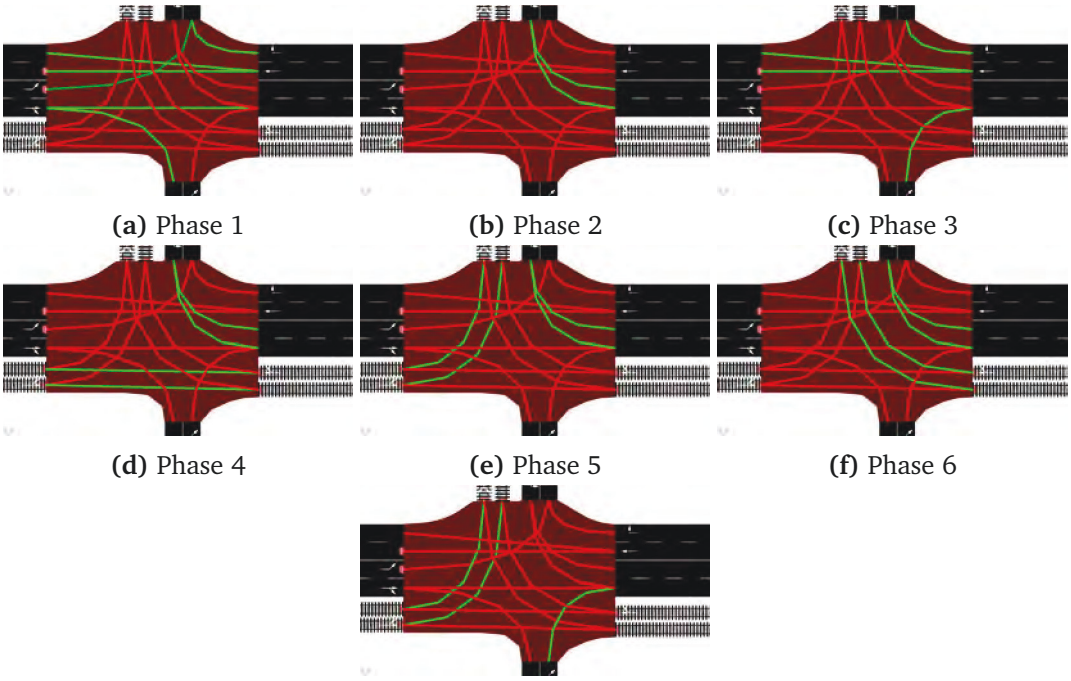
Figure 4: Representation of the implemented intersection in real-world map and SUMO.

The varying traffic demand at an urban intersection is fitted by a sine function, with the traffic demand on morning and evening being the high peaks and at midday and night being the low peaks. We therefore simplify dynamic traffic demand function to Equation (2).

$$Demand = BaseFlow \times \left(1 + sin^2\left(\frac{t}{AddFrequency_T}\right)\right). \tag{2}$$

In Equation (2) index T indicates the type of vehicle, $BaseFlow$ indicates the basic traffic flow at an intersection and $AddFrequency_T$ is the frequency of adding new vehicles in road network.

To simulate the traffic light control, we use the original signal phases and their transitions which are relevant for the observed traffic and provided by the city administration of Jena. Thus, we implemented seven different signal phases (see Figure 5) and accordingly 6×7 transitions. The main idea is that our phase controlling system could be superimposed to an existing local traffic light control, without any reconfiguration and evaluating the whole signalised intersection. Hence, we hard coded legal requirements, such as the minimum green time, clearing times and maximum blocking time, firmly in the source code to form the preset phase duration and technical evaluations keep valid, and the presented model is dynamically selecting the phase duration and next phase.



(a) Phase 1 **(b)** Phase 2 **(c)** Phase 3

(d) Phase 4 **(e)** Phase 5 **(f)** Phase 6

(g) Phase 7

Figure 5: Traffic light phases.

3 Results and Discussion

Compared to previous researches based on simplified traffic scenarios, we present the performance of our DDQL model in a more realistic and complex traffic environment and preliminary results. The agent of our DDQL model is able to learn to control a complex signalised intersection and reduce the waiting time of MIT as well as PT. The represented results are achieved with the reward parameter shown in Table 1.

In Figure 6 we compare, as preliminary results, the developed DDQL model to the original traffic control on two parallel running simulations of the same intersection with the same characteristics and traffic flows. Different vehicles are generated randomly. In the first subfigure, the total accumulated reward is depicted over an simulation of one hour. The

Table 1: Reward parameters.

	Car	**Bus**	**Tram**
η	0.001	0.01	0.03
ρ	2	2	2
C	60	10	5
ϑ	0.01	0.01	0.01

second subfigure shows the reward calculated by Equation (1) for each simulation step. The third and fourth subfigures show the waiting time for the PT and MIT. And the fifth subfigure illustrates the total amount of vehicles on all incoming lanes of the controlled intersection. Preliminary results show that for MIT and PT the waiting time is reduced, and the agent controlled intersection gains in total a higher system scores. The results we have so far are promising and we hope to extend the model for the validation in a more complex environment.

The proposed model adapts flexibly to the local traffic flows and could generate non-cyclic switching patterns respectively phase-skipping. This behaviour is required particularly in situations with low traffic volumes and demonstrates the advantage of our model over other controls.

4 Conclusions

In this paper, we propose to solve the traffic light control problem using a deep reinforcement learning model. The traffic information is collected from the road network, and the original traffic light information is provided by the city administration of Jena. The state space includes two layers of vehicle position and type, and each layer is two-dimension values that consists the index of every incoming lane and the length of lane with partition in 500 grids that each grid denotes 1 m. The actions are modeled as a Markov decision process, and the reward function is the negative cumulative waiting time and total queue length of all lanes with a normalisation parameter. To handle the complex traffic scenario in our problem, we propose a Double Deep Q-Learning (DDQL) model with novel neural network architecture and experience replay buffer. The proposed model can learn a good policy under varying traffic demand, outperform the existing traffic light control system of the city of Jena in waiting time, which is shown in an extensive simulation in SUMO and TensorFlow. In the next step of our work we will optimise the phase control even further including other types of road users such as pedestrians and cyclists.

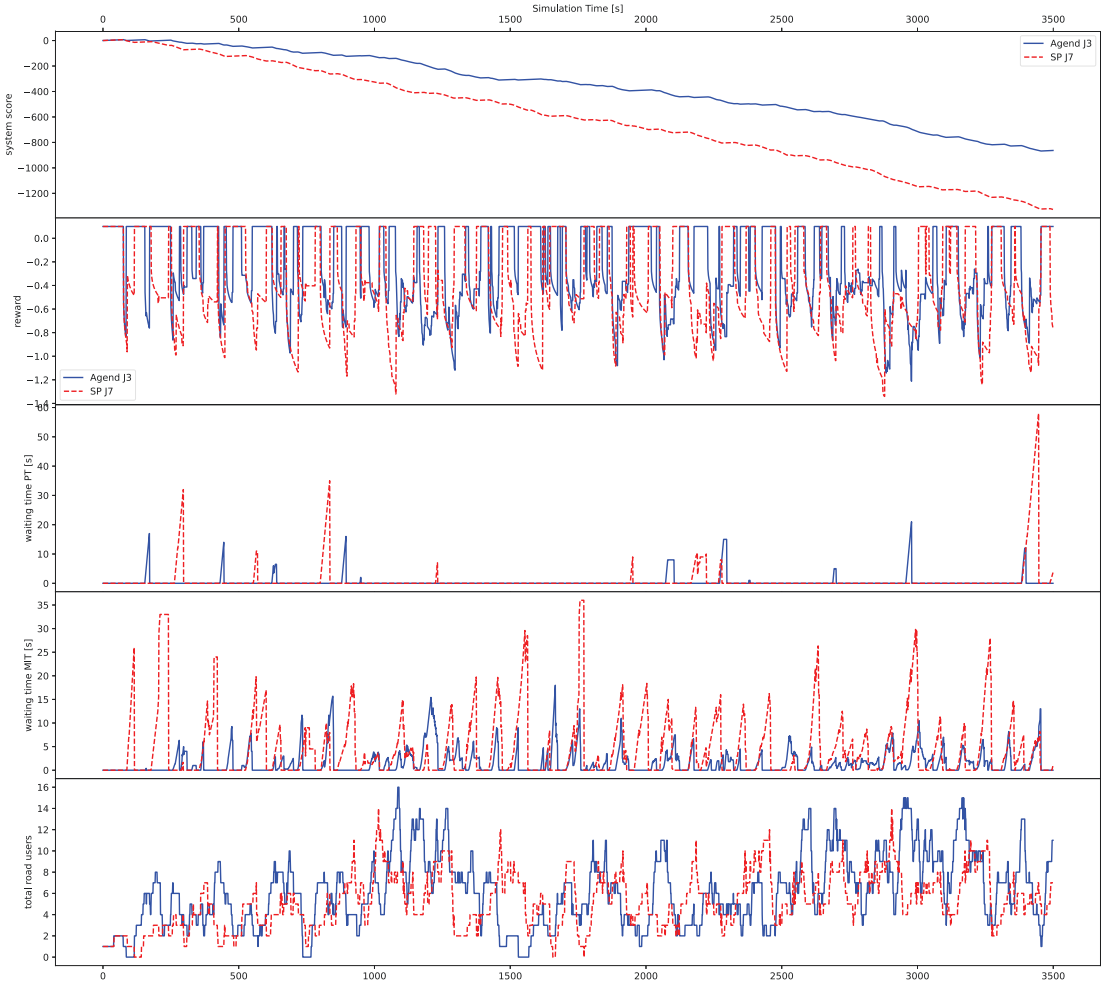


Figure 6: Comparison of the original traffic control of the city Jena (red dashed line) and traffic control via DDQL (blue line).

Acknowledgements

This research is supported by the project “5G-basierte V2X-Vernetzung zur Steigerung der Verkehrssicherheit sowie zur Optimierung des multimodalen Verkehrs und der Energieversorgung in Jena” funded by the German Federal Ministry for Digital and Transport.

References

[Guo19] M. Guo, P. Wang, C. Chan, and S. Askary: “A Reinforcement Learning Approach for Intelligent Traffic Signal Control at Urban Intersections”. In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Institute of Electrical and Electronics Engineers, 2019, pages 4242–4247.

[Has16] H. van Hasselt, A. Guez, and D. Silver: “Deep Reinforcement Learning with Double Q-learning”. In: *Proceedings of the Thirtieth AAAI Conference on Artificial*

Intelligence. Volume 30. 1. Mar. 2016, pages 2094–2100. DOI: 10.1609/aaai.v30i1.10295.

[Koo08] P. KOONCE, L. A. RODEGERDTS, K. LEE, S. QUAYLE, S. BEAIRD, C. BRAUD, J. A. BONNESON, P. J. TARNOFF, and T. URBANIK: *Traffic signal timing manual*. Technical report FHWA-HOP-08-024. U.S. Department of Transportation, Federal Highway Administration, June 2008. URL: https://rosap.ntl.bts.gov/view/dot/800.

[Lia18] X. LIANG, X. DU, G. WANG, and Z. HAN: *Deep Reinforcement Learning for Traffic Light Control in Vehicular Networks*. 2018. DOI: 10.48550/arXiv.1803.11115.

[Lin15] Y. LIN, X. YANG, N. ZOU, and M. FRANZ: "Transit signal priority control at signalized intersections: a comprehensive review". In: *Transportation Letters* 7.3 (2015), pages 168–180. ISSN: 1942-7875. DOI: 10.1179/1942787514Y.0000000044.

[Lon21] M. LONG, X. ZOU, Y. ZHOU, and E. CHUNG: *Deep Reinforcement Learning for Transit Signal Priority in a Connected Environment*. 2021. DOI: 10.2139/ssrn.3992999.

[Pol16] E. van der POL and F. A. OLIEHOEK: "Coordinated Deep Reinforcement Learners for Traffic Light Control". In: *NIPS'16 Workshop on Learning, Inference and Control of Multi-Agent Systems*. Barcelona, Spain, Dec. 2016. URL: https://www.fransoliehoek.net/docs/VanDerPol16LICMAS.pdf.

[Wei18] H. WEI, G. ZHENG, H. YAO, and Z. J. LI: "IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control". In: London, United Kingdom, 2018, pages 2496–2505. DOI: 10.1145/3219819.3220096.

Corresponding author: Tobias Nousch, Chair of Traffic Process Automation, Technische Universität Dresden, Germany, e-mail: tobias.nousch@tu-dresden.de