PROCEEDINGS OF ELMAR-2020

62nd International Symposium ELMAR-2020

14-15 September 2020, Zadar, Croatia

EDITED BY

Mario Muštra
Josip Vuković
Branka Zovko-Cihlar

University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Wireless Communications
Unska 3 / XII
10000 Zagreb
CROATIA

Published by:	Faculty of Electrical Engineering and Computing, University of Zagreb
Edited by:	Mario Muštra, Josip Vuković, Branka Zovko-Cihlar
Front Cover:	Painting by artist Mrs Ljerka Njerš
Printed by:	Ispis Ltd., Zagreb
Print ISBN:	978-1-7281-5972-0, CFP20825-PRT
XPLORE ISBN:	978-1-7281-5973-7, CFP20825-ART

Printed in Croatia.

Elmar

62nd International Symposium ELMAR-2020 is organised by:

Croatian Society Electronics in Marine - ELMAR, Zadar

University of Zagreb Faculty of Electrical Engineering and Computing, Department of Wireless Communications

ELMAR-2020 symposium Sponsors

Ministry of the Sea, Transport and Infrastructure of the Republic of Croatia

Ministry of Science and Education of the Republic of Croatia

Croatian Regulatory Authority for Network Industries

HAKOM

Town of Zadar

County of Zadar

ELMAR-2020 symposium takes place under the co-sponsorship of:

IEEE Region 8
IEEE Croatia Section
IEEE Croatia Section Chapter of the Signal Processing Society
IEEE Croatia Section Chapter of the Antennas and Propagation Society
IEEE Croatia Section Chapter of the Microwave Theory and Techniques Society
Croatian Academy of Sciences and Arts
Croatian Academy of Engineering - HATZ
University of Zagreb
University of Zagreb Faculty of Electrical Engineering and Computing
University of Zagreb, Faculty of Transport and Traffic Sciences
University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology

ELMAR-2020 symposium is supported by:

ZADAR TOURIST BOARD
ERICSSON NIKOLA TESLA
OTP BANKA HRVATSKA
FALKENSTEINER BORIK ZADAR CROATIA

ELMAR-2020 SYMPOSIUM COMMITTEES

General Chair

Branka Zovko-Cihlar, University of Zagreb, Croatia

General Co-Chair

Mislav Grgić, University of Zagreb, Croatia

Program Chair

Mario Muštra, University of Zagreb, Croatia

ELMAR-2020 SYMPOSIUM INTERNATIONAL PROGRAM COMMITTEE

Juraj Bartolić, Croatia
Narcis Behlilović, Bosnia and Herzegovina
Aura Conci, Brasil
Marek Domanski, Poland
Janusz Filipiak, Poland
Borko Furht, USA
Mohammed Ghanbari, United Kingdom
Dušan Gleich, Slovenia
Mislav Grgić, Croatia
Sonja Grgić, Croatia
Ho Yo-Sung, Korea
Bernhard Hofmann-Wellenhof, Austria
Ebroul Izquierdo, United Kingdom
Ana Katalinić Mucalo, Croatia
Ismail Khalil, Austria
Tomislav Kos, Croatia
Igor Kuzle, Croatia
Panos Liatsis, United Kingdom
Lidija Mandić, Croatia
Branka Medved Rogina, Croatia
Borivoj Modlic, Croatia
Marta Mrak, United Kingdom
Mario Muštra, Croatia
Zdeněk Němec, Slovakia
Miloš Oravec, Slovakia
Jarmila Pavlovičova, Slovakia
Fernando Pereira, Portugal
Jan Pidanič, Slovakia
Peter Planinšič, Slovenia
Pavol Podhradský, Slovakia
Darko Ratkaj, Switzerland
Fabiana Rodrigues Leta, Brasil
Gregor Rozinaj, Slovakia
Markus Rupp, Austria
Shiguang Shan, China
Thomas Sikora, Germany
Karolj Skala, Croatia
Ryszard Stasinski, Poland
Luis Torres, Spain
Frantisek Vejrazka, Czech Republic
Dijana Vitas, Croatia
Stamatis Voliotis, Greece
Krzysztof Wajda, Poland
Branka Zovko-Cihlar, Croatia

ELMAR-2020 SYMPOSIUM
INTERNATIONAL REVIEW COMMITTEE

Goran Bakalar, Croatia
Sanja Bauk, Montenegro
Alen Begović, Bosnia and Herzegovina
Marko Bosiljevac, Croatia
Jelena Božek, Croatia
Emil Dumić, Croatia
Hrvoje Gold, Croatia
Sonja Grgić, Croatia
Edouard Ivanjko, Croatia
Branimir Ivšić, Croatia
Niko Jelušić, Croatia
Juraj Kačur, Slovakia
Ana Katalinić-Mucalo, Croatia
Serđo Kos, Croatia
Tomislav Kos, Croatia
Hrvoje Leventić, Croatia
Časlav Livada, Croatia
Sadko Mandžuka, Croatia
Marko Matulin, Croatia
Štefica Mrvelj, Croatia
Mario Muštra, Croatia
Miloš Oravec, Slovakia
Jarmila Pavlovičová, Slovakia
Juraj Petrović, Croatia
Jan Pidanič, Slovakia
Snježana Rimac-Drlje, Croatia
Renata Rybárová, Slovakia
Dorsaf Sebai, Tunisia
Chakib Taybi, Morocco
Filip Turčinović, Croatia
Mario Vranješ, Croatia
Josip Vuković, Croatia

ELMAR-2020 SYMPOSIUM
ORGANISING COMMITTEE

Mislav Grgić, Croatia
Mario Muštra, Croatia
Jelena Božek, Croatia
Filip Turčinović, Croatia
Josip Vuković, Croatia

CROATIAN SOCIETY ELECTRONICS IN MARINE - ELMAR

Address: Faculty of Electrical Engineering and Computing
Croatian Society Electronics in Marine - ELMAR
Unska 3 / XII, 10000 Zagreb, CROATIA

Phone: + 385 1 6129 567

Fax: + 385 1 6129 717

E-mail: elmar2020@fer.hr

URL: www.elmar-zadar.org

Welcome speech of the President of Croatian Society Electronics in Marine - ELMAR, Zadar

It is my great pleasure to welcome you to the 62nd International Symposium ELMAR-2020 and wish you a successful work and communication during your stay in Zadar!

Even though we completed only a bit less than one half of the 2020, it can be said that it was one of the strangest in the recent history. Covid-19 disease presented a problem for almost all businesses and activities our civilization relies on. It is not different with the ELMAR, but we are very proud that 62nd ELMAR takes place even in these circumstances.

ELMAR-2020 is organized under the technical co-sponsorship of IEEE Region 8, IEEE Croatia Section, Croatian Academy of Sciences and Arts, Croatian Academy of Engineering, University of Zagreb Faculty of Electrical Engineering and Computing Department of Wireless Communications, University of Zagreb, Faculty of Transport and Traffic Sciences, and University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology.

This year the International Review Committee selected 32 papers written by authors from 9 countries. It is our great pleasure to thank all the authors on their submitted papers and the effort they made while preparing their contributions. We are sincerely grateful to reviewers for their kind help and vast contribution in selection of the papers which showed significant scientific contribution and because of it are accepted for presentation at ELMAR-2020 and publication in IEEE Xplore.

Due to these, a bit different conditions, the program of this year's ELMAR Symposium is divided into 7 regular sessions and one special session. The International Program Committee of ELMAR traditionally invites outstanding experts from fields related to the ELMAR international symposium to present a state-of-the-art in their research fields. As keynote speakers Prof. Dubravko Babić from the Faculty of Electrical Engineering and Computing will give a talk about FERSAT: Characterizing Light Pollution Using a Cubesat; Prof. Edouard Ivanjko from the Faculty of Transport and Traffic Sciences will give a talk about Development of Adaptive Traffic Signal Control Systems for Urban Environments.

I am specially obliged to thank Ministry of the Sea, Transport and Infrastructure of the Republic of Croatia, Ministry of Science and Education of the Republic of Croatia, Croatian Regulatory Authority for Network Industries, Town of Zadar, and County of Zadar, without whom would be impossible to organize this scientific event.

At last, I would like to thank all the members of the Organizing Committee for their efforts in preparing this symposium in these, different than regular, conditions.

Branka Zovko Cihlar

Professor *Branka Zovko-Cihlar*
President of the Croatian Society
Electronics in Marine - ELMAR, Zadar

Contents

ANTENNAS AND PROPAGATION

ANTENNAS AND PROPAGATION 2

MULTIMEDIA PROCESSING

MULTIMEDIA PROCESSING 2

SPECIAL SESSION: INTELLIGENT TRANSPORT SYSTEMS

WIRELESS COMMUNICATIONS AND AUTOMATION

Variable Speed Limit Control Based on Deep Reinforcement Learning: A Possible Implementation

Martin Gregurić, Krešimir Kušić, Filip Vrbanić, Edouard Ivanjko
Faculty of Transport and Traffic Sciences, University of Zagreb
Vukelićeva Street 4, HR-10000 Zagreb, Republic of Croatia
martin.greguric@fpz.unizg.hr

***Abstract*—Today's urban motorways cannot fulfill their purpose to simultaneously serve transit and local urban traffic with a high Level of Service. In the case of urban motorway infrastructure, the traditional "build only" approach is not always possible due to the lack of space. This study is focused on the Variable Speed Limit Control (VSLC) as one of the traffic control methods applicable for any type of motorway and Q-learning as one commonly used approach for designing learning based VSLC algorithms. The drawback of this methodology is the representation and exploration of the large state-action space as it is the case in its application for VSLC. This study introduces a Deep Q-Network to approximate the Q-function and presents a novel learning approach for the VSLC application with possibility to track vehicles on the microscopic level. The proposed reward function steers the learning towards the improvement of reward and prevention of oscillation among consecutive speed limits.**

***Keywords*—Traffic Control; Variable Speed Limit Control; Intelligent Transportation Systems; Learning Systems; Deep Q-Learning Network; Criteria Functions; Performance Analysis**

I. Introduction

Urban motorways are considered as an integral part of the urban traffic network. They evolved from urban bypasses. Thus, they serve transit traffic, but also the local traffic originating from the urban environment. The continuous increase of traffic demand for the use of the urban motorways leads to occasional capacity drops causing periodic congestion. Such periodic congestion often occurs near an on-ramp and is caused by local traffic entering the motorway. On-ramp flow is characterized by lower average speed compared to the speed of the mainstream flow at the motorway. Therefore, the increased inflow at an on-ramp may slow down the mainstream traffic flow due to realignment of the vehicles from an on-ramp in the mainstream flow, as well due to the realignment of vehicles already on the motorway to the left lanes on the motorway trying to avoid slower traffic in the far-right lane. This effect can produce disruption in the mainstream flow and leads to congestion creating bottlenecks, therefore, decreasing stability and safety [1].

Variable Speed Limit Control (VSLC) is a traffic control approach for motorways that computes speed limits based on the current traffic condition and posts them on a Variable Message Sign (VMS). It can reduce the propagation effect of disturbances (shock-waves) in speeds of the traffic flow by smoothing the transition of speed (harmonization) between congested downstream traffic and upstream free-flow traffic as it is described in [2]. The approach of Reinforcement Learning (RL) can provide efficient model-free control and computation of optimal policies in terms of delay minimization for the various traffic control tasks, including the VSLC, as it is shown in [3]. The goal of RL is to select the best-known action for any given state according to the previously computed ranking. Various approaches are possible in the spatial and temporal discretization of traffic states, and its adjacent actions that are based on macroscopic traffic parameters. All those traffic state assessment parameters in combination with several possible actions (e.g. all possible speed limits) can lead to the "curse of dimensionality". It is possible to apply various function approximations methods such as, for example, Neural Networks (NN) to reduce the dimensionality of the state-action space. To handle nonlinear traffic flow behavior, Deep Q-Learning (DQL) algorithm, and its various modifications have recently started to be used in traffic control as described in [4], and [5]. One of the most prominent obstacles towards efficient DQL implementation in the traffic control problems is appropriate reward function design [6]. It steers learning convergence towards the desired objective. The focus of this paper is the validation of the proposed complex multi-objective reward function in DQL for VSLC through a modified learning process.

This paper is organized as follows. Section II describes the basics of VSLC and Section III the basics of Deep Q-Learning (D-QL). Following Section IV discusses the application of D-QL for VSLC. Section V presents and comments on the obtained results. The last section VI presents the conclusion and possible future work on the subject matter.

II. Variable Speed Limit Control

The purpose of VSLC is to increase the motorway operational capacity and decrease the accident risk as it is mentioned in [7]. The impact of VSLC on aggregate traffic flow is most evident through the homogenization of speeds, which means fewer variation between speeds of vehicles as well between lanes. Results summarized in [8] show that applying a slightly reduced speed limit, when the traffic flow is close to the operational capacity, will produce a temporary decrease of the mainstream flow. The results also show that a sufficiently low speed limit value applied to traffic flow where density is beyond critical one causes a permanent decrease of capacity in the fundamental diagram (Fig. 1). Thus, the area under the VSLC becomes an actively controllable mainstream

978-1-7281-5973-7/20/$31.00 ©2020 IEEE

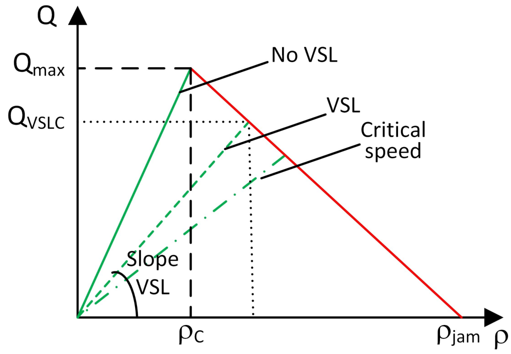

Figure 1. Influence of speed limits on the fundamental diagram

bottleneck that limits the traffic inflow into the congested downstream area. It ensures that the bottleneck is operating near its operational capacity, thus preventing overloading. The effect of reducing traffic outflow from the area under VSLC is the basis for developing control strategies for VSLC.

There are several strategies for VSLC algorithm design as it is described in [8]: (1) based on the predefined values for specific traffic conditions, (2) rule-based algorithms such as the one based on fuzzy-logic in [9], and (3) local feedback-based mainstream traffic flow control with main idea of keeping the detected occupancy as close as possible to the capacity level, and thereby avoiding a capacity drop. Examples for the later are Mainline Virtual Metering (MVM) and Simple Proportional Speed Controller (SPSC) compared in [10]. MVM is based on the fundamental diagram and critical density, while SPSC adequately responds to the changes in downstream density instead of maintaining a fixed desired density near the critical one. The latest approaches based on machine learning techniques and the application of DQL will be elaborated with more details in continuation.

III. Deep Q-learning

In [11], it is stated that Deep Reinforcement Learning (DRL) has emerged in the past few years as a novel control technique for highly nonlinear, stochastic, and data-rich problems. The general RL framework is based on an unsupervised machine learning structure that is embedded into an autonomous agent. It obtains and analyzes the state of its environment and computes an adequate action that will affect the same environment. The objective of the agent is to maximize the accumulated rewards that are computed according to its action's effects on the environment. One of the most used RL approaches is Q-Learning (QL). The learned Q-value function $Q: \ S \times A \mapsto \mathbb{R}$ in QL represents a mapping of state-action pairs to achieve a long-term goal by executing a specific (optimal) action in a given state [12]. In a non-deterministic environment, QL iteratively updates the optimal Q-value by using the newly received training sample $(s_t, a_t, s_{t+1}, r_{t+1})$ according to (1):

$$\begin{aligned} Q^*(s_t, a_t) &= Q(s_t, a_t) \\ &+ \alpha_n(r_t + \gamma \max_{a' \in A} Q(s_{t+1}, a') - Q(s_t, a_t)), \end{aligned} \quad (1)$$

where $Q(s_t, a_t)$ is the Q-value for the respective state-action pair (s_t, a_t) at time step t, γ is discount rate which is reduced over time in order to model process in which the earlier rewards are worthier than the rewards in the future, r_{t+1} is the reward received after performing the action a_t in state s_t and inducing a change to the new state s_{t+1}, and α_n is the learning rate that controls how fast the Q-values are altered. However, the size of Q-table, where all Q-values are stored in relation with states and actions, for a certain complex state-action representations can overgrew and lead to a poor scalability and computational infeasibility.

Recently, Deep Neural Networks (DNN) have shown the ability to process high-dimensional data due to their multi-layered structure. The DNN structure contains convolution layers that are used for feature extraction from data-intensive inputs such as images or image like representations. Afterward, a set of fully-connected layers in the DNN structure is used for classification purposes. Thus, stand-alone DNN is commonly used for natural language processing, semantic segmentation, and classification of images, as it is described in numerous recent studies presented in [13]. The study in [14] has shown that one setup of DNN for QL (known as the Deep Q-Networks (DQN)) can be used for non-linear approximation of a Q-function in game control problems.

Prior to the learning process of DQN, it is necessary to conduct an observation phase in which a Replay Memory (RM) is assembled. At each time step t during the observation phase, the QL agent tracks its current environment and creates observed interaction experience $E_t = (s_t, a_t, r_t, s_{t+1})$. This observed experience is placed into the RM as $M = (M_1, M_2, ..., M_t)$. When RM capacity is reached, the learning starts, and the oldest memory entry is discarded after each learning iteration. Thus, the newest observation can be included in it. According to [15], in order to learn DNN features/parameters θ such that outputs $Q(s, a; \theta)$ approximates the optimal Q-value $Q^*(s, a)$, the QL agent needs training data: where each input data set $X = \{(s_t, a_t) : t \geq 1\}$ corresponding with the target $y = \{Q^*(s_t, a_t) : t \geq 1\}$. It is possible to notice that (1) is replaced with (2) in DQL:

$$Q^*(s_t, a_t) = r_{t+1} + \gamma \max_{a' \in A} Q(s_{t+1}, a'; \theta). \quad (2)$$

At this point it is necessary to estimate the value $r_{t+1} + \gamma \max_{a' \in A} Q(s_{t+1}, a'; \theta a')$ instead of the above mentioned target. In [16], target DQN is introduced as a learning stabilizer and learning divergence reducer. Here $Q(s_{t+1}, a'; \theta a')$ is the output of a separate target DQN with parameters $\theta a'$ and the input s_{t+1} of the target DQN is the same as the input for the other DQN. Also $Q(s_{t+1}, a'; \theta a') = 0$ if training episode is over at time step $t+1$ (e.g. one simulation). Original and target

DQN's have the same architecture. Therefore, targets can be computed as $y = \left\{ \gamma \max_{a' \in A} Q(s_{t+1}, a') : t \geq 1 \right\}$. During the learning process agents learn features/parameters θ of the first DQN in every learning iteration while target DQN features/parameters θ' are updated periodically. Agents learn based on the minimization of the Mean Squared Error (MSE) as it is shown with (3):

$$MSE(\theta) = \frac{1}{m} \sum_{t=1}^{m} \{ r_{t+1} + \gamma \max_{a' \in A} Q(s_{t+1}, a'; \theta') - Q(s_t, a_t; \theta) \}^2, \quad (3)$$

where m is the size of input data set X. It is possible to conclude that if m is large enough, it will consume a significant amount of computational cost. To overcome this problem, this study uses the advanced learning optimization algorithm Adam [17], and a mini-batch learning concept. Mini-batch represents a small stack of elements (or individual experience entries) that are randomly selected from the RM and presented to the DQN during each learning iteration. The learning process contains an exploration phase to find state-action pairs without a Q-value and to compute it. The opposite term is exploitation, and it denotes using of already known information to maximize the rewards. Exploitation-Exploration trade-off analysis is based on ϵ-Greedy algorithm in this study. According to [18], it generally exploits the best available option, but in random moments the ϵ-Greedy algorithm explores the other available options. Probability for performing explorations (or generating random actions) is reduced over time since the DQN is gaining more knowledge.

IV. Applying Deep Q-Learning in VSLC

It is possible to conduct various Q-function approximations for dimensionality reduction of a state-action space for QL based VSLC. One of the approaches presented in [3] is to use feature-based state representations such as Coarse and Tile coding, and Radial Basis Function. According to [4], those approaches provide adequate results only if the states are described in low-dimensional fashion and handcrafted with linear value or policy functions. A most significant study in this direction for VSLC is described in [19]. It uses an actor-critic architecture for DQL to learn a large number of discrete speed limits in a continues action space for purposes of differential VSLC. In paper [11], a per lane VSLC, based on Lagrangian control using Deep-RL is proposed. Since the traffic flow modeled in [11] contains Autonomous Vehicles (AV), the VSLC can directly adjust the speed limits of AVs within specific traffic lane and in this way, control remaining traffic flow, instead of using classical VMSs. The main goal of research in [11] is to investigate the possibility of using DQN to approximate the state-action space function for a VSLC controlled traffic environment.

This study uses a modified learning procedure and complex reward function for DQL in the VSLC application. During the observation phase, a learning sample is stored in RM under

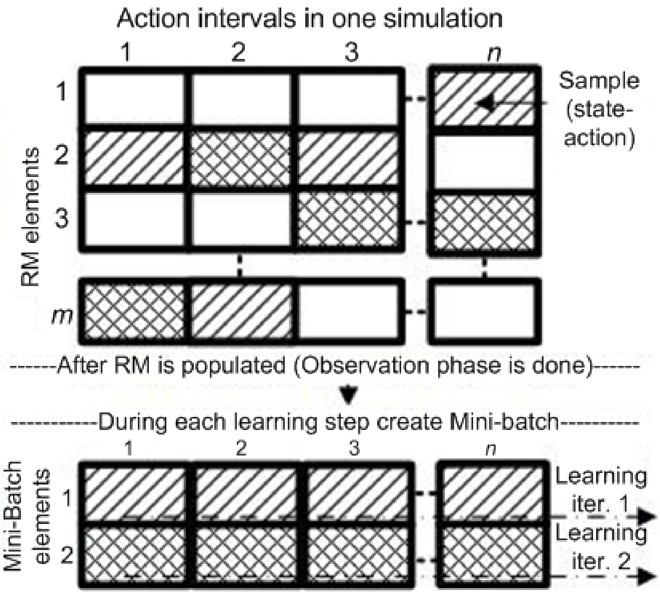

Figure 2. Proposed Mini-Batch design with size 2

an adequate simulation action interval if the measured density for the current sample is smaller than the value of critical density multiplied with the empirically found constant 1.43. Each element in the mini-batch extracted from RM contains several samples that correspond with the number of action intervals in simulation. Furthermore, each element in the mini-batch is made of samples that are randomly selected according to its corresponding action interval in simulation. From a temporal perspective, elements of a mini-batch cover all specific traffic characteristics during one simulation run. The proposed approach enables a more comprehensive learning process. Schematic visualization of the applied Mini-Batch design with size two can be seen in Fig. 2.

Semi-random generation of speed limits in the observation phase is done based on current density at the congested area (effectual bottleneck) shown in Fig. 4. Speed limit values are generated uniformly randomly around the current average speed if a measured density exceeds the critical one. Otherwise, the speed limit will be generated randomly according to the triangle distribution (with a higher probability for the largest speed limit). In case of VSLC for this study, outputs/actions are 7 predefined speed limits that range between 60 (km/h) and 130 (km/h) increasing with a step of 10 (km/h). In this study, the exploration phase in the learning process is starting with a higher probability for choosing semi-random speed limits compared to the ones computed by DQN in ratio 70 to 30. The number of semi-randomly posted speed limits is gradually reduced over time until the start of the exploitation phase. The exploitation phase lasts considerably shorter compared to the exploration phase. A combination of semi-randomly generated speed limits and long exploration phase is done for the sake of a faster learning process and a more comprehensive exploration of the state-action space.

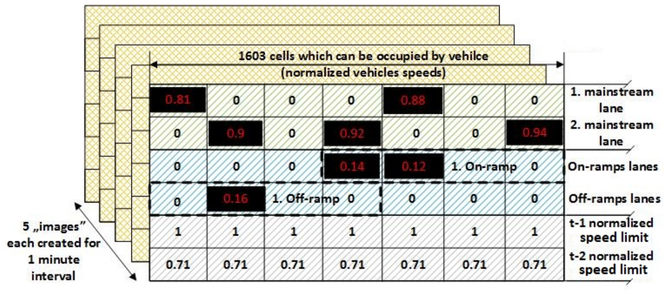

Figure 3. Schematic visualization of "image" like state representation

The efficiency of VSLC based on DQN also depends upon adequate coding of traffic states in an "image" like fashion, the structure of DQN, and reward function design.

A. State Representation

The current traffic situation is presented as a stack of five consecutive "images" of traffic flow on the controlled motorway (Fig. 3). The time-span between each image is $1\ min$. Thus, it is possible to capture the dynamics of vehicle motion. The mentioned image is constructed by dividing the whole simulated motorway into $5\ m$ long cells that represent an average length of one vehicle. If a cell is occupied by a vehicle, it contains its normalized speed, otherwise, it contains the value zero. The so created image can be represented as a matrix with dimensions 6 by 1603. The number of matrix columns represents the number of cells. The first two rows of the matrix denote two traffic lanes on the motorway mainstream, and the following two rows represent on- and off-ramps. Total lengths of the on- and off-ramps are positioned and aligned according to the mainstream lanes for the sake of adequate modeling using a matrix representation. The last two rows represent normalized speed limits posted during the two previous control periods. These additional pieces of information are important since they give insight in speed limits, that were posted before the current control period to prevent oscillations of the control output (posted speed limits).

B. Structure of Deep Q-Network

The best results achieved so far are with the structure of a DQN with three convolution layers. One fully connected layer that contains 512 neurons, and the output layer presented by a fully connected layer with 7 neurons. Each neuron in the output layer represents the Q-value of one possible action or in this case speed limit value. Furthermore, each output of the convolution layer is normalized and passed to another. All convolution and fully connected layers are using Rectifier Nonlinear activation functions (ReLU). Pooling layers are not included in the DQN structure since it is important to be aware of the traffic bottleneck location and location of other traffic anomalies, which can lead to congestions. Parameters of convolution layers are shown in table I.

According to table I, it is possible to conclude that the shape of filters and configuration of strides is adjusted in a line with the structure of traffic scenario "image" on a particular motorway section. Described structure enables a separate assessment of all three parts in the traffic scenario "image". The shape of filters and strides in every other layer is designed according to the aggregation result of previous layers by keeping in mind the original significance of the rows in the traffic scenario "image".

C. Reward Function

A crucial part of QL is the reward function. It steers the Q-values in the direction of the favorable goal. In this study, the reward function is divided into three parts. The first part of the reward function r_{fd} is computed based on the fundamental diagram (Fig. 1) that represents the relation between density and flow. The fundamental diagram contains three segments. The first segment includes free-flow speed between density zero and bi-stable traffic flow, which is set at 75% of critical density for both mainstream traffic lanes. Critical density ρ_c is set to be $56\ veh/km$ while bi-critical density ρ_{bi} is set to $42\ veh/km$. The second segment of the fundamental diagram is located between bi-stable and critical flow, and the final third segment starts at the critical density and ends at the value of maximal density. Maximal possible number of vehicles n_{max} measured each $3\ s$ is set to 3000 for a $3\ min$ long action interval on the whole controlled motorway. Computation of the first reward function part r_{fd} related to the characteristics of fundamental diagram is done according to (4):

$$r_{fd} = \begin{cases} \mid v_{VMS} - v_{max} \mid \cdot C_{v1} & if \rightarrow \rho \leq \rho_{bi} \\ Norm_s \cdot vd_{dif} & if \rightarrow \rho_{bi} < \rho \leq \rho_c \\ \begin{cases} n_{max} & v_{VMSi} > v_{VMSmid} \\ n & otherwise \end{cases} & if \rightarrow \rho > \rho_c \end{cases}, \tag{4}$$

where v_{max} is the maximal speed limit of 130 km/h, v_{VMS} is the posted speed limit, and v_{VMSi} is the index of posted speed limit ranging between 0 and 6 (index 0 represents the speed limit of 60 km/h up to the index 6 that represents 130 km/h), v_{VMSmid} is the index of 3 which denotes the middle index value in the posted speed limit range between 0 and 6, vd_{dif} denotes weighted difference between the current speed v and density ρ in order to alleviate dimensionality problem, $Norm_s$ is normalization factor of desired posted speed indexes in bi-stable traffic flow computed according to the linear mapping between density in bi-stable flow and indexes of speed limit, C_{v1} is the weight added to the difference between the posted and maximal allowed speed, and it is set to 85 according to the executed simulation trials.

TABLE I. Parameters of convolution layers

Convolution layers	First	Second	Third
Number of filters	128	64	16
Dimension of filters	2 x 8	1 x 5	3 x 3
Strides	2 x 3	1 x 3	1 x 3

The second part of reward function r_{dl} is given with (5). It focuses on the difference between the posted speed limit and the currently measured speed. This difference must be below the threshold of 25 km/h so the drivers can safely adjust their speed without the need for excessive breaking that can induce a negative disruption in traffic flow:

$$r_{dl} = \begin{cases} \mid v - v_{VMS} \mid \cdot C_{v2} & \mid v - v_{VMS} \mid > 25 \\ 0 & otherwise \end{cases}, \quad (5)$$

where C_{v2} is the weight added to the difference between the measured and posted speed limit, and it is set to 55 according to the simulation trials results. The third part of the reward function r_{cs} described with (6) reduces the reward if a high oscillation in speed limits is detected regarding two previous actions v_{t-2} and v_{t-1}. Value of this parameter is reset to zero after each performed action.

$$r_{cs} = \begin{cases} r_{cs} + 400 & v_{t-2} - v_{t-1} \geq 20 \\ r_{cs} + 600 & v_{t-1} - v_t \geq 20 \\ r_{cs} + 400 & r_{cs} > 1000 \end{cases} \quad (6)$$

Final reward r, given with (7), is computed as a composite criteria function. Each of the three reward parts contributes to the final value according to its importance (denoted with a weight coefficient), and therefore steers the reward towards a specific goal. In this study, the largest weight factor is assigned to the part of the reward that is based on the fundamental diagram (r_{fd} given with (4)). The final reward output is normalized in the interval $[-3, 0]$ for the stabilization of the learning process.

$$r = 0.5 \cdot r_{fd} + 0.3 \cdot r_{dl} + 0.2 \cdot r_{cs} \quad (7)$$

V. OBTAINED SIMULATION RESULTS

The motorway model with one VMS for VSLC is developed in the microscopic traffic simulator VISSIM by using synthetic traffic data. The main idea is to evaluate the convergence characteristics of DQL applied in VSLC towards the aims embedded in the proposed reward function. According to the used reward function, the final result should show a reduction in speed limit oscillation and a simultaneous increase of average mainstream speed.

A. Simulation framework, model and traffic data

The simulation framework contains two parts. The first of them is an application unit that utilizes a VSLC and data processing shell. The second part represents a simulation unit based on the VISSIM traffic simulator. Raw data for each vehicle from VISSIM are delivered by using the VISSIM COM interface to the data processing shell that creates images of the motorway network and store data. It stacks these images and delivers them along with other relevant data to the application core (VSLC algorithm).

The used motorway model contains two on-ramps and one off-ramp with a total mainstream length of 8 km (Fig. 4). The duration of each simulation run is 2 h. Synthetic traffic data are created in a way that a sudden increase in traffic demand after one hour of simulation at mainstream is created and lasts for 20 min causing congestion. This study is based on 10080 simulation runs with the first 500 simulations used for the observation phase.

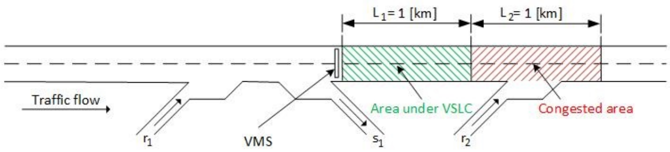

Figure 4. Motorway model

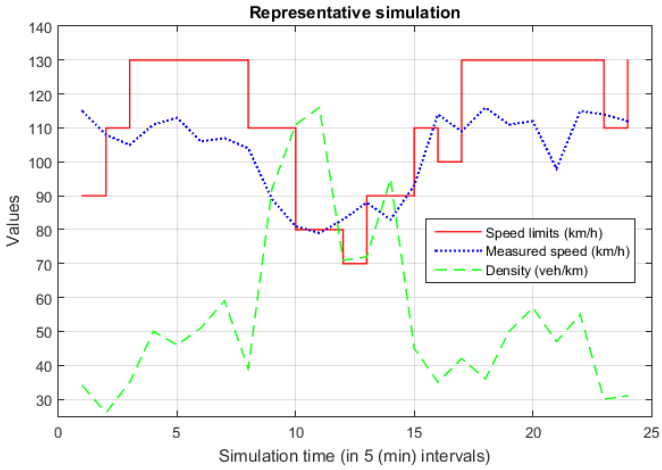

Figure 5. Relation of the measured speed and density, and posted speed limits

B. Simulation results and discussion

Figure 5 shows the relation of the measured speed and density, and posted speed limits. It is possible to conclude that speed limits are gradually decreased when the measured speed starts to decrease its value and consequently density increases its value. Speed limits are continuously and gradually reduced to limit the traffic inflow into the congestion area until the density drops below the critical one. When the congestion is dissolved and density is below critical one, the VSLC starts to produce larger values of the speed limit. It is possible to notice that proposed elements of the reward function, which tend to reduce fluctuations in speed limits in two consecutive time steps, and reduce large differences between measured and posted speed limits, have produced satisfactory results.

In Fig. 6, it is possible to see the convergence of the reward function towards higher values. This tendency is represented also by the linear regression line. It is noticeable that the proposed algorithm achieves lower rewards with intense fluctuation at the start of the learning process. The reward is gradually increased and related fluctuations are reduced as the learning process moves on.

Furthermore, table II shows a comparative analysis of average speed and density in the congested area for the

TABLE II. Comparison of simulation results

	Average Speed (km/h)	Average Density (veh/km)
No control	96.4	28.85
SPSC	97.1	28.86
DQN	109.4	21.65

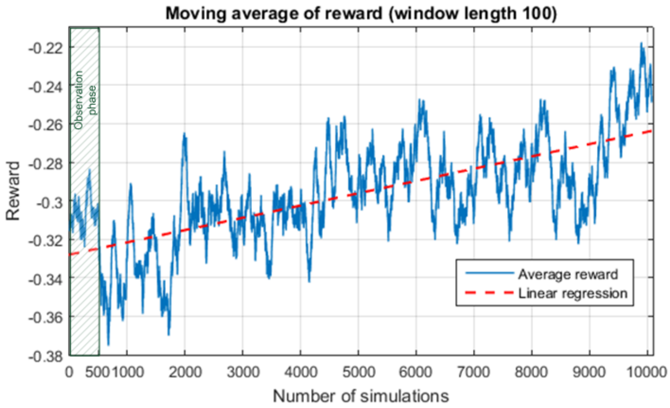

Figure 6. The increase of average reward

same traffic scenario without VSLC, traffic responsive VSLC algorithm SPSC [10], and most representative DQL result. Initial results show that the DQL algorithm produces better results in comparison with the other evaluated scenarios i.e. in comparison to a simple traffic responsive VSLC.

VI. Conclusion and Future Work

This study proposes a VSLC algorithm based on DQL with an adjusted learning process and complex reward function composed of three separate goals. At the same time, it tries to increase the throughput of the motorway by increasing the average mainstream speed, increase safety by decreasing the difference between measured speed and posted speed limit, and minimize oscillation of speed limit in consecutive action intervals. Presented results show that this kind of learning setup can increase average mainstream speed, reduce density and gradually increase overall reward during the learning process. On the other hand, due to the learning setup, speed and stability of convergence during the learning process are low for both previously mentioned parameters. Provided analysis of the posted speed limits and measured speeds shows that the proposed VSLC has managed to prevent significant oscillation between consecutive speed limits, and large differences between posted and measured speed.

Future work on this topic will include the application of a multi-agent DQL system on several consecutive speed limit areas on the motorway. An analysis of the optimal structure of the DQN and reward function to increase the stability of the learning process and convergence rate will also be conducted.

Acknowledgment

This work has been partly supported by the University of Zagreb and Faculty of Transport and Traffic Sciences under the grants "Investigation of the impact of autonomous vehicles on urban traffic flow characteristics" and "Innovative models and control strategies for intelligent mobility", and by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS). The authors thank the company PTV Group for providing a research license of the simulator VISSIM. This research has also been carried out within the activities of the Centre of Research Excellence for Data Science and Cooperative Systems supported by the Ministry of Science and Education of the Republic of Croatia.

References

[1] E. Grumert, A. Tapani, and X. Ma, “Characteristics of variable speed limit systems,” *European Transport Research Review*, vol. 10, 2018.

[2] E. R. Müller, R. C. Carlson, W. Kraus, and M. Papageorgiou, “Microsimulation analysis of practical aspects of traffic control with variable speed limits,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, Feb 2015.

[3] K. Kušić, E. Ivanjko, and M. Gregurić, “A comparison of different state representations for reinforcement learning based variable speed limit control,” in *2018 26th Mediterranean Conference on Control and Automation*, June 2018.

[4] S. M. A. Shabestary and B. Abdulhai, “Deep learning vs. discrete reinforcement learning for adaptive traffic signal control,” *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018.

[5] Y. Wu, H. Tan, Z. Jiang, and B. Ran, “ES-CTC: A Deep Neuroevolution Model for Cooperative Intelligent Freeway Traffic Control,” *ArXiv*, 2019.

[6] M. Gregurić, M. Vujić, C. Alexopoulos, and M. Miletić, “Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data,” *Applied Sciences*, vol. 10, no. 11, 2020.

[7] R. Ziolkowski, “Effectiveness of Automatic Section Speed Control System Operating on National Roads in Poland,” *PROMET - Traffic & Transportation*, vol. 31, no. 4, 2019.

[8] M. Papageorgiou, E. Kosmatopoulos, and I. Papamichail, “Effects of variable speed limits on motorway traffic flow,” *Transportation Research Record: Journal of the Transportation Research Board*, pp. 37–48, 2008.

[9] D. Li and P. Ranjitkar, “A fuzzy logic-based variable speed limit controller,” *Journal of advanced transportation*, vol. 49, 06 2015.

[10] E. Ivanjko, K. Kušić, and M. Gregurić, “Simulational analysis of two controllers for variable speed limit control,” *Ahead of print: Proceedings of the Institution of Civil Engineers - Transport*, 2019.

[11] E. Vinitsky, K. Parvate, A. Kreidieh, C. Wu, and A. Bayen, “Lagrangian Control through Deep-RL: Applications to Bottleneck Decongestion,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 759–765, Nov 2018.

[12] A. A. Sherstov and P. Stone, “Function approximation via tile coding: Automating parameter choice,” in *Proceedings of the 6th International Conference on Abstraction, Reformulation and Approximation*, SARA’05, (Berlin, Heidelberg), pp. 194–205, Springer-Verlag, 2005.

[13] S. Zhou, Q. Chen, and X. Wang, “Discriminative deep belief networks for image classification,” in *2010 IEEE International Conference on Image Processing*, Sep. 2010.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. A. Riedmiller, “Playing Atari with Deep Reinforcement Learning,” *CoRR*, 2013.

[15] J. Gao, Y. Shen, J. Liu, M. Ito, and N. Shiratori, “Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network,” *CoRR*, 2017.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, and et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, pp. 529–533, 2015.

[17] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *International Conference on Learning Representations*, Dec 2014.

[18] M. Aslani, M. S. Mesgari, S. Seipel, and M. Wiering, “Developing adaptive traffic signal control by actor–critic and direct exploration methods,” *Proceedings of the Institution of Civil Engineers - Transport*, vol. 172, no. 5, 2019.

[19] Y. Wu, H. Tan, and B. Ran, “Differential variable speed limits control for freeway recurrent bottlenecks via deep reinforcement learning,” *ArXiv*, vol. abs/1810.10952, 2018.