Preprints, 10th IFAC International Symposium on
Advances in Automotive Control
Columbus, Ohio, USA, August 28-30, 2022

Dynamic and Interpretable State Representation for Deep Reinforcement Learning in Automated Driving

Bilal Hejase [*] **Ekim Yurtsever** [*] **Teawon Han** [**]
Baljeet Singh [**] **Dimitar P. Filev** [***] **H. Eric Tseng** [***]
Umit Ozguner [*]

[*] *Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: hejase.2@osu.edu, yurtsever.2@osu.edu, ozguner.1@osu.edu).*
[**] *Ford Greenfield Labs, Palo Alto, CA 94304 USA (e-mail: than10@ford.com, bsing124@ford.com)*
[***] *Ford Motor Company, Dearborn, MI 48121 USA (e-mail: dfilev@ford.com, htseng@ford.com)*

Abstract: Understanding the causal relationship between an autonomous vehicle's input state and its output action is important for safety mitigation and explainable automated driving. However, reinforcement learning approaches have the drawback of being black box models. This work proposes an interpretable state representation that can capture state-action causalities for an automated driving agent, while also allowing for the underlying formulation to be general enough to be adapted to different driving scenarios. It also proposes encoding temporally-extended information in the state representation for better driving performance. We test this approach on a reinforcement learning agent in a highway simulation environment and demonstrate that the proposed state representation can capture state-action causalities in an interpretable manner. Experimental results show that the formulation and interpretation can be used to adapt the behavior of the driving agent to achieve desired, even unseen, driving behaviors after training.

Keywords: autonomous vehicles; reinforcement learning control; state representation; interpretability; generalization; deep learning

1. INTRODUCTION

Learning-based approaches for automated driving systems (ADS) have shown promise due to their ability to generalize over scenarios and learn behaviors from real data as compared to rule-based approaches.

Imitation learning (IL) has emerged as a popular supervised learning approach for ADS due to its simplicity and ability to train end-to-end on offline data. In such approaches, a driving agent learns an optimal policy from expert or human demonstrations. These demonstrations are often collected by driving a vehicle through diverse traffic and weather conditions or through simulated environments. It has been applied to a variety of driving tasks such as driving behavior prediction in Han et al. (2019), lane following in Bojarski et al. (2016), and urban driving in Codevilla et al. (2018). However, such methods have struggled to solve complex driving tasks and have been limited to specific driving functions as they require the collection of large amounts of driving data and the learned behaviors are limited to the driving scenarios in the data. Prakash et al. (2020) also show that IL approaches suffer from covariate shift and generalize poorly to new environments. Codevilla et al. (2019) further show that these methods suffer from dataset bias and causal confusion.

Mnih et al. (2015) first proposed a framework that combined reinforcement learning (RL) with deep learning to achieve human-level control on a wide variety of tasks. Since then, deep RL methods have found success in ADS due to their ability to handle sequential decision-making problems and to generalize knowledge to unseen scenarios; Yurtsever et al. (2020). Rather than train on collected data, RL methods train a neural network by collecting and sampling interactions with the environment. These interactions are usually performed in simulation environments which allows the agent to experience edge cases such as collisions more frequently. Kendall et al. (2019) demonstrate the success of deep RL in driving a real vehicle along a countryside road with no traffic. Bewley et al. (2019); Osiński et al. (2020) demonstrate simulation-to-reality transfer of a learned driving policy for steering control on a closed road. Toromanoff et al. (2020) demonstrated the ability to handle complex urban driving including lane changes, vehicles, pedestrians, and traffic lights. Deep RL has also been applied as end-to-end systems that directly map raw sensory input to actuation signals; see Chen et al. (2019, 2021).

[⋆] The Ohio State University (OSU) authors have been supported by a Ford grant 60076067 to OSU.

For safety-critical applications, it is important to be able to interpret and understand a taken action. However, deep learning methods lack interpretability due to their black-box nature. Intermediate state representations have been proposed to alleviate the issue of interpretability and to improve sample efficiency. Müller et al. (2018) find that low-level abstractions of the environment, as compared to raw input, improve sim-to-real transfer, domain generalization, and sample complexity. Chen et al. (2015) propose an interpretable, low-dimensional state representation of the driving scene using human-designed perception indicators termed *affordance indicators*. These perception indicators can represent the curvature of a lane or the distance to a vehicle. They coin *direct perception* as a paradigm that maps an input image directly to a set of affordance indicators using a deep Convolutional Neural Network under supervised learning. This representation has been applied to the urban setting in Sauer et al. (2018) and to the highway setting in Nageshrao et al. (2019).

While the affordance indicators provide a useful representation of the driving scenario, it is not able to capture the intrinsic motivation that causes a vehicle to perform a certain action nor adapt to new information after training. Against this backdrop, this work proposes the driving forces as a dynamic state representation that is low-dimensional, interpretable, and can encode both the intrinsic motivation and the environment. The formulation is such that the state representation can be generalized to different road conditions making it useful for learning-based agents. This work also proposes encoding temporal information into the proposed state representation to improve safety. We train an RL-based agent in a highway environment on this state representation and illustrate the capability to generate different driving behaviors through the formulation, even after training.

2. PROBLEM FORMULATION

Consider the task of highway driving on a three-lane road with traffic vehicles and the agent architecture in Fig. 1. The vehicle receives the affordance indicators (I_t) generated by a direct perception module at each instant. As this work is not concerned with the training of the direct perception module, affordance indicators are represented by ground-truth labels provided by the simulator. The *Driving Forces Module* generates the driving forces (D_t) from the received affordance indicators. A *DDQN* RL agent (see Van Hasselt et al. (2016)) generates an appropriate high-level longitudinal and lateral action from the state representation. An *Explicit Safety Check* ensures that the agent only explores safe actions as defined by a set of hand-crafted rules to optimize the search on the state space. The safety controller is as proposed in Nageshrao et al. (2019). The *Low-level PID controller* generates the ego control commands that correspond to the high-level action received.

This sequential decision-making problem is modeled as a Markov Decision Process (MDP) with states S, discrete action set A, transition function T, and reward function R. The agent at state $s \in S$ chooses an action $a \in A$, which is revised by the rule-based safety check to ensure a safe action $a = \hat{a}$. The action determines the new state s' through the transition function $T(s, a)$, and receives a reward $R(s, a, s')$. The goal of the reinforcement learning driving agent is to find the optimal policy that maximizes the cumulative reward. The state S_t is a function of the observations, $S_t = f(I_t)$, where this function maps the observations to the driving forces.

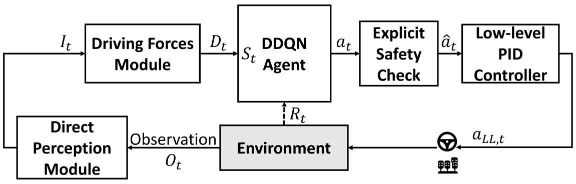


Fig. 1. Overall agent architecture.

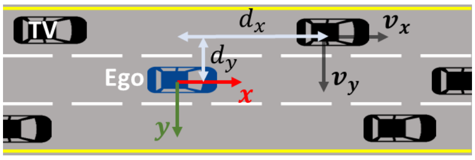


Fig. 2. Three-lane highway simulation scenario. Affordance indicators for a traffic vehicle are shown.

3. DYNAMIC STATE REPRESENTATION

In this section, the driving forces as an interpretable low-dimensional state representation for a learning-based agent is presented. The affordance indicators are introduced followed by the formulation for the driving forces.

3.1 Affordance Indicators

The work of Nageshrao et al. (2019) defines a set of affordance indicators to sufficiently describe the road scenario in a three-lane highway setting. Following the same formulation, the affordances for a traffic vehicle in the front and rear of the left, right, and center lanes is given by:

$$\{d_x, d_y, v_x, v_y\} \tag{1}$$

where d_x, d_y, v_x, and v_y represent the longitudinal distance, lateral distance, longitudinal velocity, and lateral velocity with respect to the ego vehicle (see Fig. 2). Three affordance indicators are also defined for the ego vehicle state: the lateral position, the longitudinal velocity, and the lateral velocity. This state representation contains a total of 27 affordance indicators, $I_t \in \mathbb{R}^{27}$, as it is assumed the ego vehicle can see a maximum of six vehicles.

3.2 Driving Forces

The process of understanding the motivation of an ego vehicle can be divided into several factors: (1) environmental factors such as the road layout and static obstacles; (2) inter-agent factors such as the effect of surrounding traffic on the ego vehicle; (3) intra-agent factors such as the desire to follow a certain speed. The driving forces are formulated as a set of artificial potential functions that capture these factors similar to the work of Fredette and Özguner (2016). The formulation needs to also be low-dimensional, interpretable, applicable to learning-based agents, and generalizable to arbitrary road conditions, layouts, and traffic vehicles. Note that to generalize the

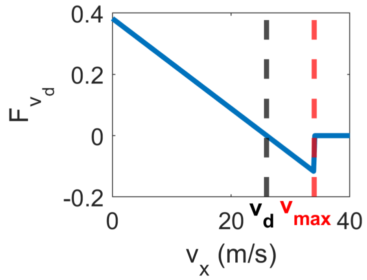


Fig. 3. Velocity following force. The force increases in the positive or negative sense when deviating from the desired speed.

forces to different formulations and retain their meaning, it is important that the values produced are normalized. Based on these requirements, the driving forces from the affordance indicators are presented. In all formulations, $u(\cdot)$ denotes the step function, x the longitudinal direction, and y the lateral direction.

Velocity following force The velocity following force, F_{v_d}, represents the internal motivation for the vehicle to follow the desired speed. The desired speed can either be a constant or some speed profile to be followed while driving. This force is formulated as follows:

$$F_{v_d}(v_x) = \frac{v_d - v_x}{v_{max}} u(v_{max} - v_x) u(v_{max} + v_x) \qquad (2)$$

where $F_{v_d} \in [-1, 1]$. v_d, v_x, and v_{max} represent the desired speed, longitudinal speed, and maximum speed respectively. This force scales with deviation from the desired speed as shown in Fig. 3. The term $u(v_{max} - v_x)$ ensures that the force is not infinitely increasing or decreasing.

Road profile force The road profile force, F_{RA}, is defined as the motivation of the vehicle to move to a certain lateral position on the road. The magnitude of the potential is used to represent the desired lateral positions. Such a formulation confines the agent to the bounds of the road. Different formulations can be considered depending on the scenario at hand. The following formulation corresponds to an arbitrary lane road:

$$F_{RA}(y) = \sum_i h_i e^{\frac{-(L_i - y)^2}{l \omega_R}} \qquad (3)$$

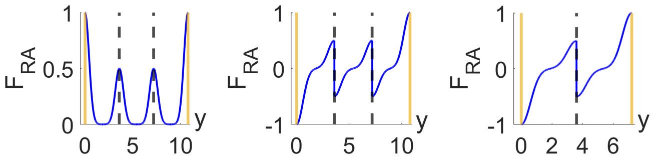
where i is the lane marker index, L_i is the lane marker position, h_i is the lane marker gain, l is the lane width, and ω_R is the force variance. The parameter ω_R can be selected to control how much freedom the vehicle can deviate from the centerline before the force begins to exponentially increase. For example, a larger variance would result in a higher force when deviating from the centerline. Alternatively, one can also compute the road profile force as:

$$F_{RA}(y) = \sum_i \operatorname{sgn}(L_i - y) h_i e^{\frac{-(L_i - y)^2}{l \omega_R}} \qquad (4)$$

where,

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

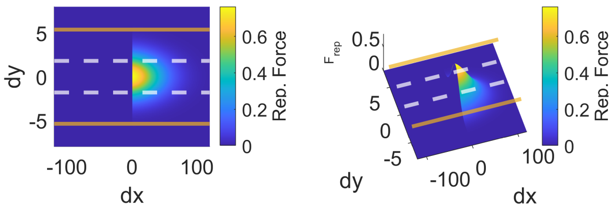
In the simulations performed, we represent the force using (3). The road profile force for the case of a three-lane highway ($i = 4$) and a two-lane highway ($i = 3$) is shown in Fig. 4. First plot uses (3) and the other two plots use (4).



(a) (b) (c)

Fig. 4. Road profile force for a three-lane and two-lane road configuration. Lane markings shown with dashed lines.

The formulation can also be extended to road closures, restricted lanes, and on-road obstacles by representing them as areas of large force. One can also consider lane biasing to the road profile to induce a preferred lane change direction. In comparison, the affordance indicators often assign a state feature for each lane to be detected and can not extend to arbitrary road layouts.



(a) (b)

Fig. 5. Forward-looking inter-agent repulsion force along the longitudinal axis centered around the ego vehicle. The lane boundaries are shown in the white dashed lines.

Inter-agent repulsion force The inter-agent repulsion force captures the influence of traffic vehicles on the ego vehicle. We define the forward-looking force as follows:

$$F_{rep}(d_x, d_y) = \sum_j \left[u(d_{x,j}) e^{-\frac{d_{x,j}^2}{\sigma_x^2}} d_{x,j} e^{-\frac{d_{y,j}^2}{\sigma_y^2}} \right] \qquad (5)$$

where j is the traffic vehicle index, $d_{x,j}$ and $d_{y,j}$ are the relative lateral and longitudinal distance to traffic vehicle j, σ_x^2 and σ_y^2 are the force variance along the longitudinal and lateral directions.

Figure 5 shows a visualization of the forward-looking repulsion force. Different car-following behaviors can be induced by changing the magnitude of the force through the variance. For example, increasing the variance increases the effect of the force. The formulation can also be applied to an arbitrary number of traffic vehicles allowing the incorporation of other sources of information such as interconnected vehicles.

Lane change force The lane change force, F_{LC}, captures the motivation of the ego vehicle to perform a lane change. This can be further expanded as the motivation to reach a desired speed and to avoid slow traffic ahead. We define this force as follows taking into consideration the

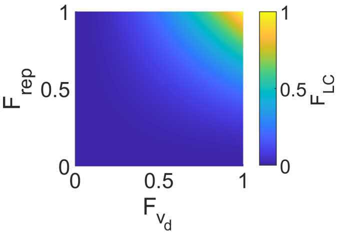


Fig. 6. Lane change force profile. Magnitude of the force depends on the deviation from desired speed and the traffic ahead.

availability of the lane and traffic vehicles:

$$F_{LC}(F_{v_d}, F_{rep}) = G_r F_{v_d}^2 F_{rep}^2 \prod_j^N G_{v,j}, \tag{6}$$

where,

$$G_r = \mathbf{1}_{\{\text{if a lane change is available}\}}$$

and

$$G_{v,j} = \mathbf{1}_{\{\text{vehicle j does not block the target lane}\}}$$

The force is divided into a left lane change component and a right lane change component. This allows for different definitions for each direction in the case that a target lane is preferred. The parameter G_r captures whether the target lane is available without considering the traffic vehicles. $G_{v,j}$ determines whether a traffic vehicle j blocks the target lane. This could be using the relative distance or an external safety module. In this work, we consider the rule-based safety check. Figure 6 visualizes the lane change profile using (6). The force peaks when the vehicle is not following the desired speed and a slow-moving vehicle is present ahead. If the vehicle is following the desired speed, then there is no motivation to perform a lane change. Similarly, if there is no traffic vehicle ahead.

3.3 Temporally-extended driving forces

Predicting the effect of temporally-extended actions, such as a lane change, within an MDP is hard for RL agents. One method of addressing this issue is by integrating temporally-extended information into the state representation. A temporally-extended driving force that captures information useful for lane changing is defined and this predictive force is coined the *lane change risk.*

Consider a vehicle and obstacle model to predict the consequence of future actions for a certain time horizon N such as the point mass model. Then, for the ego vehicle:

$$y(k + \Delta k) = y(k) + v_{ey}\Delta k \tag{7}$$

where v_{ey} is the ego vehicle lateral velocity. Similarly, for the obstacle with the ego vehicle as reference:

$$\begin{aligned} x(k + \Delta k) &= x(k) + (v_x - v_{ex})\Delta k \\ y(k + \Delta k) &= y(k) + v_y \Delta k \end{aligned} \tag{8}$$

where k is the current step in the prediction horizon and Δk is the sampling for the horizon. We assume that the velocities in the longitudinal and lateral directions are fixed for the horizon and defined by the values at the start of the horizon. The lateral ego velocity, v_{ey} is defined by the action path: left or right lane change.

The risk at the current position in the horizon is modeled using a risk potential field similar to Raksincharoensak et al. (2016). Given the (x_o, y_o) position of an obstacle and the (x, y) position of the ego vehicle, the risk potential field is defined as a two-dimensional Gaussian function:

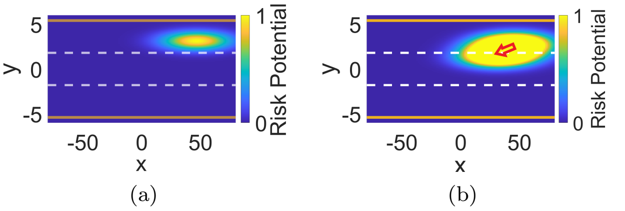


Fig. 7. Risk potential field profile of an obstacle over a horizon of 5 seconds. Left shows the initial step, right shows the final step.

$$U(x, y, x_o, y_o) = e^{-\frac{(x-x_o)^2}{2\sigma_{ox}^2}} e^{-\frac{(y-y_o)^2}{2\sigma_{oy}^2}} \tag{9}$$

where σ_{ox}^2 and σ_{oy}^2 are the longitudinal and lateral variance. This potential function represents a blob that spreads outwards from the center of the obstacle. Figure 7 shows the risk potential profile as defined in (9). The lane change risk is then defined as the sum of risks from all traffic vehicles in the target lane over the prediction horizon. It captures the temporally-extended risk associated with the vehicle shown in Fig. 7. This force is defined for each of the left and right target lanes as follows:

$$\begin{aligned} F_{H,llc}(x_k, y_k) &= \sum_{k=1}^{N} \sum_{j_l} U_{ol}(x_k, y_k, x_{j_l,k}, y_{j_l,k}) \\ F_{H,rlc}(x_k, y_k) &= \sum_{k=1}^{N} \sum_{j_r} U_{or}(x_k, y_k, x_{j_r,k}, y_{j_r,k}) \end{aligned} \tag{10}$$

where j_l, j_r are the traffic vehicle index on the left and right lane, U_{ol} represents the potential function under the assumption that v_{ey} corresponds to a left lane change, and U_{or} under the assumption of a right lane change. It can be interpreted as the risk down-the-line for taking a lateral action at the current position in time.

4. SIMULATION

The simulation is based on the three-lane highway environment in Fig. 2. The vehicle is modeled as:

$$\begin{aligned} x(t + \Delta t) &= x(t) + v_x(t)\Delta t \\ y(t + \Delta t) &= y(t) + v_y(t)\Delta t \\ v_x(t + \Delta t) &= v_x(t) + a_x(t)\Delta t \\ v_y(t + \Delta t) &= v_y(t) + a_y(t)\Delta t \end{aligned} \tag{11}$$

where t is the time index, v is the velocity, a is the acceleration, and Δt is the sampling time. At any instant, the ego vehicle can be surrounded by a maximum of six traffic vehicles which are captured by the affordance indicators. To introduce diverse traffic scenarios, the speed of all vehicles is randomly initialized in the range of $[22, 32]\, m/s$ and up to N_t traffic vehicles are randomly placed in the environment. N_t is chosen to be a uniform random number between 5 and 21. In the training phase, the ego vehicle follows an $\epsilon - greedy$ policy to choose the next action given the current state. Traffic vehicles may randomly perform lane changes taking into account the relative distance and speed to nearby vehicles to avoid a collision. The episode terminates when a collision occurs or the total episode length is reached. In the evaluation

Table 1. Simulation Parameters

Parameter Name	Value
Nominal desired speed v_d	32 m/s
Maximum speed v_{max}	34 m/s
Solid lane, h_i	1.0
Broken lane, h_i	0.5
Lane width l	3.6 m
Road force variance ω_R	0.16
Repulsion force (x, y) variance (σ_x^2, σ_y^2)	(400, 5)
Risk potential (x, y) variance $(\sigma_{ox}^2, \sigma_{oy}^2)$	(400, 0.5)

phase, the trained policies are frozen and the best action is selected based on the driving condition.

The action is a combination of high-level lateral commands {*maintain, change lane left, change lane right*} and longitudinal commands {*accelerate, maintain, brake, hard brake*}. A total combination of 12 actions are possible. A low-level PID controller executes the high-level action given by the agent.

The reward function depends on the deviation from the desired speed, the lane centerline, the distance to the leading vehicle, and the presence of a collision. These reward components are calculated as follows:

$$\begin{aligned} r_v &= e^{-\frac{(v_{e_x}-v_{des})^2}{10}} - 1, \\ r_y &= e^{-\frac{(d_{e_y}-y_{des})^2}{10}} - 1, \\ r_x &= \begin{cases} e^{-\frac{(d_{lead}-d_{safe})^2}{10 d_{safe}}} - 1 & \text{if } d_{lead} < d_{safe}, \\ 0 & \text{otherwise,} \end{cases} \\ r_{col} &= -2 \end{aligned} \tag{12}$$

where d_{lead} is the distance to the lead vehicle and d_{safe} is the minimum safe distance. The total reward becomes

$$R = r_v + r_y + r_x + r_{col} \tag{13}$$

Two variations for the agent state space are considered:

- *DF Model:* Five driving forces without temporally-extended driving forces, $S_t = \{F_{vd}, F_{RA}, F_{rep}, F_{LC_{left}}, F_{LC_{right}}\}$
- *DF Hazard:* The DF Model is complemented with the lane change risk forces, $F_{H,llc}$ and $F_{H,rlc}$.

5. EXPERIMENTS AND RESULTS

A DDQN agent is trained to test the proposed state representation in the highway environment. An agent is trained for each state space variation for 10,000 episodes with a discount factor of 0.9. The exploration, ϵ, is annealed from 1.0 to 0.2 over 7,000 episodes and then kept fixed for the rest of the training. Each episode consists of 200 steps. The Q-network is a fully-connected network with two hidden layers with 100 neurons each and Leaky ReLu activation. The Adam optimizer is used with a learning rate of $1e^{-4}$. Simulation parameters are given in Table 1.

Effect of temporally-extended driving forces The effect of the temporally-extended driving forces is investigated for three different prediction horizons under the DF Hazard model: (i) 5 seconds (*DF Hazard Short*) (ii) 10 seconds (*DF Hazard Moderate*) (iii) and 15 seconds (*DF Hazard Long*). DF Hazard Short represents the average lane change duration, see Toledo and Zohar (2007).

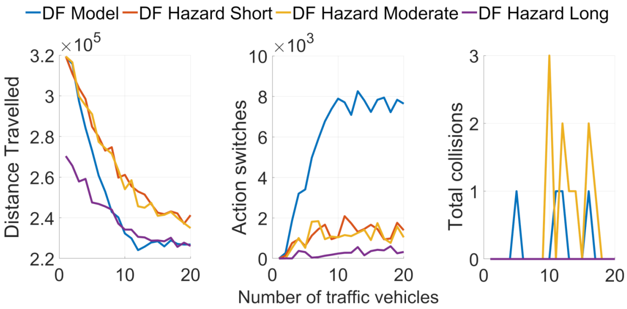


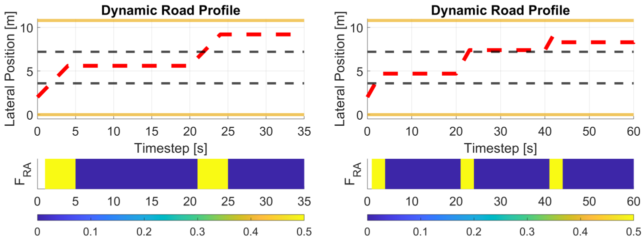
Fig. 8. Evaluation of total number of collisions, average distance travelled, and action-switching under different traffic conditions for 50 episodes.

All agents are evaluated under different traffic conditions by varying the number of traffic vehicles from 1 to 20 vehicles for 50 episodes each. The simulation results given in Fig. 8 show that incorporating temporally-extended information in driving forces results in a longer distance travelled and safer decision-making with careful selection of the prediction horizon. The temporal information encoded leads to a drop in collisions for the case of lane changing. Note that the choice of the prediction horizon can also deteriorate safety performance. The best prediction horizon was found to be 5 seconds, which corresponds to the average duration of a lane change. Action switching is defined as switching between accelerate and brake; or left and right lane change. The action switching exhibited by the models show that the incorporation of temporally-extended information results in a better quality of actions.

Dynamic properties of the state representation The ability of the driving forces to generalize the RL policy to new road profiles and behaviors without further training is investigated. Fig. 9 shows that manipulating the road profile force can achieve a desired lateral position. Lane changes are induced by increasing the magnitude of the force at the current position. In Fig. 9b, the agent exhibits the ability to follow new road profiles for the same policy. In this case, narrower lanes are assumed. Next, the ability to induce desired longitudinal behaviors through manipulation of the velocity following force is shown, see Fig. 10. The agent achieves two desired longitudinal behaviors: (i) following different desired speeds (ii) inducing intermittent step accelerations. This further shows that the formulation and interpretation of the driving forces can be used to generalize a trained, frozen policy over unseen scenarios in an online fashion, which has adaptability implications.

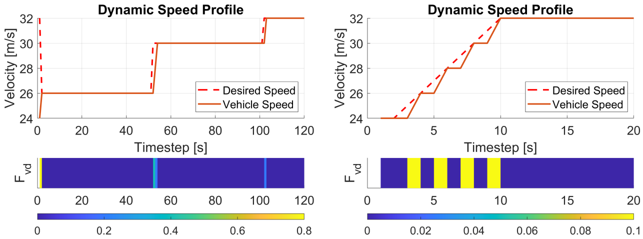
6. CONCLUSION

In this paper, the driving forces as a dynamic, low-dimensional, interpretable state representation were presented. It was shown that the formulation can be leveraged to generalize the policy of the RL agent to unseen scenarios and behaviors. Furthermore, the formulation of the driving forces is general enough to be applied to arbitrary road layouts, traffic vehicles, and to incorporate temporally-extended information. Temporally-extended information in the state representation was shown to be important for achieving better decision making. This work shows that the proposed driving forces as a state representation can be



(a) (b)

Fig. 9. Different lateral behaviors beyond what was seen during training are induced through manipulation of the road profile force.



(a) (b)

Fig. 10. Different longitudinal behaviors are induced through manipulation of the velocity following force.

used to improve the issue of explainability, interpretability, and real-time adaptation of RL agents. In future work, we plan to further explore this state representation for safety mitigation of an RL agent.

REFERENCES

Bewley, A., Rigley, J., Liu, Y., Hawke, J., Shen, R., Lam, V.D., and Kendall, A. (2019). Learning to drive from simulation without real world labels. In *2019 International conference on robotics and automation (ICRA)*, 4818–4824. IEEE.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J., et al. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Chen, C., Seff, A., Kornhauser, A., and Xiao, J. (2015). Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE international conference on computer vision*, 2722–2730.

Chen, J., Li, S.E., and Tomizuka, M. (2021). Interpretable end-to-end urban autonomous driving with latent deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*.

Chen, J., Yuan, B., and Tomizuka, M. (2019). Model-free deep reinforcement learning for urban autonomous driving. In *2019 IEEE intelligent transportation systems conference (ITSC)*, 2765–2771. IEEE.

Codevilla, F., Müller, M., López, A., Koltun, V., and Dosovitskiy, A. (2018). End-to-end driving via conditional imitation learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 4693–4700. IEEE.

Codevilla, F., Santana, E., López, A.M., and Gaidon, A. (2019). Exploring the limitations of behavior cloning for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9329–9338.

Fredette, D. and Özguner, Ü. (2016). Swarm-inspired modeling of a highway system with stability analysis. *IEEE Transactions on Intelligent Transportation Systems*, 18(6), 1371–1379.

Han, T., Jing, J., and Özgüner, Ü. (2019). Driving intention recognition and lane change prediction on the highway. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, 957–962. IEEE.

Kendall, A., Hawke, J., Janz, D., Mazur, P., Reda, D., Allen, J.M., Lam, V.D., Bewley, A., and Shah, A. (2019). Learning to drive in a day. In *2019 International Conference on Robotics and Automation (ICRA)*, 8248–8254. IEEE.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540), 529–533.

Müller, M., Dosovitskiy, A., Ghanem, B., and Koltun, V. (2018). Driving policy transfer via modularity and abstraction. *arXiv preprint arXiv:1804.09364*.

Nageshrao, S., Tseng, H.E., and Filev, D. (2019). Autonomous highway driving using deep reinforcement learning. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2326–2331. IEEE.

Osiński, B., Jakubowski, A., Ziecina, P., Miłoś, P., Galias, C., Homoceanu, S., and Michalewski, H. (2020). Simulation-based reinforcement learning for real-world autonomous driving. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 6411–6418. IEEE.

Prakash, A., Behl, A., Ohn-Bar, E., Chitta, K., and Geiger, A. (2020). Exploring data aggregation in policy learning for vision-based urban autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11763–11773.

Raksincharoensak, P., Hasegawa, T., and Nagai, M. (2016). Motion planning and control of autonomous driving intelligence system based on risk potential optimization framework. *International journal of automotive engineering*, 7, 53–60.

Sauer, A., Savinov, N., and Geiger, A. (2018). Conditional affordance learning for driving in urban environments. In *Conference on Robot Learning*, 237–252. PMLR.

Toledo, T. and Zohar, D. (2007). Modeling duration of lane changes. *Transportation Research Record*, 1999(1), 71–78.

Toromanoff, M., Wirbel, E., and Moutarde, F. (2020). End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7153–7162.

Van Hasselt, H., Guez, A., and Silver, D. (2016). Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Yurtsever, E., Lambert, J., Carballo, A., and Takeda, K. (2020). A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8, 58443–58469.