Review

An Overview of Reinforcement Learning Methods for Variable Speed Limit Control

Krešimir Kušić ^{*,†} , Edouard Ivanjko [†] , Martin Gregurić and Mladen Miletic

Faculty of Transport and Traffic Sciences, University of Zagreb, Vukelićeva 4, HR-10 000 Zagreb, Croatia; edouard.ivanjko@fpz.unizg.hr (E.I.); martin.greguric@fpz.unizg.hr (M.G.); mladen.miletic@fpz.unizg.hr (M.M.)

* Correspondence: kresimir.kusic@fpz.unizg.hr

† These authors contributed equally to this work.

Received: 1 July 2020; Accepted: 15 July 2020; Published: 17 July 2020

Abstract: Variable Speed Limit (VSL) control systems are widely studied as solutions for improving safety and throughput on urban motorways. Machine learning techniques, specifically Reinforcement Learning (RL) methods, are a promising alternative for setting up VSL since they can learn and react to different traffic situations without knowing the explicit model of the motorway dynamics. However, the efficiency of combined RL-VSL is highly related to the class of the used RL algorithm, and description of the managed motorway section in which the RL-VSL agent sets the appropriate speed limits. Currently, there is no existing overview of RL algorithm applications in the domain of VSL. Therefore, a comprehensive survey on the state of the art of RL-VSL is presented. Best practices are summarized, and new viewpoints and future research directions, including an overview of current open research questions are presented.

Keywords: intelligent transportation systems; urban motorways; variable speed limit; reinforcement learning; deep learning; multi-agent systems

1. Introduction

Today's available real-time traffic data from motorways enables a practical application of services from Intelligent Transportation Systems (ITS) to be used as traffic control measures for improving traffic [1]. This improvement is made in terms of better utilization of existing infrastructure rather than building an additional one. This is specially done on urban motorways that present an integral part of the urban traffic network. Urban motorways serve as bypasses for transit traffic, but they also serve the local traffic originating from the city or connect city districts. They are well connected with the local urban traffic network through a larger number of near on- and off-ramps. Periodically, they become oversaturated, mostly near an on-ramp caused by local traffic entering the motorway. This happens because urban motorways are also heavily used by the local commuters that also generate severe congestion at its on- and off-ramps.

The cost of indirect consequences of traffic congestion is represented by the form of time spent by passengers in the traffic jams, air pollution, and accidents. This presents a significant problem today. For example, emissions from road traffic are responsible for 72% of total greenhouse gas emissions from the transport sector in EU [2]. Thus, the EU plans to reduce emissions produced by the transport sector significantly, and any such reduction on motorways can lead to tremendous gains for the cost-effectiveness in general. Building additional road infrastructure is one way to adopt the traffic network to the constant increase in traffic demand. However, expanding the motorway infrastructure is not always a solution as proved by the theory of induced traffic demand, where increases in motorway capacity will induce additional traffic demand [3]. This is known as the Braess paradox [4].

Two of the commonly used traffic control strategies for the mainstream traffic flow control on urban motorways are Variable Speed Limit (VSL) control and Ramp Metering (RM). VSL influences the traffic flow dynamics by changing the allowed speed limit on a particular motorway section [5]. RM influences the motorway mainstream by allowing only a portion of the traffic flow from a particular on-ramp to enter the motorway according to the current traffic flow conditions [6]. Thus, preventing the occurrence of congestion on the motorway mainstream. The focus of this study is on VSL. VSL control systems can be classified as static or dynamic as it is explained in [7]. Static VSL is based on hourly or seasonal changes in the speed limit, while the dynamic one is directly dependent on the current traffic flow or weather conditions. The main objective of VSL control systems is to improve throughput and traffic safety on motorways due to the concept of speed harmonization [8–10]. VSL aims to ensure stable traffic flow in motorway areas affected by recurrent bottlenecks. Thus, it has a twofold influence of preventing and alleviating congestion as well. Several classic strategies for VSL have been developed in order to optimize the traffic flow [11–13]. The control logic of such VSL controllers is based on the classical feedback control theory. They only react to imminent changes in the traffic, which results in a delayed response. Additionally, they require information of the traffic flow dynamics in terms of the fundamental diagram (relation between flow, speed, and density) to tune the controller parameters [14]. The fundamental diagram has to be estimated for each controlled section to ensure the optimal behavior of the controller.

Machine Learning (ML), specifically Reinforcement Learning (RL) techniques [15,16], are promising alternatives for modeling VSL and addressing the mentioned problems. Especially the problem of knowing the fundamental diagram of the controlled section. Various RL algorithms have been successfully applied for solving traffic flow control on motorways [17–23]. It is also worthy to mention other applications of RL in sustainable mobility, like RL-based RM [24], traffic signal control [25], vehicle relocation and ride request assignment in shared autonomous mobility-on-demand systems [26] or RL-based assisted power management [27] for electric hybrid bicycle in bike-sharing systems [28]. However, the efficiency of the combined RL and VSL (RL-VSL) approach is highly related to the learning process. The learning process is performed in traffic simulations since it applies random speed limit values in the beginning. Thus, potentially creating dangerous situations for a real-world implementation. Simulations cover scenarios depending on the initial parameters with a drawback of not covering all possible relevant traffic states (incidents, driver behavior, the influence of different weather situations). The learning process should be performed in a structured manner [29], complementing existing traffic scenarios with synthesized ones that evoke or replicate substantial aspects of all relevant real traffic scenarios [30]. Although the desirable learning convergence can be achieved by using a smaller number of traffic scenarios, it can be infeasible to apply RL-VSL in real-world motorway environments due to the stochastic and continuous nature of traffic flows. Thus, choosing the appropriate learning process is crucial and there is a need for a systematic review of applying RL for VSL.

To the best of our knowledge, there is no comprehensive systematic overview of applications of RL algorithms in the domain of VSL. An overview of VSL control strategies, including classical VSL strategies and theoretical background of influences of VSL on traffic flow and its applications, is presented in [5]. However, it lacks the part related to RL based approaches. So far, the main motivation for this research is to present the state of the art of currently used RL algorithms for solving the VSL optimization problems on urban motorways. This research provides a clear, comprehensive overview of the concept of RL with applications in the field of VSL, pointing out open questions. Therefore, the main contributions of this paper are:

- In this study, the systematic literature review approach is applied. A keyword-based search was used, and 12 primary studies were systematically identified from search results.
- Traffic management studies on motorways focused on RL approaches to solve the VSL optimization problem are covered by this study.

- Unlike existing studies such as [31], which focus on summarizing the approaches of VSLs traffic management systems, this study tries to assess how well the present RL-VSL approaches work based on the provided results. First, the objectives of the approaches, such as improving efficiency or safety, were identified and categorized. Then, different approaches were compared on how well they meet a specific goal. In addition, the RL methods used to solve VSL and how the VSL problem is being modeled for a particular objective for intelligent traffic management on motorways were identified and summarized.

2. Application of RL in VSL Control

The application of methodologies from the domain of artificial intelligence in traffic control provides a new opportunity for addressing the issues with current VSL control approaches. Optimization of VSL requires the determination of an optimal policy for posting speed limits as actions on Variable Message Signs (VMS) or sending them directly to vehicles in the case of the Vehicle to Infrastructure (V2I) communication environment, Autonomous Vehicles (AV) or to Connected Autonomous Vehicles (CAV). For this, RL can be applied without the need to explicitly specify how that task is to be achieved, as shown in [17–19,21–23,32–37].

2.1. Variable Speed Limit Control

VSL is a traffic control approach that optimizes the mainstream traffic on motorways by adapting the speed limit according to real-time traffic and weather conditions. Nowadays, the speed limit is being displayed on VMS installed at motorways notifying drivers about the current speed limit. Correctly positioned VMSs on the motorway enable the VSL to operate efficiently in terms of flow harmonization or to alleviate congestion occurring in the areas near on- and off-ramps [11]. The significant contributions of VSL on the traffic flow is seen through the speed harmonization [8,9,38], which is reflected in the smaller speed differences of vehicles between lanes and within the lane, as well as in reduced speed variance between upstream free-flow and the downstream flow in the congested area. Consequently, a safer and more stable traffic flow is obtained [39–41].

Preventing traffic breakdown by managing the flow rate entering the bottleneck area by inducing an artificial "controllable" bottleneck using VSL upstream of the real bottleneck is studied in [11,42,43]. The reduced mean speed of vehicles under the active VSL area produces higher traffic density and a decrease in the mainstream flow. A reduction in flow will be temporary for higher speed limits and lower traffic loads [44], while permanent reduction is evident for lower speed limit values at higher traffic loads [45]. Such artificially created bottleneck located upstream of occurrence of an unmanageable bottleneck (congestion) allows the VSL controller to limit the inflow rate to the congested area to the approximate value of the bottleneck operational capacity ($\approx q_{cap}$) which prevents traffic flow breakdown and, thus, enables higher throughput through a congested area. This is the basic working principle of VSL when used for bottleneck control, as presented in Figure 1.

Figure 1. Application of Variable Speed Limit (VSL) for the bottleneck control.

As shown in [11], in addition to obtaining speed limits, the positioning and choosing of the appropriate length of the section covered by VSL is the crucial part for optimal utilization of the VSL

system. A recent overview of VSL control strategies, which summarizes the VSL influence on the managed traffic flow, can be found in [31].

Reinforcement Learning

The benefit of using RL over classical feedback control methods is that there is no need for knowing the explicit model of traffic flow dynamics and the influence of VSL on it. RL in general faces this problem by a learning agent that learns to act by trial-and-error interactions within its environment [16,46] (see Figure 2). In the standard RL paradigm, an agent is connected to its environment via the perception and action framework. At each step of interaction, the agent senses the environment and then selects an action to change the state of the environment. This state transition generates a reinforcement signal (reward or penalty) received by the agent [47]. The agent's environment is assumed to be modeled as a Markov Decision Processes (MDP). An MDP is a tuple $\langle S, A, P, R \rangle$ where S is a finite set of n discrete states and A is a finite set of actions available to the agent. The actions are stochastic and Markovian in the sense that an action a_t in a given state $s_t \in S$ results in a state s_{t+1} with fixed probability $P(s_{t+1}|s_t, a_t)$, while R represents the reward received after state transition. Policy π is a function representing a mapping from states to actions $\pi : S \mapsto A$, which optimizes performance, for example, expected accumulated reward received. RL methods differ according to the exact measure and optimization criteria (total reward optimization, discounted reward optimization, or average reward optimization) used to select actions.

Figure 2. Reinforcement learning framework for VSL control.

While taking actions (e.g., speed limits in the case of VSL agent) by trial-and-error, the agent incrementally learns a “value function” over state-action pairs $Q_i(s_t, a_t)$, which indicates their utility to that agent. Q-Learning (QL) [47] is one of the most implemented *value iteration* RL algorithms. The value function updating rule for QL is as follows:

$$Q_i(s_t, a_t) := (1 - \alpha_Q)Q_i(s_t, a_t) + \alpha_Q(r_{t+1} + \gamma \max_{a' \in A} Q_i(s_{t+1}, a')), \quad (1)$$

where t is the discrete time step, action a_t in state s_t induces a state change to the new state s_{t+1} . Depending on that transition, the agent receives the reward r_{t+1} . The goal is to find an optimal policy π^* that maximizes the expected discounted reward for state s . Parameter α_Q is the learning rate that controls how fast the Q -values are altered. Discount factor γ controls the importance of future rewards. Finally, an agent looks one step ahead at the maximum $Q_i(s_{t+1}, a')$ value given by optimal action a' in new state s_{t+1} . At any time, the RL methods use one-step lookahead with the current value function to choose the best action in each state by some kind of maximization. The agent can explore optimal actions or exploit its current control policy. Therefore, the policies that RL methods learn are called *greedy* with respect to their value functions. In addition to such greedy actions, RL methods also take

some directed or random (exploratory) actions. This ensures that all states are visited plenty of times so that the learning method does not get stuck in a local optimum. There are several exploration strategies to ensure this. The random exploration strategy takes random action with a fixed probability, giving high probabilities to actions with high values. The counter-based exploration prefers to execute actions that lead to infrequently visited states. Recency-based exploration promotes actions, which have not been executed recently in a given state. A commonly used scheme for action selection is ϵ – greedy strategy, which takes a random action with probability ϵ and a greedy action with probability $1 - \epsilon$.

The Deep Reinforcement Learning (DRL) approach represents the extension of the classical RL used as the latest approach to tackle specific VSL control problems such as differential VSL where each lane receives its own speed limit value [37]. It is based on the learning representations of data. This approach attempts to model high-level abstractions in the data by using multiple processing layers with complex structures or otherwise composed of multiple nonlinear transformations [48]. The model, which contains multiple processing layers with complex structures based on the concept of the human cerebral cortex, is known as Deep Neural Network (DNN). The DNN integrates feature extraction, and classification (or prediction) processes in a single framework by using information-dense input datasets [1]. The DRL can also be used as the framework that integrates several different traffic control methods on motorways such as VSL control, RM, and lane change control [36].

3. Research and Implementation

3.1. Research Method and Research Questions

As shown in Table 1, this study is focused on pure RL-VSL on motorways. Papers that proposed methodologies to improve motorway performance by considering traffic control using VSL were reviewed. To achieve the set objectives, three main research questions were formulated:

- RQ1. What factors address speed limit management studies in terms of utilizing RL to solve the VSL problem?
- RQ2. What kinds of methodologies have been proposed to address the potential problems related to intelligent VSL systems?
 - RQ2.1. What kinds of RL algorithms have been proposed for VSL when traffic flow is consisting of only AVs?
 - RQ2.2. What kinds of RL algorithms have been proposed for VSL when traffic flow consists of a mixture of autonomous and human-driven vehicles?
- RQ3. What challenges and open questions arise?

Articles available online and published in English between January 2010 and 20 June 2020 were in the focus of this study. The following digital libraries were included:

- Scopus;
- IEEE;
- Web of Science (WoS).

We used keyword-based searches to identify primary studies to filter relevant articles based on appropriately selected keywords (variable speed limit, traffic flow, control, reinforcement learning) used in the WoS digital library.

3.2. Conducting the Review

In Table 1, chronologically listed most representative approaches for VSL design based on the RL framework are shown. In the same table, improvements are denoted according to the first control methodology listed in the column “Compared against”.

Table 1. The most representative Reinforcement Learning (RL) frameworks for VSL control.

Paper	Year	RL Method	Compared against	Control Strategy	Improvements
[35]	2020	Q-Learning	no-control	VSL	19.4% lower crash risk
[49]	2020	W-Learning	no-control, single and independent VSL agents	multi-agent VSL	18.1% lower bottleneck density
[34]	2019	Distributed Q-Learning (kNN-TD)	no-control	multi-agent VSL	51% lower <i>TTS</i>
[19]	2019	Maximax MARL	no-control, feedback, independent, hierarchical MARL	multi-agent RM and VSL	11% lower <i>TTS</i>
[23]	2018	Deep-RL (GRU)	no-control, feedback RM	centralized agent VSL to AVs	25% higher bottleneck throughput
[20]	2018	Q-Learning linear approx.	no-control, Q-tabular, (Tile, Coarse, RBF)	VSL	10% lower overall <i>TTS</i>
[21]	2017	Q-Learning (kNN-TD)	no-control, FB-VSL	VSL	21.8% lower <i>TTT</i>
[18]	2016	O-Learning (ANN)	no-control, certain fixed amount of VSL	VSL	30% lower <i>TTS</i>
[17]	2014	R-MART	no-control	VSL	18% lower <i>TTS</i>

3.3. Results of Research Questions

Studies listed in Table 1 have shown that RL-VSL has been extensively studied as a promising alternative for solving the VSL optimization problem. In this subsection, the overview of each particular study is given, starting with the earliest ones.

Studies in [17,18] have demonstrated formulations of the VSL problem as RL problems and eventually solved the VSL problem using RL algorithms. Later on, in [17], the so-called Reinforcement-Markov Average Reward Technique (R-MART) algorithm was used to manage VSL. The state space for the learning algorithm is described with four discrete traffic density values describing the free flow, lightly congested, moderately congested, and heavily congested traffic flow. Actions are defined by a set of eight elements with values ranging from 44 to 129 km/h that increase or decrease by the constant amount 12 km/h. Although the algorithm itself is similar to QL, it differs in that the value in the Q-function $Q(s, a)$ is obtained as the expected mean of the rewards collected for each time step. In general, the R-MART algorithm uses the concept of long-term mean rewards instead of deferred rewards, as is the case with QL. The reward function is formulated towards the minimization of the Total Time Spent (*TTS*). In addition to measuring the effectiveness of the proposed solution, the vehicle emission gas CO₂ was measured, highlighting the correlation between vehicle speeds influenced by VSL. This parameter was not included in the reward function as objective. The research is significant because it was made on the model of the wider traffic city network of the city of Sioux Falls (SD, USA), while other research is mainly based on the analysis of the effect of RL based VSL on a specific isolated segment of an urban motorway. Using the R-MART based VSL controller, a reduction of 18% for *TTS* and about 20% less exhaust gas CO₂ was achieved compared to the case without VSL (static speed limit 72 km/h). The results show that the problem of smooth changes in the output values of VSL controllers (speed limits), that occasionally oscillate and at some point two adjacent limits take large speed differences, which can affect safety, is not solved.

Later on, in [18], the QL algorithm is used to learn to manage VSL in order to optimize traffic flow on motorways. To describe the agent's environment, the state vector $\vec{s}$ contains six components: two previous actions (speed limits) and currently measured speeds at four consecutive sections around

the area where congestion occurs. The set of actions contains four speed limits $\{60, 80, 100, 120\}$ km/h. The reward function is proportional to the negative *TTS* measured between the two control intervals (duration of each action). An additional condition is added to the reward that prevents oscillations of the speed limit, and at the same time, penalizes changes in speed limit in which the consecutive absolute difference is greater than 20 km/h. An element is also added, representing a reward (amount 0) if the agent recognizes the free-flow conditions and allows vehicles to drive at a maximum speed of 120 km/h. The macroscopic model METANET [50] was used to simulate the proposed solution. The Tile coding method was used to generalize the continuous state variables in the case of linear Q-function approximation [16]. The results were also compared with a nonlinear approximation using an Artificial Neural Network (ANN). In the case of ANN, an additional variable describing the traffic density on the managed segment was added to classify the state conditions on the managed motorway. Further, an important difference between previous work is that additional variables were included in the state vector, such as the predicted speed calculated using a parallel METANET simulation based on the current traffic situation and expected traffic demand. As shown by the results, in this case, they were better by 2% in using cell coding, and 0.8% in the case of ANN compared to the case without speed prediction. The no-control case was taken as a starting point, in which the mentioned methods showed a decrease in *TTS* by approximately 30%. The case for the best *TTS* achieved for a certain fixed amount of speed limit is also taken into account. In this case, the VSL controllers failed to achieve a better result suggesting that the proposed VSL learning system needs to be further improved. Also, the authors performed an analysis of the robustness of action selection strategies (control policy) by adding noise (modeled using the Gaussian distribution) to the measured traffic parameters speed and density. Control policy behaves well for noise up to 10% otherwise the learned control policy can no longer be used to appropriately select speed limits.

Research in [21] has shown that VSL based on the QL algorithm (QL-VSL) can achieve better results compared with close loop-based algorithms, which uses speed limits adjusted based on control feedback loops (FB-VSL). The FB-VSL sets a speed limit slightly after the traffic jam has already been formed. This can be too late because the control error can be large. Thus, in order to correct the control error, the VSL controller must significantly reduce the speed limit to reduce inflow rate into the congested area in order to maintain traffic parameters (measured density ρ in the bottleneck) around the desired value ρ_d (often critical density ρ_c for which the maximal traffic flow is achieved). Occasional, more restrictive actions (such as lower speed limits) can negatively impact the flow upstream of the controlled sections and induce new congestion in front of the VSL area. A comparative analysis of controllers has shown that QL-VSL can recognize a potentially problematic condition leading to bottleneck activation and, thus, proactively correct the speed limit before congestion occurs. It can be said that the agent, through interaction with the environment through trials and errors, has the ability to successfully learn how to recognize certain critical traffic (pattern) situations. In that way, QL-VSL is able to predict disturbances in traffic flow and the potential to act preventively somewhat earlier than the classic FB-VSL controllers. QL-VSL succeeds with much less intervention (less speed limit correction) to achieve a better effect in preventing bottleneck activation or capacity drop on the segment of the urban motorway where congestion occurs. Proactive VSL control reduces the likelihood of the creation of larger congestion and reduces the need for stronger actions (lower speed limit values). The environment in which the QL-VSL agent operates is described by the discrete values of measured traffic density in the bottleneck area (mainstream motorway segment immediately after the on-ramp) and the density at the problematic on-ramp. The final set of actions in [21] is ranging from 20 to 65 mph with an increment of 5 mph. The agent's indirect goal is to minimize the Total Travel Time (*TTT*) by preventing the occurrence or alleviating the consequences of an already active bottleneck. The effect of each action at a specific traffic flow condition is assessed by the reward function defined by the Poisson density distribution (veh/km) in the bottleneck area stimulating the agent to strive to maintain a flow around ρ_c at maximum vehicle flow and thus minimize *TTT*. In order to accelerate the convergence of the learning process, an additional parameter "prize" (+200) is introduced for two adjacent states

around ρ_c . To penalize the congested state (density much higher than the ρ_c), an additional “penalty” value of -400 is added to the reward function. The learning process takes place offline, i.e., in the background on continuously collected data (states, actions, and rewards). In this way, the VSL agent is able to periodically refresh its knowledge based on continuously collected new traffic data and thus adapt its control policy if needed. In the case of the overspeed issue in terms of relaxing the driver compliance rate to the speed limits, the “continuous learning” approach for QL-VSL outperformed the basic QL-VSL approach regarding reduction in measured *TTT*. Eventually, the proposed QL-VSL strategy outperforms baselines (FB-VSL and no-control) with an improvement of *TTT* up to 21.84% in the fluctuating traffic demand scenario.

A drawback in modeling VSL as MDP on a longer motorway stretch is a large number of needed state-action pairs that lead to the problem of the exponential growth of the solution space. It becomes impossible to search for an optimal solution in real-time. This is known as the *curse of dimensionality* [51]. As shown in [18], this can be solved with function approximation techniques. In [20], three different approaches for feature-based state representation (Coarse and Tile coding, and Radial Basis Function (RBF)) having a linear function approximator for *Q*-function were compared for solving the QL-VSL problem. The mentioned methods were tested on a synthetic motorway model modeled in the microscopic traffic simulator VISSIM. Results show that function approximation methods outperform QL-based VSL formulated with a *Q*-table by an average improvement of 10% regarding the convergence of reward function (*TTS* in this case), where feature extraction methods (Coarse and Tile coding) showed a slightly faster learning rate and a more stable control policy.

In [19], an algorithm for simultaneously solving VSL and RM by means of Multi-Agent RL (MARL) was proposed in order to reduce congestion in the bottleneck area produced by the interference of entering vehicles from the on-ramp with vehicles in the mainstream. Traffic conditions or “states” of the agent’s environment in case of VSL are described with the densities measured directly downstream ρ_{ds} of the on-ramp (bottleneck location), the density ρ_{app} measured directly upstream of the bottleneck “applied VSL area” (VSL_{app}), and the third state variable is the density ρ_{us} measured on the motorway section further upstream from that “upstream section” comprising the area considered for the second state variable ρ_{app} . In the case of RM, only the first two density values were taken as well as the length of the vehicle queue at the on-ramp. Actions (allowed speed limits) are defined by the set $\{90, 100, 110, 120\}$ km/h. Learning algorithms based on temporal-difference learning and weighted k-Nearest Neighbors for linear function approximation denoted as kNN-TD [52] were used to approximate the *Q*-function. Depending on the Euclidean distance between the entry point of the state and the points that represent the centers of each group, the corresponding *Q*-value is refreshed depending on the severity of the state belonging to each group. Thus, the simplest method tested considering RM and VSL agents learn independently. In the second case, a hierarchical MARL approach was used where an agent with a higher hierarchy is the first to bring an action, and this action is communicated to the second-highest ranked agent who takes this action into account and selects its own action accordingly. The third approach is the so-called maximax MARL based on the principle of locality of intersection among agents [53]. In it, a mapping of the effect of an agent’s actions to the global value function (in this case, only the neighboring agents are considered) is being done [54], in which agent *i* and neighboring agent *j* look for actions that maximize the *Q*-value gain. Only the agent whose action has the largest contribution to the increase in joint payoff may change the action, while other neighboring agents may not. The process is repeated until the maximum joint payoff is achieved. It can be said that agents cooperate locally and their usefulness is further used in the calculation of the global *Q*-value. By applying the maximax MARL coordinated control of the VSL and RM using kNN-TD learning algorithm, an additional improvement of *TTS* compared to all baselines was achieved. Almost 11% compared to the no-control case. In order to reduce the difference

in speed limit from 120 km/h to VSL_{app} , the speed limit at the upstream section is adjusted using the rule given with the following Equation (2):

$$VSL_{us} = \max[(VSL_{app} + \delta), 120], \quad (2)$$

where $\delta = 10$ km/h and denotes the maximal difference between the upstream VSL_{us} and applied VSL_{app} speed limit. This ensures a smoother speed transition of vehicles between the adjacent upstream (free-flow) segment and the segment managed by VSL. Thus, preventing sudden changes in vehicle speed reduces the likelihood of shock waves on the mainstream. A microscopic simulation model is based on a motorway stretch of the N1 national motorway outbound from Cape Town in South Africa's Western Cape Province and was developed within the AnyLogic 7.3.5. University Edition software suite using built-in Road Traffic and Process Modeling Libraries.

In [34], a novel VSL control algorithm under the V2I environment aiming to optimize the motorway traffic mobility and safety is presented. The control system is a multi-agent system consisting of several VSL agents. The agents work cooperatively using the proposed distributed RL approach DQL-VSL where kNN-TD algorithm is used as the general function approximator (see Figure 3). The goal of the control algorithm is to improve traffic mobility by maintaining motorway traffic density slightly under the critical point ρ_c to produce the maximum traffic volume. In the case of traffic safety, the objective is to reduce the speed difference between adjacent segments (Figure 4). The control system is tested using the open-source microscopic traffic simulation software MOTUS. Separated results were analyzed regarding the motorway section and motorway bottleneck as well regarding the mentioned two objectives. Results revealed that compared with no-control cases, the proposed DQL-VSL can noticeably decrease the system TTT and increase the bottleneck outflow. Moreover, the speed difference between motorway segments indicating the potential rear-end collision risk is significantly reduced. In the case of traffic mobility control and measured traffic parameters in the bottleneck, TTT is lower for 51%, mean speed is higher for 35% and outflow from the bottleneck is higher for 24%. Interesting results are obtained regarding the average bottleneck speed difference of speed between bottleneck and the adjacent segment where lower reduction is achieved by using traffic safety control strategy compared to traffic mobility control. This could imply a lack in the modeling of the safety control objective.

Figure 3. Scheme of traffic state collection and speed limit executing process [34].

Figure 4. Time to the collision between two adjacent segments [34].

In [22], a control strategy based on RM and VSL was proposed. To calculate the RM and VSL control output, the Eligibility Traces based Reinforcement Learning (ETRL) algorithm was used. The state variables are density measured within each section, while action contained RM rate and VSL values at the same time. Therefore, there is no actual cooperation between RM and VSL as it was introduced in the work of [19]. The reward function is modeled as the *TTT* minimization problem. The motorway model is based on M62 eastbound motorway and traffic data provided by England Highway Agency. Testing is done in MATLAB by using macroscopic simulations based on the Cell Transmission traffic flow Model (CTM). During the rush hours, the average speed is maintained above 70 km/h, which is how the authors state high average speed for rush-hour congestion.

In [49], authors introduced W-learning based VSL (WL-VSL): a novel multi-agent RL-based VSL control approach. WL-VSL is implemented on a motorway simulation scenario where two agents are learning, using the W-learning algorithm [55], to control two segments upstream of a congested area jointly (Figure 5). The reward function (3) for each agent is based on the agent's local performance by sensing of average speed $\bar{v}_{i,t+1}$ within agent's segments as well as the downstream bottleneck by sensing the *TTS* value measured in sections L_1 , L_2 and L_3 , as denoted in Figure 5.

$$r_{i,t+1} = - \left[\left(1 - \frac{\bar{v}_{i,t+1}}{\delta} \right) \left(\frac{1}{2} (TTS_1 + TTS_2) + TTS_3 \right) \right] \quad (3)$$

Figure 5. Motorway model [49].

WL-VSL is evaluated on microscopic simulations using two traffic scenarios with dynamic and static traffic demand in the SUMO simulator. WL-VSL was compared with base cases: no control, single agent, and two independent agents. WL-VSL outperforms these baselines with respect to measured traffic parameters (*TTS*, density, and average speed) in the downstream congested area of the simulated urban motorway. Improvement of traffic parameters up to 18.18% regarding the decrease of traffic density and 7.3% higher average speed is achieved in the bottleneck during the peak hour in the case of multi-agent WL-VSL.

The study [35] proposed an RL-based VSL control algorithm to reduce crash risks associated with oscillations near the bottleneck. More precisely, the state, action, and reward for the QL-based VSL control were designed to improve safety. Density within three motorway sections are used to be the state variables (Figure 6). The speed limit ranges from 20 to 65 mph with an increment of 5 mph. The objective of QL-VSL control is to minimize the total risks of crashes associated with oscillations near motorway bottlenecks. The total crash risk was calculated using a crash risk prediction model developed by the same authors [56], taking into account the characteristics of vehicle

deceleration trajectories. The QL-VSL was trained to learn the optimal speed limit for various traffic states to achieve a goal of safety optimization. The CTM based simulation model was modified as the simulation platform for evaluating the control effects. The results showed that after the training process, the proposed QL-VSL control successfully reduced the crash risks by 19.4% while only increased the TTT by 1.5%. A continuous online learning function was developed in RL to enhance the robustness of the control strategy regarding the overspeed issue. The results showed that with continuous learning, the QL-VSL control performed reasonably well under lower driver compliance situations.

Figure 6. Determination of state set in the RL agent [35].

In [23], a per lane VSL, based on Lagrangian control using Deep-RL (DRL) is proposed. It is a learning structure based on Recurrent Neural Network (RNN) [57] where the application of the so-called Gated Recurrent Unit (GRU) [58] mechanism solves the problem of vanishing gradient (decay of information through time) that occurs when learning a classic RNN using the backpropagation through time procedure. GRU is a simplified version of the Long-Short Term Memory (LSTM) recurrent network introduced in [59] that combines forget and input gates to form an ‘update’ gate. In general, the LSTM can maintain temporal information in the state for a long number of timesteps and is widely used in sequential data analysis. Since the traffic flow used in research contains AVs, a centralized agent based on deep reinforcement learning DRL-VSL can directly adjust the speed to AVs within specific traffic lanes and, in this way, control remaining traffic flow, instead of using classical VMSs. State space of the controller is described with: density and average speed of human drivers in each lane for each observed section, density and average speed of AVs in each lane for each observed section, and the outflow at the final bottleneck. The reward function is modeled to sense the outflow from the bottleneck. The proposed solution was tested using *Flow*, which is a new library for application of DRL to traffic micro-simulators SUMO. RM, the cycle time of which is controlled by a feedback controller, has been used as a baseline to compare the efficiency of proposed DRL-VSL in controlling the inflow rate into the bottleneck. The results show that 10% penetration rate of controlled AV can improve the throughput of the two-stage bottleneck (four lanes reduce to two and then reduce to one) for $\approx 25\%$ but only in the case of high inflows ($inflow \geq 1600 \text{ veh/h}$) compared to the no-control. At lower inflow rate, the DRL-VSL worsens the traffic outflow compared to both no-control and feedback RM.

In [37], the Differential Variable Speed Limit (DVSL) system was tested for variable speed control. Depending on the traffic situation on the motorway segment, DVSL can assign different speed limit values for each lane separately. The traffic model for DVSL testing is part of the San Bernardino Highway in California, USA. It consists of five traffic lanes and one on- and off-ramp. The off-ramp is located immediately after the on-ramp, causing a disturbance in the far right traffic lanes as drivers start re-arranging earlier to be able to get off the motorway. Thus, they intertwine with the flow of vehicles entering the on-ramp, creating congestion. The far left lanes often remain used under their capacity. By applying the classic VSL, the same speed limit value is mandatory in each traffic lane, which can unnecessarily disturb the flow of less occupied traffic lanes due to lower limit values. Hence the solution of applying DVSL that would manage the speed limit depending on the load of a particular lane. Traffic demand in the simulated model was taken from the PeMS (Performance Measurement System) traffic database. A Deep Learning (DL) structure based on the policy iteration

algorithm known as *actor-critic architecture* [60] was used to train the agent for DVSL. The actor generates a speed limit, and the critic evaluates the executed action of the actor. The architecture makes it possible to avoid evaluating the whole set of actions for each lane. The set of actions consists of six speed limits 50, 55, 60, 65, 70, 75 mph. Given the five traffic lanes, the number of combinations is relatively large ($6^5 = 7776$) and is a problem for classical Q-Learning or DQL. Therefore, the Deep Deterministic Policy Gradient (DDPG) algorithm optimizes parameters in the actor-critic architecture since it enables work with large sets of discrete action values. The algorithm first gives a continuous action, and based on it finds a set of the closest discrete speed constraints. To overcome the issue with temporal correlation between samples ($s_t, s_{t+1}, a_t, r_{t+1}$) and rapid forgetting of possibly useful experiences, the prioritized experience replay [61] is used to address these problems by storing experience into a replay memory. The experiences are constantly sampled from the replay memory to update the agent's knowledge, which stabilized the learning of neural networks in DRL. The traffic situation is described by a 15-dimensional vector that describes the occupancy in the area immediately after the flow interference (mainstream-entry ramp), in front of the congestion area (mainstream), and the occupancy at the on-ramp. Four reward functions were tested in which the value evaluates the agent's performance during each control time step (duration of the action is 1 min). The first reward function is proportional to the negative value of the measured *TTS*, the second function is positive, and describes the mean speed in the bottleneck. The third reward function is proportional to the negative number of sudden braking of the vehicles, and the detection threshold of such events is a deceleration of 4.5 m/s^2 . The fourth reward function takes the environmental aspect into account, and is proportional to the negative sum of the scaled values of *CO*, *HC*, *NO_x* and *PM_x* emissions (according to the Euro VI standard). Three baseline cases were taken to compare the results. The first without VSL, the second with QL structure similar to that in [21] with only the reward function being proportional to the negative *TTS* and the third case in which the agent is trained using the DQL algorithm. Both starting cases (QL and DQL) have the possibility of setting the same speed limit for all traffic lanes, i.e., they do not have the possibility of assigning a different speed limit for each individual traffic lane as in the case of DVSL controllers. All learning-based VSL controllers were trained through 150 episodes. Each episode is a simulation lasting 18 simulation hours, where the SUMO microscopic traffic simulator was used for simulations. Upon completion of the learning process, the operation of all RL controllers was simulated on an equal set of simulations (using 50 different simulation seeds) to ensure a valid comparison of the obtained results. The analysis of the results included a comparison of the average cumulative values of the reward functions and the average travel time. By applying DVSL based on the DDPG algorithm, a better result was achieved compared to QL, DQL, and the case without VSL. For example, the DDPG with a reward function proportional to *TTS* reduced the mean cumulative value of *TTS* by approximately 3% compared to the case without VSL. The analysis of the optimal control policy (DDPG approach), depending on the reward function, proved that the most stable learned control policy was the case when the reward function was formulated through *TTS*. The authors note that such a system is not yet possible in reality because it requires CAV technology. This research and the obtained results provide a good insight into the possibility of using AV as excellent actuators for managing the entire traffic system on motorways as well as the design of the deep learning algorithm structure related to the number of actions and defining the appropriate criterion function (reward function).

As shown in [18], predicted traffic parameters can be used as input to the RL-VSL control algorithm to improve the learning and control process. As shown in [62], Convolutional Neural Network (CNN) can be used for speed prediction on a motorway segment. CNN has been proven as suitable for processing 2D data, such as images. Ideally, spatial characteristics of traffic flow on the motorway segment can be represented as an image containing information about traffic speed, density or vehicle position, etc. (see Figure 7). One image represents a two-dimensional matrix. The sequence of traffic environment measurement samples (several images) is most commonly used. In this case, the entrance to the CNN network is a three-dimensional matrix filled with traffic data of several

consecutively collected images in time (channels). Images are continuously sent to the input of a deep neural network.

Figure 7. Traffic-to-image conversion on an urban motorway network.

However, other DL algorithms, such as LSTM can be used for the prediction of traffic parameters as well. For example, in [63], LSTM was used to process time-series traffic data from the motorway segment, thus adding the spatial-temporal information to capture the long time dependency of a nonlinear traffic dynamic in an effective manner. In general, CNN can output one value, while the LSTM structure enables the forecast of a sequence of new traffic states for several future time steps.

4. Discussion

The application of advanced traffic management solutions in the field of ITS enables further improvement of the level of service on urban motorways. One such solution is the VSL control system. VSL can produce a satisfying performance when traffic on the urban motorway is consistent in a spatial-temporal context. However, the effectiveness of VSL decreases in cases when the traffic conditions are exposed to the rapid oscillations in traffic demand or in the case when the motorway capacity is being reduced. RL as one of the approaches for online ML provides an optimal trade-off between the complexity and efficiency among various model-free data-driven traffic control methods. As shown, it is worth applying RL-VSL since its continuous self-adaptation features are able to tackle control problems related to the new, not yet anticipated traffic conditions [18,21,35].

From the completed survey, several potential research directions to address the limitations of the existing methods were identified. The limitations of current FB-VSL controllers are the need for an accurately estimated fundamental diagram and delay. The emphasis of the survey is placed on the description of VSL controllers in which algorithms are based on the RL type of ML. The main efficiency drawback of classical VSL algorithms is their inability to adapt control policy to a new traffic situation in which case they operate sub-optimally [21]. During the last five years, according to Table 1, it is

possible to notice an increase in the number of studies looking forward to improving and proposing the new RL-based VSL control algorithms. Regarding different RL techniques and state-action modeling of the environment, including various types of different objectives, the development of RL-VSL strives to bring such approaches closer to a real-world application.

A few RL based approaches have been successfully applied in VSL control. Mostly used is the QL algorithm, which has been used in its base form using Q-table. It is well integrated with a more sophisticated method such as function approximation (linear [19,20], nonlinear [18]) for Q-function generalization. Even though the authors in [22] stated that the ETRL algorithm can find a better optimum compared to the QL, the results obtained by ETRL-VSL are insufficient to confirm that, since there is a lack of comparison with the QL-VSL. Likewise, in [17], further studies are needed to assess if the R-MART based VSL has any benefits over the QL-VSL. An example of the kNN-TD algorithm proposed in [19] has been successfully applied for Q-function approximation in [35]. As shown in [23,37], an application of DL can be useful when state and action space in the RL-VSL model becomes too large in which case the basic form of RL-VSL (Q-table) cannot be used efficiently.

In [19], an interesting approach is explained regarding the third state variable (density ρ_{us}) measured on the motorway section further upstream from the primary VSL area. This variable is chosen originally to provide a predictive component in terms of motorway demand. Also, it is expected to provide the agent with an indication of the severity of the congestion in cases where it has split back beyond the application area of VSL. In that way, there is no need for predictive models (like in [18]). In that way, the VSL agent can recognize a platoon of vehicles coming into the VSL and downstream bottleneck area. This allows the VSL agent to react even earlier (proactively).

The newly proposed concept in [35] has incorporated a risk model into RL-VSL for safer traffic in the case of traffic oscillations on motorways. It is tested on a macroscopic CTM based simulator. Results showed that after the training process, the proposed QL-VSL control successfully reduced the crash risks by 19.4% while only increasing the *TTT* by 1.5%. Since the proposed risk model is taking into account the characteristics of vehicle deceleration trajectories, potentially a more reliable evaluation will be achieved through microscopic simulations taking microscopic dynamic parameters of each individual vehicle in the calculation. The weather condition as a state variable for QL-VSL conceptually could be the way for further research on risk control on motorways since the weather has a significant impact on traffic dynamics and safety.

In [34], results showed that there could be more than one optimal traffic equilibrium according to different control objectives (safety and flow maximization) due to separately implemented and tested objectives criteria. It will be interesting to have both included as a multi-objective optimization problem or multi-agent MDP modeled VSL problem where each agent seeks to find a balance between its local goal while at the same time not deviating from the global goal. In that case, it could be possible to have only one unique equilibrium regarding the optimal control policy.

In [21], a continuous offline learning function was proposed in RL to enhance the robustness of the control strategy regarding the overspeed issue. The results showed that with continuous learning, the QL-VSL control performed reasonably well under lower driver compliance situations regarding reduction in *TTT*. This is desirable since, in a real environment on motorways, the compliance rate can vary a lot and needs to be considered in the calculation. Similarly, in [18], an analysis of the robustness of action selection strategies (control policy) by adding noise to the measured traffic parameters (speed and density) was performed. Control policy behaves well for noise up to 10% otherwise learned control policy can no longer be used to appropriately select speed limits. This indicates that in a real environment, the placement of sensors has to be carefully chosen to be able to reconstruct missing data if one or more sensors break down.

Regarding the hierarchical MARL approach in [19] for solving multi-agent RM and VSL problem, the action is communicated to the second-highest ranked agent. As shown, the result of this communication considering hierarchical MARL, the state-action space of the second agent grows by a factor equal to the number of actions available to the first agent. In contrast, as proposed in [49]

the WL-VSL strategy is based on the idea that there is no need for any global controller analyzing the behavior of all agents. Similarly to DQL-VSL [34], the WL-VSL does not require additional communication between agents. The agents are incrementally setting their so-called *W*-values, using only local information (same in QL). An agent will be aware of its competition indirectly by the interference they cause. As stated in [34], the distributed system can deploy the controllers flexibly along the motorway, and there is no concern about the breakdown of the central traffic controller. Due to the low number of available research (until now [34,49]) considering developing and testing the pure multi-agent RL-VSL systems, further investigation is needed in those directions.

5. Conclusions

The identified open areas of application of VSL based on RL methods are as follows. From the traffic engineering perspective, the concept of establishing a multi-agent RL-VSL system based on the cooperation of agents that could simultaneously position and determine the length of the VSL area as well as calculate the speed limit jointly has to be considered in the future studies. These three decisions could be calculated simultaneously by three different agents within the VSL. Such dynamic positioning of the VSL area on the urban motorway depending on the current spatial-temporal characteristics of the traffic flow, as well as integrating several connected VSLs on a longer motorway segment could achieve smoother speed transition. Therefore, enabling more effective traffic flow control, limiting the need for more restrictive actions such as lower speed limits directly in front of congestion which negatively impact the free-flow upstream of the controlled section. From the algorithmic side, an open area is visible in MARL where decentralized approaches have to be further investigated since centralized approaches are prone to the curse of dimensionality.

Until now, all analyses are based on discretized time sampling of traffic states and discrete actions used for the RL algorithm. Real-time traffic management is one of the areas of interest given that in the real environment of the motorway, the position and speed of the vehicle changes as a continuous variable. The application of continuous variables to describe agent actions and states would allow smoother speed control of an individual vehicle, which is interesting for testing simulation models that support a CAV and AV environment. This will trigger-off the usage of the policy iteration algorithm for the case of continuous-valued action and RNN architecture like LSTM and GRU, which are able to take spatial-temporal behavior of the traffic as input into the learning process for a more accurate DRL-VSL control. The weather condition has a significant impact on traffic dynamics and safety, and it should be examined in-depth as a state variable in the RL-VSL control as well.

Author Contributions: The conceptualization of the study was done by K.K., M.G., and E.I., who also did the funding acquisition. The writing of the original draft and preparation of the paper was done by K.K., and M.M. All authors contributed to the writing review and final editing. The supervision was done by E.I. Visualizations were done by K.K. All authors have read and agreed to the published version of the manuscript.

Funding: 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).

Acknowledgments: 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.

Conflicts of Interest: The authors declare no conflict of interest. The funding institutions had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
AV	Autonomous Vehicles

CAV	Connected Autonomous Vehicles
CTM	Cell Transmission traffic flow Model
CNN	Convolutional Neural Network
DDPG	Deep Deterministic Policy Gradient
DL	Deep Learning
DRL	Deep Reinforcement Learning
DQL	Deep Q-Learning
DVSL	Differential Variable Speed Limit
GRU	Gated Recurrent Unit Neural Net
ETRL	Eligibility Traces based Reinforcement Learning
FB-VLS	Feedback Variable Speed Limit
ITS	Intelligent Transportation Systems
kNN	k-Nearest Neighbors
LSTM	Long-Short Term Memory
ML	Machine learning
MDP	Markov Decision Processes
QL	Q-Learning
RL	Reinforcement Learning
R-MART	Reinforcement-Mark Average Reward Technique
RM	Ramp Metering
RNN	Recurrent Neural Network
TCA	Traffic Control Agent
TTT	Total Travel Time
TTS	Total Time Spent
VMS	Variable Message Signs
VSL	Variable Speed Limit
WoS	Web of Science

References

- Gregurić, M.; Vujić, M.; Alexopoulos, C.; Miletić, M. Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data. *Appl. Sci.* **2020**, *10*, 4011. [\[CrossRef\]](#)
- European Environment Agency. *Greenhouse Gas Emissions from Transport in Europe*; Technical Report; EEA: Copenhagen, Denmark, 2019.
- Noland, R. Relationships between highway capacity and induced vehicle travel. *Transp. Res. Part A Policy Pract.* **2001**, *35*, 47–72. [\[CrossRef\]](#)
- Braess, D. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung* **1968**, *12*, 258–268. (In German) [\[CrossRef\]](#)
- Lu, X.Y.; Shladover, S.E. Review of Variable Speed Limits and Advisories: Theory, Algorithms, and Practice. *Transp. Res. Rec.* **2014**, *2423*, 15–23. [\[CrossRef\]](#)
- Papageorgiou, M.; Kotsialos, A. Freeway ramp metering: An overview. *IEEE Trans. Intell. Transp. Syst.* **2002**, *3*, 271–281. [\[CrossRef\]](#)
- Castro, Á.G.; de Cáceres, A.M. Variable Speed Limits. Review and development of an aggregate indicator based on floating car data. In Proceedings of the 13th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’13); Riga, Latvia, 16–19 October 2013; Transport and Telecommunication Institute: Riga, Latvia; pp. 117–127.
- Strömberg, P.; Lind, G. Harmonization with Variable Speed Limits on Motorways. *Transp. Res. Procedia* **2016**, *15*, 664–675. [\[CrossRef\]](#)
- van den Hoogen, E.; Smulders, S. Control by variable speed signs: Results of the Dutch experiment. In Proceedings of the Seventh International Conference on ‘Road Traffic Monitoring and Control’, London, UK, 26–28 April 1994; pp. 145–149. [\[CrossRef\]](#)
- Gao, C.; Xu, J.; Li, Q.; Yang, J. The Effect of Posted Speed Limit on the Dispersion of Traffic Flow Speed. *Sustainability* **2019**, *11*, 3594. [\[CrossRef\]](#)

11. Müller, E.R.; Carlson, R.C.; Kraus, W.; Papageorgiou, M. Microsimulation Analysis of Practical Aspects of Traffic Control with Variable Speed Limits. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 512–523. [\[CrossRef\]](#)
12. Jeon, S.; Park, C.; Seo, D. The Multi-Station Based Variable Speed Limit Model for Realization on Urban Highway. *Electronics* **2020**, *9*, 801. [\[CrossRef\]](#)
13. Dadashzadeh, N.; Ergun, M. An Integrated Variable Speed Limit and ALINEA Ramp Metering Model in the Presence of High Bus Volume. *Sustainability* **2019**, *11*, 6326. [\[CrossRef\]](#)
14. Lighthill, M.J.; Whitham, G.B. On kinematic waves II. A theory of traffic flow on long crowded roads. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **1955**, *229*, 317–345. [\[CrossRef\]](#)
15. Watkins, C.J.C.H. Learning from Delayed Rewards. Ph.D. Thesis, King's College, Cambridge, UK, 1989.
16. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; A Bradford Book; The MIT Press: Cambridge, MA, USA, 1998.
17. Zhu, F.; Ukkusuri, S.V. Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *41*, 30–47. [\[CrossRef\]](#)
18. Walraven, E.; Spaan, M.T.; Bakker, B. Traffic flow optimization: A reinforcement learning approach. *Eng. Appl. Artif. Intell.* **2016**, *52*, 203–212. [\[CrossRef\]](#)
19. Schmidt-Dumont, T.; van Vuuren, J.H. A case for the adoption of decentralised reinforcement learning for the control of traffic flow on South African highways. *J. S. Afr. Inst. Civ. Eng.* **2019**, *61*, 7–19. [\[CrossRef\]](#)
20. Kušić, K.; Ivanjko, E.; Gregurić, M. A Comparison of Different State Representations for Reinforcement Learning Based Variable Speed Limit Control. In Proceedings of the MED 2018—26th Mediterranean Conference on Control and Automation, Zadar, Croatia, 19–22 June 2018; pp. 266–271.
21. Li, Z.; Liu, P.; Xu, C.; Duan, H.; Wang, W. Reinforcement Learning-Based Variable Speed Limit Control Strategy to Reduce Traffic Congestion at Freeway Recurrent Bottlenecks. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3204–3217. [\[CrossRef\]](#)
22. Tabadkani Aval, S.S.; Shojaei Ghandeshtani, N.; Akbari, P.; Eghbal, N.; Noori, A. An Eligibility Traces based Cooperative and Integrated Control Strategy for Traffic Flow Control in Freeways. In Proceedings of the 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 24–25 October 2019; pp. 40–45.
23. Vinitsky, E.; Parvate, K.; Kreidieh, A.; Wu, C.; Bayen, A. Lagrangian Control through Deep-RL: Applications to Bottleneck Decongestion. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 759–765. [\[CrossRef\]](#)
24. Lu, C.; Huang, J.; Gong, J. Reinforcement Learning for Ramp Control: An Analysis of Learning Parameters. *Promet—Traffic Transp.* **2016**, *28*, 371–381. [\[CrossRef\]](#)
25. Dusparic, I.; Monteil, J.; Cahill, V. Towards autonomic urban traffic control with collaborative multi-policy reinforcement learning. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 2065–2070.
26. Guériau, M.; Dusparic, I. SAMoD: Shared Autonomous Mobility-on-Demand using Decentralized Reinforcement Learning. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 1558–1563.
27. Hsu, R.C.; Liu, C.; Chan, D. A Reinforcement-Learning-Based Assisted Power Management with QoR Provisioning for Human–Electric Hybrid Bicycle. *IEEE Trans. Ind. Electron.* **2012**, *59*, 3350–3359. [\[CrossRef\]](#)
28. Lee, J.; Jiang, J. Enhanced fuzzy-logic-based power-assisted control with user-adaptive systems for human-electric bikes. *IET Intell. Transp. Syst.* **2019**, *13*, 1492–1498. [\[CrossRef\]](#)
29. Schumann, R.; Tamarcaz, C. Towards systematic testing of complex interacting systems. In Proceedings of the First Workshop on Systemic Risks in Global Networks co-located with 14. Internationale Tagung Wirtschaftsinformatik (WI 2019), Siegen, Germany, 24 February 2019; pp. 55–63.
30. Zambrano-Martinez, J.L.; Calafate, C.T.; Soler, D.; Cano, J.C. Towards Realistic Urban Traffic Experiments Using DFRouter: Heuristic, Validation and Extensions. *Sensors* **2017**, *17*, 2921. [\[CrossRef\]](#)
31. Khondaker, B.; Kattan, L. Variable speed limit: An overview. *Transp. Lett.* **2015**, *7*, 264–278. [\[CrossRef\]](#)
32. Wang, W.; Yang, Z.S.; Zhao, D.X. Control model of variable speed limit based on finite horizon Markov decision-making. *Jiaotong Yunshu Gongcheng Xuebao/J. Traffic Transp. Eng.* **2011**, *11*, 109–114.
33. Fares, A.; Gomaa, W.; Khamis, M.A. MARL-FWC: Optimal Coordination of Freeway Traffic Control Measures. *arXiv* **2018**, arXiv:1808.09806. Available online: <http://xxx.lanl.gov/abs/1808.09806> (accessed on 20 June 2020).

34. Wang, C.; Zhang, J.; Xu, L.; Li, L.; Ran, B. A New Solution for Freeway Congestion: Cooperative Speed Limit Control Using Distributed Reinforcement Learning. *IEEE Access* **2019**, *7*, 41947–41957. [CrossRef]
35. Li, Z.; Xu, C.; Pu, Z.; Guo, Y.; Liu, P. Reinforcement Learning-Based Variable Speed Limits Control to Reduce Crash Risks near Traffic Oscillations on Freeways. *IEEE Intell. Transp. Syst. Mag.* **2020**. [CrossRef]
36. Wu, Y.; Tan, H.; Jiang, Z.; Ran, B. ES-CTC: A Deep Neuroevolution Model for Cooperative Intelligent Freeway Traffic Control. *arXiv* **2019**, arXiv:1905.04083. Available online: <http://xxx.lanl.gov/abs/1905.04083> (accessed on 20 June 2020).
37. Wu, Y.; Tan, H.; Ran, B. Differential Variable Speed Limits Control for Freeway Recurrent Bottlenecks via Deep Reinforcement learning. *arXiv* **2018**, arXiv:1810.10952. Available online: <http://xxx.lanl.gov/abs/1810.10952> (accessed on 20 June 2020).
38. Smulders, S. Control of freeway traffic flow by variable speed signs. *Transp. Res. Part B* **1990**, *24*, 111–132. [CrossRef]
39. Allaby, P.; Hellinga, B.; Bullock, M. Variable speed limits: Safety and operational impacts of a candidate control strategy for freeway applications. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 671–680. [CrossRef]
40. Pauw, E.; Daniels, S.; Franckx, L.; Mayeres, I. Safety effects of dynamic speed limits on motorways. *Accid. Anal. Prev.* **2017**, *114*. [CrossRef]
41. Ziolkowski, R. Effectiveness of Automatic Section Speed Control System Operating on National Roads in Poland. *Promet—Traffic Transp.* **2019**, *31*, 435–442.
42. Iordanidou, G.; Roncoli, C.; Papamichail, I.; Papageorgiou, M. Feedback-Based Mainstream Traffic Flow Control for Multiple Bottlenecks on Motorways. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 610–621. [CrossRef]
43. Lu, X.Y.; Varaiya, P.; Horowitz, R.; Su, D.; Shladover, S.E. A new approach for combined freeway variable speed limits and coordinated ramp metering. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, ITSC 2010, Funchal, Portugal, 19–22 September 2010; pp. 491–498.
44. Hegyi, A.; De Schutter, B.; Hellendoorn, J. Optimal coordination of variable speed limits to suppress shock waves. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 102–112. [CrossRef]
45. Papageorgiou, M.; Kosmatopoulos, E.; Papamichail, I. Effects of Variable Speed Limits on Motorway Traffic Flow. *Transp. Res. Rec. J. Transp. Res. Board* **2008**, 37–48. [CrossRef]
46. Bae, H.; Kim, G.; Kim, J.; Qian, D.; Lee, S. Multi-Robot Path Planning Method Using Reinforcement Learning. *Appl. Sci.* **2019**, *9*, 57. [CrossRef]
47. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
48. Benuwa, B.B.; Zhan, Y.Z.; Ghansah, B.; Wornyo, D.K.; Banaseka Kataka, F. A Review of Deep Machine Learning. *Int. J. Eng. Res. Afr.* **2016**, *24*, 124–136. [CrossRef]
49. Kušić, K.; Dusparic, I.; Guériau, M.; Gregurić, M.; Ivanjko, E. Extended Variable Speed Limit control using Multi-agent Reinforcement Learning. In Proceedings of the 23rd IEEE International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 22–23 September 2020.
50. Messmer, A.; Papageorgiou, M. METANET: A macroscopic simulation program for motorway networks. *Traffic Eng. Control* **1990**, *31*, 466–470.
51. Belletti, F.; Haziza, D.; Gomes, G.; Bayen, A.M. Expert Level Control of Ramp Metering Based on Multi-Task Deep Reinforcement Learning. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1198–1207. [CrossRef]
52. Martin H, J.A.; Asiaín, J.; Maravall, D. Robust high performance reinforcement learning through weighted k-nearest neighbors. *Neurocomputing* **2011**, *74*, 1251–1259. [CrossRef]
53. Nair, R.; Varakantham, P.; Tambe, M.; Yokoo, M. Networked distributed POMDPs: A synthesis of distributed constraint optimization and POMDPs. *Proc. Natl. Conf. Artif. Intell.* **2005**, *1*, 133–139.
54. El-Tantawy, S.; Abdulhai, B.; Abdelgawad, H. Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1140–1150. [CrossRef]
55. Humphrys, M. Action Selection Methods Using Reinforcement Learning. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 1996.
56. Li, Z.; Ahn, S.; Chung, K.; Ragland, D.R.; Wang, W.; Yu, J.W. Surrogate safety measure for evaluating rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks. *Accid. Anal. Prev.* **2014**, *64*, 52–61. [CrossRef] [PubMed]
57. Han, Z.; Zhao, J.; Leung, H.; Ma, K.F.; Wang, W. A Review of Deep Learning Models for Time Series Prediction. *IEEE Sens. J.* **2019**. [CrossRef]

58. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078. Available online: <http://xxx.lanl.gov/abs/1406.1078> (accessed on 20 June 2020).
59. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
60. Barto, A.G.; Sutton, R.S.; Anderson, C.W. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. Syst. Man Cybern.* **1983**, *SMC-13*, 834–846. [CrossRef]
61. Schaul, T.; Quan, J.; Antonoglou, I.; Silver, D. Prioritized Experience Replay. *arXiv* **2015**, arXiv:1511.05952. Available online: <http://xxx.lanl.gov/abs/1511.05952> (accessed on 20 June 2020).
62. Ma, X.; Dai, Z.; He, Z.; Wang, Y. Learning Traffic as Images: A Deep Convolution Neural Network for Large-scale Transportation Network Speed Prediction. *arXiv* **2017**, arXiv:1701.04245. Available online: <http://xxx.lanl.gov/abs/1701.04245> (accessed on 20 June 2020).
63. Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C Emerg. Technol.* **2015**, *54*. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).