Transportation Research Part C 125 (2021) 103008

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning

Xuan Di [a,b,*], Rongye Shi [a]

[a] *Department of Civil Engineering and Engineering Mechanics, Columbia University, United States*
[b] *Center for Smart Cities, Data Science Institute, Columbia University, United States*

A R T I C L E I N F O

Keywords:
Artificial intelligence (AI)
Autonomous vehicle (AV) control
Mixed autonomy

A B S T R A C T

This paper serves as an introduction and overview of the potentially useful models and methodologies from *artificial intelligence* (AI) into the field of *transportation engineering* for autonomous vehicle (AV) control in the era of mixed autonomy when AVs drive alongside human-driven vehicles (HV). It is the first-of-its-kind survey paper to comprehensively review literature in both transportation engineering and AI for mixed traffic modeling. We will discuss state-of-the-art applications of AI-guided methods, identify opportunities and obstacles, and raise open questions. We divide the stage of AV deployment into four phases: the pure HVs, the HV-dominated, the AV-dominated, and the pure AVs. This paper is primarily focused on the latter three phases. Models used for each phase are summarized, encompassing game theory, deep (reinforcement) learning, and imitation learning. While reviewing the methodologies, we primarily focus on the following research questions: (1) What scalable driving policies are to control a large number of AVs in mixed traffic comprised of human drivers and uncontrollable AVs? (2) How do we estimate human driver behaviors? (3) How should the driving behavior of uncontrollable AVs be modeled in the environment? (4) How are the interactions between human drivers and autonomous vehicles characterized? We also provide a list of public datasets and simulation software related to AVs. Hopefully this paper will not only inspire our transportation community to rethink the conventional models that are developed in the data-shortage era, but also start conversations with other disciplines, in particular robotics and machine learning, to join forces towards creating a safe and efficient mixed traffic ecosystem.

1. Introduction

We are transitioning into a big data era from a data-shortage era, thanks to the popularity of ubiquitous sensors, such as GPS (Di et al., 2017; Liao et al., 2018; Shou and Di, 2018; Meinrenken et al., 2020), blue tooth (Allström et al., 2014), and smart phones (Herrera et al., 2010; Shou et al., 2020a). Autonomous vehicles (AV) or Level-5 automated vehicles, mounted with sensors like camera and LiDAR, will potentially provide exploding volumes of transportation data (SAS, 2015). While moving from a data-sparse to a data-rich era, we, the transportation community, urgently need a methodological paradigm shift from physics-based models to *artificial intelligence* (AI)-guided methods, which can project future traffic dynamics comprised of AVs driving alongside human-driven vehicles

* Corresponding author at: Department of Civil Engineering and Engineering Mechanics, Columbia University, United States.
E-mail address: sharon.di@columbia.edu (X. Di).

https://doi.org/10.1016/j.trc.2021.103008
Received 8 July 2020; Received in revised form 25 December 2020; Accepted 1 February 2021
Available online 11 March 2021

(HV) and assist in socially optimal policy-making. **Physics-based** (or rule-based (Zhou and Laval, 2019)) models refer to all the scientific hypotheses about the movement of cars or traffic flow, including traffic models on micro-, meso-, and macro-scale; while **AI-guided** methods refer to cutting-edge models that mimic human intelligence, leveraging deep neural networks, reinforcement learning, imitation learning, and other advanced machine learning methods.

This paper serves as an introduction and overview of the potentially useful models and methodologies from *AI* into the field of *transportation engineering* in the era of mixed autonomy. We will discuss the state-of-the-art applications of AI-guided methods to AV controls, identify opportunities and obstacles, and raise open questions, It is the first-of-its-kind survey paper to comprehensively review literature in both transportation engineering and AI for mixed traffic modeling. Hopefully this paper will not only inspire our transportation community to rethink the conventional models that are developed in the data-shortage era, but also start conversations with other disciplines, in particular robotics and machine learning, to join forces towards creating a safe and efficient mixed traffic ecosystem.

Vehicle automation consists of six levels, according to SAE *J*3016 standard (International Standard J3016, 2016): from Level-0 "no automation" to Level-5 "fully automation". Car manufacturers target different automation levels. Advanced driver assistance systems (ADAS) in commercial production vehicles, such as Tesla AutoPilot (Tesla, 2018), are Level-1 or 2 systems. Waymo driverless cars (Waymo, 2020) belong to Level-4 or Level-5 automation. Level-5 automated vehicles, fully automated vehicles, autonomous vehicles, self-driving cars, driving in autonomous mode, are interchangeable. Most academic papers on driving policies may not explicitly state which level of driving automation their models work on. But by default one level of automated systems is usually depicted by one particular set of driver models. Level-1 or 2 AV controller is usually modeled as linear or nonlinear controllers, the parameters of which are characterized using stability analysis. Level-4 or 5 AVs are commonly modeled using AI methods, including deep learning (DL), reinforcement learning (RL), and game theory. The mapping between driver models and automation levels is summarized in the second column in Table 2.

Vehicles' driving choices contain three levels: operational level (including pedal and brake control, turn signal), tactical level (including lane-changing, lane-keeping), and strategic level (including routing). This paper is mainly focused on the operational and tactical controls of AVs. The operational and tactical controls can be further categorized into longitudinal control (i.e., car-following, lane-keeping) and lateral control (i.e., lane-change). Longitudinal control has been studied in various scenarios, including: platooning (Gong et al., 2016; Zhou et al., 2017b; Wei et al., 2017; Li et al., 2018d), speed harmonization (Ma et al., 2016; Malikopoulos et al., 2018; Arefizadeh and Talebpour, 2018), longitudinal trajectory optimization (Wei et al., 2017; Li et al., 2018b), and eco-approach and departure at signalized intersections (Altan et al., 2017; Hao et al., 2018; Yao et al., 2018). Most of the existing studies are limited to a single AV navigating along a highway or dense with human drivers, or all AVs dominate the road with negligible interactions with HVs (Katrakazas et al., 2015).

1.1. Modeling complexity

To date, the vast majority of existing research has focused - perhaps unsurprisingly - on two polar scenarios, where either a single AV navigates in an ecosystem dense with human drivers, or a platoon of AVs move along a highway, with negligible interaction with human-controlled counterparts. Much less attention has been accorded to the far more realistic, yet challenging transition path

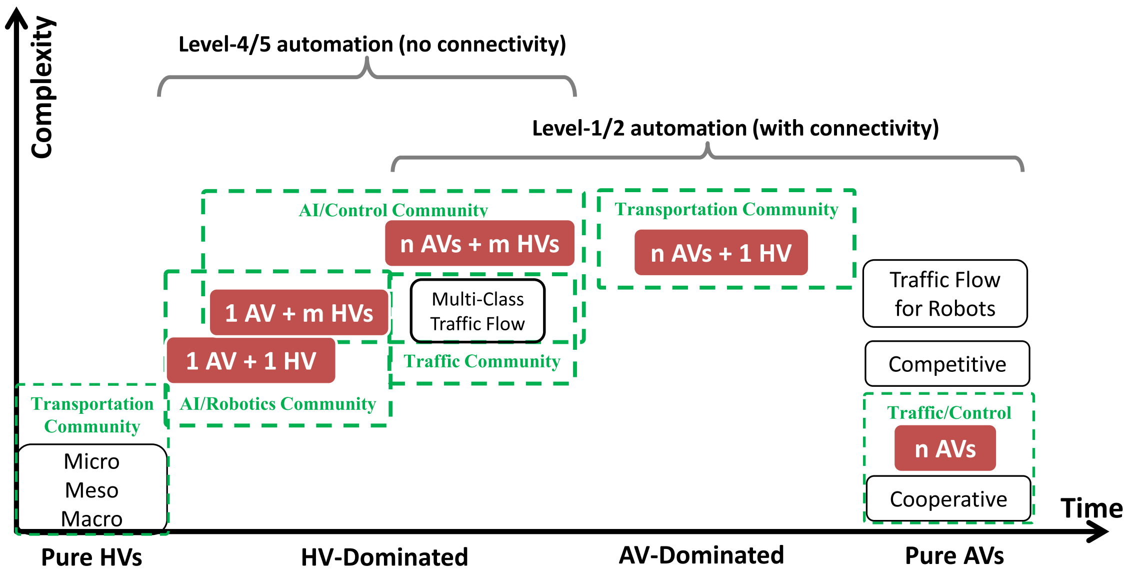


Fig. 1. Modeling complexity at each stage (Four research questions will be answered through models of each phase. White boxes indicate the model types for each phase. Research communities are enclosed by dashed green boxes.).

between these two scenarios. However, it is precisely this hybrid human–machine space that deserves our concerted attention now, so-called "mixed autonomy" (Wu et al., 2017b).

We divide the stage of AV deployment into four phases: the pure HVs, the HV-dominated, the AV-dominated, and the pure AVs. This paper focuses on the latter three phases. Fig. 1 demonstrates the modeling complexity for each phase. It is most challenging to model the *HV-dominated* and *AV-dominated* phases, in other words, mixed autonomy. This is an understudied phase due to the unknown and complex interactions among different types of vehicles. We further divide mixed autonomy by the relative proportion of AVs and HVs using the following notions (indicated in red boxes in Fig. 1):

- 1 AV + 1 HV: one AV interacts with one HV;
- 1 AV + m HVs: one AV navigates the HV-dominated traffic environment;
- n AVs + m HVs: multiple AVs navigate the HV-dominated traffic environment; n AVs + 1 HV: multiple AVs interact with one HV in the AV-dominated traffic environment; It is one special case of n AVs + m HVs.
- n AVs: a pure AV market where all vehicles are replaced by AVs. Accordingly, AVs interact with one another.

Remark 1.1. In the process of preparing this paper, we discover another survey paper (Zhou and Laval, 2019) on longitudinal control of AVs and their impact on traffic congestion. The main difference is that Zhou and Laval (2019) focus on training a single AV on an empty highway or with a few HVs surrounded (corresponding to 1 AV + m HVs), while this paper reviews a broader literature in mixed autonomy.

In Fig. 1, the community associated with each phase is enclosed by a green dotted box. The transportation community has primarily focused on modeling the pure HVs, the AV-dominated, and the pure AVs, while AI and control communities are more focused the HV-dominated phase where a single AV or a finite number of AVs navigate the traffic environment. We also indicate the modeling assumptions about automation levels and vehicle connectivity on the upper part of the figure. The models developed for the HV-dominated phase by the AI community tend to assume Level-4 or 5 automation without connectivity, while those developed for the AV-dominated and the pure AV phases, along with specific platooning scenarios in the HV-dominated phase, tend to use Level-1 or 2 automation with connectivity. Below we will further elaborate how these communities diverge in autonomous control modeling.

1.2. Divergence in the communities

Researchers from the transportation community and the robotics community refer to same quantities with different terminologies (Fernandez Fisac, 2019). Here we present all the relevant terminologies across the communities in Table 1.

When it comes to the autonomous driving controller design, these two communities take two different paths. First of all, these two communities share different goals. Transportation researchers aim to understand the influence of AVs on the transportation system performance (from the SYSTEM perspective), such as traffic congestion (Zhou and Laval, 2019), while robotics researchers are primarily focused on the development of optimal driving policies for AVs to learn and adapt in a stochastic environment (from the VEHICLE perspective). As a consequence, these communities investigate problems on different scales.

The robotics community aims to design AV controllers with human-in-the-loop. In particular, researchers on human-cyber-physical systems design AVs that actively influence human drivers through mutual interactions, in order to achieve efficient driving. Their impact on system level performance remains unknown though. The transportation systems community aims to understand the influence of AVs on the transportation system performance, including travel time, traffic delay, traffic safety, and emissions. Because multi-class microscopic models are not scalable for the varying topology of mixed vehicle types, researchers are more focused on modeling mixed traffic using the multi-class approach on a macro scale (Talebpour and Mahmassani, 2016; Levin and Boyles, 2016; Melson et al., 2018; Chen et al., 2016; Chen et al., 2017; Kockelman, 2017). On road networks, static (Chen et al., 2016; Chen et al., 2017) or dynamic (Dresner and Stone, 2007; Levin and Boyles, 2015; Levin and Boyles, 2016; Patel et al., 2016; Melson et al., 2018)

Table 1
Terminologies across communities (partly adapted from Fernandez Fisac (2019)).

Transportation	Control	AI	Game Theory
Traffic	System	Environment	Game
Traffic evolution	Dynamics	Transition	Dynamics
Traffic state	State space	State	State
CAV	CACC	AV	AV
Car-driver unit	Controller	Agent	Player
Vehicle control	Control	Action	Play
Vehicle control	Control	Action	Play
Vehicle control law	Control law	Policy	Strategy
Vehicle control objective	Cost	Reward	Payoff
Car-following	Longitudinal control	-	-
Lane-change	Lateral control	-	-
Driving behavior	Driver model	Driver intent	Rationality
Traffic outcome	Optimal control	Optimal policy	Equilibrium

Table 2
Mapping categories in the car-following scenario by communities.

Models			Level of automation	Goal	Input features	Behavioral difference				Community (Sample references)	Pros	Model selection criteria
						information	speed or acceleration	following distance	reaction time			
Physics-Based	HV	CFM	Level-0	capture traffic characteristics	$v_{i-1} - v_i, h_i$	less, local	continuously changing	long	long (∼ 1.5 sec)	Transportation (Newell, 1961; Gipps, 1981; Treiber et al., 2000; Kesting et al., 2010)	data-efficient, interpretable, generally mathematically tractable	small available data, limited input features
	AV	Linear or nonlinear controllers	Level-1/2 ADAS	string stability, improved system performance (capacity, congestion, safety), robust to uncertainty	linear $h_i - h_i^*$, $s_i - s_i^*$, nonlinear $v_{i-1} - v_i, h_i$, others' acceleration	rich, global	set from different speed options, heterogeneous acceleration rate	short	short (0.5 s) or heterogeneous	Transportation & control (Schakel et al., 2010; Naus et al., 2010; Ploeg et al., 2011; Milanés et al., 2014; Milanés and Shladover, 2014; Orosz et al., 2010; Jin and Orosz, 2014; Qin and Orosz, 2017; Chen et al., 2019a)		
AI-Based	HV	deep learning	-	human-like performance	(time series of) $v_{i-1} - v_i$, h_i, reaction delay	local	unstable behavior, including asymmetric driving behavior, traffic oscillation			Transportation (Panwai and Dia, 2007; Khodayari et al., 2012; Zhou et al., 2017a; Huang et al., 2018; Zhu et al., 2018b)	amenable to massive data, adaptive to uncertain and volatile traffic environments, and are generalizable to generic driving tasks	big, high-dimensional data
	AV	Deep (Reinforcement) Learning	Level-4/5 (e.g., Waymo)	optimal decision-making for safety and efficiency	camera images	local	end-to-end nonlinear, unintepretable controllers			Robotics (Lillicrap et al., 2015; Zhang et al., 2016; Sallab et al., 2017; Perot et al., 2017; Jaritz et al., 2018; Gu et al., 2020)		

traffic assignment models are developed to capture AVs' intersection coordination and routing behavior. Those models on a macroscopic scale may lack detailed interpretation of how different types of vehicles interact at the micro scale.

Second, with different goals, different AV decision-making frameworks have been employed. Academic researchers have to make various assumptions to implement AV components in their models or simulations, because real-world AVs are primarily developed and tested by private companies which are not willing to reveal how the existing AV test fleets on public roads are actually programmed to drive and interact with other road users. Accordingly, different driving models lead to different driving behavior and traffic patterns.

The transportation and the control communities assume AVs are particles or fluids following the physics-based models, including both the micro- and macroscopic traffic models that were originally developed for human drivers, and tailors AV behavior on that of HVs in which AVs are essentially human drivers but react faster, "see" farther, and "know" the road environment better. For instance, a majority of studies equate AVs to ADAS or commercial semi-autonomous functionality (e.g., Tesla's Autopilot). Accordingly, models of the dynamic response of these systems are used as AV driving models (Naus et al., 2010; Qin and Orosz, 2013; Shladover et al., 2015; Delis et al., 2016; Zhou et al., 2020). Otherwise AVs are treated like humans but with modified parameters (Schakel et al., 2010; Naus et al., 2010; Ploeg et al., 2011; Milanés et al., 2014; Milanés and Shladover, 2014; Jin and Orosz, 2014). Also, because these automated driving systems are only enabled in designated traffic scenarios, such as platooning, control models are thus constrained to these scenarios, not applicable to a generic traffic environment.

The robotics community, on the other hand, treats AVs like AI robots or agents who can continuously explore environments and exploit optimal actions (Sadigh et al., 2016b; Liu and Tomizuka, 2016). When the environment is observable, AVs select optimal strategies based on predefined reward functions in cooperative or non-cooperative games. RL, a cutting-edge learning paradigm initially developed for optimal control of robotics, has been naturally deployed for AVs. In this framework, human drivers are modeled as part of the environment where AVs move and explore, using either an Markov decision process (Mukadam et al., 2017) or a simulated model-free environment (Wu et al., 2017b; Wu et al., 2017d; Wu et al., 2017a; Wu et al., 2018; Kreidieh et al., 2018b).

The aforementioned modeling difference arises from the fundamentally different assumptions of vehicle automation levels. The transportation and control community is primarily focused on Level-1 or 2 automation (International Standard J3016, 2016), which is also coupled with vehicular connectivity. The most studied scenario is collaborative platooning on highways, in which connected and automated vehicles (CAVs) drive in a platoon to optimize certain systematic performance measures. In contrast, the robotics and AI community is mainly focused on Level-5 AVs, mostly without communication with neighboring vehicles, which are operated in a variety of generic traffic scenarios such as lane-keeping, lane-change, merging, and crossing. Such a difference across communities originates from their radically different beliefs of the feasible technological path: the former takes a gradual, incremental step to automation, from Level-1 to 2, and complements automation with connected vehicle technologies, which is pushed forward constantly by the U.S. Department of Transportation (USDOT). In contrast, AV tech companies exemplified by Waymo LLC take a leap from Level-0 directly to the full autonomy and do not rely on any communication infrastructure.

Remark 1.2. We would like to point out that automation and connectivity are two distinct technologies. AVs may or may not have connectivity, while connected vehicles may or may not have automation. *Connected vehicle* refers to the vehicular technology that enables users to communicate with one another within surface transportation ecosystems (USDOT, 2019). Vehicle-to-vehicle (V2V) (or vehicular ad hoc networks (VANETs) used in the communication society) and vehicle-to-infrastructure (V2I) communication are the most studied scenarios. The enabling mechanisms include dedicated short range communication (DSRC) and cellular communication like 5G. The DSRC infrastructure has primarily been established by USDOT on pilot testbeds (USDOT, 2020) instead on a large real-world scale, while cellular networks are primarily rolled out by private companies. The existing Adaptive Cruise Control (ACC) systems (i.e., Level-1) does not posses the capability of communication, but extensive studies have been focused on communication-enabled ACC, which is Connected Adaptive Cruise Control (CACC). Otherwise research on Level-4 or 5 automation does not generally assume connectivity.

Third, behavioral modeling of HVs and AVs is different. The transportation community differentiates HVs and AVs with different models: HVs tend to exhibit unstable, stochastic behavior, while AVs can overcome traffic instability with stable controller design. In contrast, because the robotics community does not account for collective traffic patterns, they believe human drivers are intelligent for AVs to emulate. Thus those studies do not usually distinguish between HVs and AVs. Instead, both HVs and AVs are modeled as AI agents.

Fourth, there is a discrepancy in how the interactions between AVs and HVs are modeled. The transportation community does not formalize how AVs interact with HVs in driving processes. On the microscopic level, car-following models (CFM) are applied that implicitly encode how one follows its immediate or far upstream leaders. On the macroscopic level, usually multi-class traffic models are adopted, which do not define micro level interactions in detail. The robotics community tries to explicitly design microscopic interactions between one or a few AVs and one or multiple HVs.

In summary, these two communities formulate the autonomous driving control problem with different goals, employ different models, and use different algorithms on different datasets.

1.3. Physics- v.s. AI-based traffic models

Studies on physics-based models have dated back to 1930s (Greenshields et al., 1935) and have made breakthrough in modeling and perdiction of real-world traffic phenomena (Zhou and Laval, 2019). Thus the value of physics-based models should not be undermined, even though the focus of this paper is AI-based models. Physics-based models are efficient to calibrate due to the existence of only a few parameters, are easy to interpret with all the physical terms, and are generally mathematically tractable for stability analysis and safety guarantee. Therefore, they are particularly useful when the available driving data size is small and the knowledge of input features is limited, and when the research objective is to develop stability-based controllers that will be deployed in designated traffic scenarios (e.g., platooning).

Physics-based models, however, suffer from major shortcomings (Chen et al., 2019c). First, different driving tasks are described by different models. For example, car-following and lane-change behaviors are usually modeled separately. The model developed for one behavior has to be redesigned manually for different scenarios and tasks. Second, the predefined motion heuristics usually make strong assumptions about driving behaviors using a small set of parameters, which may not be able to capture human's strategic planning behaviors and may not generalize well to diverse driving scenarios in a highly interactive environment. Third, these models highly rely on existing scientific hypotheses and physics principles, which might be proposed and tested in the era when limited traffic data were collected and might likely capture partial information of real-world traffic. These issues have become more outstanding in the big data era. Thus, we need a transformative not progressive approach to address the above issues, and we believe AI-based models is one promising option.

AI-based models are amenable to massive data, adaptive to uncertain and volatile traffic environments, and are generalizable to diverse driving tasks and generic traffic environments. They are extremely powerful to extract useful information from high-dimensional complex driving data. AI-based models are suitable when driving data are high volume and high dimension and when the research goal is to train AVs to achieve human-like performances in various traffic environments or to make optimal decision-making to achieve safety and efficiency. However, AI is not "a hammer for every nail". In traffic models, it could fail to respect physical constraints and lack interpretability, compared to physics-based models. Because the field of AI is still evolving, hopefully some of these problems will be addressed in the future and some promising directions will be provided in Section 10.

Remark 1.3. We are not in a position to judge what stream of methods are more advantageous over the other, provided that both streams are still under development. Rather, we try to compare these two methodological paradigms from a neutral point of view and offer some insights into under what conditions should each method be considered.

1.4. Driving policy mapping: Overview

To further demonstrate the divergence of two communities, we provide a generic mathematical form of driving policy mappings. Denote $\mathscr{S}\subseteq\mathbb{R}^m_+$ and $\mathscr{A}\subseteq\mathbb{R}^n_+$ as state and action spaces, respectively, where $m \in \mathbb{N}_+$ and $n \in \mathbb{N}_+$ are the dimension of state and action vectors, respectively. A driving behavior model, or a driving control or policy is a mathematical mapping parameterized by θ, denoted as π_θ, from states $\mathbf{s} \in \mathscr{S}$ (i.e., observations of the traffic environment) to actions $\mathbf{a} \in \mathscr{A}$ (i.e., acceleration and steering angle):

$$\pi_\theta : \mathbf{s} \rightarrow \mathbf{a} \tag{1.1}$$

The mapping π_θ can be obtained either from the first principles (physics based) or from data. We summarize a variety of mapping forms for longitudinal control policies in Table 2. The reason we select longitudinal control here is that they are the most extensively studied models and there are a lot of recent effort in bringing in machine learning to such behavior modeling. The driving policy mapping is categorized into physics-based and AI-based. Physics-based mapping can be characterized by mathematical formulas, while AI-based mapping is usually represented by a variety of machine learning models. Within each mapping type, we also compare how HVs and AVs are modeled differently. As pointed out before, when both HVs and AVs are modeled by physics-based mappings, the main difference between HVs and AVs lie in the parameters, reflecting that AVs "sense" better, "see" farther, and "react" faster. The AI-based mappings assume there exists a complex, highly nonlinear mapping from driving perception to machine activation. In these mappings, the difference of HVs and AVs may not be so notable because the goal is to train AVs to exhibit human-like performance. In the second last column, we list communities along with sample references for each mapping category. Most of the listed references may be revisited in the rest of the paper. In the last column, we highlight the pros of each category of models and the recommended criteria of model selection, summarized in subSection (1.3).

Physics-based driving models, widely used by the transportation and control communities, assume each unit behaves like an automated particle or automaton, within which human cognitive process and the machine's mechanical dynamics are highly simplified. Car-following is the most studied driving behavior, which is categorized into three types: microscopic driving models, mesoscopic traffic flow models, and macroscopic traffic flow models.

Microscopic models characterize the movement of individual vehicles with one's position, speed, and acceleration, assuming that cars select their driving velocity and acceleration dynamically based on the following distance from their immediate leader, speed difference, and/or other features. The mathematical tool is ordinary differential equation. Some of the widely used microscopic car-following models include Newell (Newell, 1961), Gipps' model (Gipps, 1981), IDM (Treiber et al., 2000; Kesting et al., 2010), and OVM (Orosz et al., 2010; Jin and Orosz, 2014; Qin and Orosz, 2017).

Macroscopic models treat traffic as a continuum flow, which is characterized by aggregate traffic density and velocity. The evolution of traffic density and velocity are determined using partial differential equation. Popular traffic flow models include LWR (Lighthill and Whitham, 1955; Richards, 1956), PW (Payne, 1971), and ARZ models (Aw and Rascle, 2000).

Mesoscopic models, linking microscopic to macroscopic models, characterize traffic flows using a probability distribution function of vehicle velocities (van Wageningen-Kessels et al., 2015). The most popular mesoscopic models are gas-kinetic models (Hoogendoorn, 1999; Hoogendoorn and Bovy, 2001; Hoogendoorn and HL Bovy, 2003). Since mesoscopic models lack clear physical interpretation and cannot be directly applied in simulations, they are not as well-studied as other models by the transportation community. By far they mainly serve as a mathematical tool to derive macroscopic models from corresponding microscopic driving behaviors.

1.5. AI for decision-making of AVs

Instead of hypothesizing explicitly how AVs would drive, we believe the futuristic AVs should be designed to act as rational, utility-optimizing agents that play best strategies at each level of driving choices. By doing so, it would allow AVs to react according to the impending traffic situations and closely mimic human drivers' intelligence. However, the major advantage AVs will have over human driver is its ability to access the situation promptly with a better set of information, and thereby enable AVs to react in an optimal way compared to a human driver. Natural traffic experiments are, however, costly and highly risky to perform. We thus seek an innovative *AI-guided* methodological framework for complex multi-agent learning and adaptation.

Despite a significant amount of machine learning efforts given to computer vision, the intelligence of AVs lies in its optimal decision-making at the stage of motion planning. We believe the key to empowering AVs' driving intelligence is AI or even a broader area "Artificial general intelligence" (AGI) (Ramamoorthy and Yampolskiy, 2018). We are seeing a growing number of studies that have employed AI methods to discover humans' driving behaviors, including deep learning (Tanaka, 2013; Colombaroni and Fusco, 2014; Zhou et al., 2017a; Wang et al., 2018; Zhu et al., 2018a), reinforcement learning (VanderWerf et al., 2001), and imitation learning (Kuefler et al., 2017; Bhattacharyya et al., 2018). More recently, many attempts also focus on human behavior prediction, such as lane changing (Kumar et al., 2013; Woo et al., 2017; Wei et al., 2019; Shou et al., 2020c), merging (Rios-Torres and Malikopoulos, 2016; Bevly et al., 2016), and stop behavior (Kumagai and Akamatsu, 2006), to predict with high confidence when a human would change lanes. However, applications of AI to the decision-making processes of AVs are still emerging and remain understudied.

Game theory, a mature field for modeling strategic interactions of rational players, has empowered intelligence of multiple interacting machines and is revolutionizing the field of AI (Tennenholtz, 2002). Fortunately, we have seen a gradual convergence in the control, transportation, and AI communities that have employed game-theoretic models to design algorithmic decision-making processes for AVs (Yoo and Langari, 2012; Yoo and Langari, 2013; Kim and Langari, 2014; Talebpour et al., 2015; Yu et al., 2018; Huang et al., 2019; Huang et al., 2020a; Huang et al., 2020b). We also believe gaming traffic would be a key feature of future AVs to strategically interact with and navigate through a complex traffic environment.

1.6. Organization of the paper

The remainder of the paper is organized as follows: In Section 2, we will provide a general problem statement for AV control in mixed traffic, along with the existing knowledge gaps. We will then examine the existing models and methods for AV control in Sections 3–6. Section 7 presents the methods and models of human and autonomous driving policy learning, respectively. Section 9 summarizes all the models that have been reviewed. In Section 10, we present the challenges and insights into modeling the mixed traffic with AI methods and provide potential research areas.

2. AI-guided driving policy learning for AVs

2.1. Multi-vehicle systems (MVS) with mixed-autonomy

A mixed traffic system is comprised of a large number of networked intelligent agents, including AVs, human drivers, and other road users such as pedestrians and cyclists. They dynamically select driving actions while interacting with the traffic environment. Their actions are interdependent in the sense that one's driving action depends on others', via either coupled reward functions, the common traffic environment state, or the action constraints. Due to this inter-agent coupling, the mixed transportation system is a multi-agent system (MAS) - a widely used term in the control and robotics community. Specifically, we call it a "multi-vehicle system (MVS)." Below we will present a high-level AV control problem of MVS in an abstract manner.

Definition 2.1. (AV control problem statement of MVS with mixed-autonomy.) In a multi-vehicle system with mixed traffic, there are N controllable AVs indexed by $n \in \{1, 2, \ldots, N\}$ driving along a stretch of a road, with initial states $s_0^{(1)}, \ldots, s_0^{(N)}$. The n^{th} car aims to select a sequence of optimal driving controls (including acceleration and/or steering angle) in discrete time intervals (i.e., $a_1^{(n)}, \ldots, a_T^{(n)}$) or continuous time (i.e., $a^{(n)}(t), t \in [0, T]$) over a predefined planning horizon $[0, T]$. Within this generic setting, we can define how the AV control problem is formulated and how the state of each car is updated. We will present the discrete-time version throughout the paper, but it can easily be generalized to the continuous-time version.

- **Control optimization**:

  The AV control problem of MVS is to establish a driving policy mapping $\pi_\theta(\cdot)$ parametrized by θ from states $\mathbf{s} = \{s_t^{(n)}\}_{t \in [0,T]}^{n=1,\ldots,N}$ to actions $\mathbf{a} = \{a_t^{(n)}\}_{t \in [0,T]}^{n=1,\ldots,N}$.

  Depending on the selection of control models, such a mapping can be specified as a dynamical system or an optimal control problem or a model predictive control (MPC) (by the transportation and control community), a game-theoretic model (by the robotics community), or an AI-based model (by the robotics and control community).

- **State update and vehicle motion**:

  At time t, given a selected control $\mathbf{a}_t^{(n)}$, car n updates its state $\mathbf{s}_{t+1}^{(n)}$ based on a dynamical system:

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{a}_t; \mathbf{s}_{t-1}, \mathbf{a}_{t-1}; \cdots). \tag{2.1}$$

This mapping can be specified by a deterministic kinematic state-space model (mostly by the transportation and control community) or as a stochastic dynamic characterized by a transition matrix, for instance $\mathscr{P}(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$ in a Markov decision process, or a model-free environment (mostly by the robotics and control community).

The kinematic model computes the change of vehicles' positions (i.e., first-order), velocities (i.e., second-order), and accelerations (i.e., third-order) using state-space models. To simply modeling of vehicle dynamics, the point-mass (which treats a car as a particle) and bicycle or front wheel steering (which treats a car as a 2D) models are commonly used (Mueller, 2019).

Remark 2.1. The AV control problem in Definition (2.1) can be categorized based on traffic environments (including highways and urban roads) and driving tasks (including merging, diverging, lane-keeping, lane-changing, turning). Traffic environments determine driving tasks (via the action space), interacting agents (via the state space), and state update dynamics. Driving tasks determine the control type, in other words, what control or the combination of controls to use. On highways, AVs need to control both longitudinal (including car-following, lane-keeping) and latitudinal accelerations (lane-changing). On urban roads, traffic compositions differ and include not only cars, but also other road users. The driving tasks can differ. At unsignalized intersections, AVs need to pass while avoiding colliding to any cars or road users in the conflicting directions. At signalized intersections, AVs need to observe traffic lights while also paying attentions to those in conflicting zones. At roundabouts, AVs need to pass while avoiding colliding to any cars or road users in the conflicting directions. In summary, interacting agent types, traffic dynamics, and vehicle control objectives can differ in different traffic environments. To specify how the AV control problem may vary in different settings, subsequently we will define each component mathematically and point out the variants.

With the high-level problem statement, we will materialize mathematical components that form the AV control problem of MVS. Note that different communities may use different mathematical notions and notations. We try to reconcile them in a unified framework.

- $\mathscr{E}$. The mixed traffic environment composed of interacting agents, which are various road users. The traffic environment can be highways or urban streets. On highways, AVs interact with other AVs or HVs in the same lanes or adjacent lanes. On urban streets, AVs interact with other vehicles and vulnerable road users such as pedestrians and cyclists.
- $\mathbf{s}, \mathscr{S}$. The state vector and state space of the mixed traffic environment. Denote $\mathbf{s}^{(n)} \in \mathscr{S}^{(n)}$ as the state for the n^{th} car. The physical state of the n^{th} car at time t can be represented by $\mathbf{s}_t^{(n)} = (x_t, y_t, \theta_t, v_t, a_t)$, where x_t, y_t are its longitudinal and lateral positions, θ_t is the axle angle, v_t, a_t are speed and acceleration. The joint state of N cars is denoted as $\mathbf{s} = (s^{(1)}, \ldots, s^{(N)})$. The joint state space is $\mathscr{S} = \mathscr{S}^{(1)} \times \mathscr{S}^{(2)} \times \cdots \times \mathscr{S}^{(N)}$. The internal state of the n^{th} car is always observable to itself, but the state of other cars can be fully, partially, or not observable to the ego car.

  One's state is usually associated with physical or safety constraints, including driving within a lane and not colliding to other cars.
- $\mathbf{o}, \mathscr{O}$. The observation vector and observation space. If the state of a traffic environment is fully observable to the ego car, the observation space equals the state space. If the state of a traffic environment is partially observable, the observation space is usually a subset or subspace of the state space.

  An observation $\mathbf{o}^{(n)} \in \mathscr{O}^{(n)}$ of the n^{th} car contains the observed information of its surrounding road users. For instance, in a platoon without V2V communication, the observation vector for the n^{th} car at time t includes not only its own state but also that of its immediate leader, which is $\mathbf{o}_t^{(n)} = (x_t^{(n-1)}, y_t^{(n-1)}, \theta_t^{(n-1)}, v_t^{(n-1)})$. In a connected platoon with the states of both the immediate leader and the platoon leader transmitted to all other cars (Ge and Orosz, 2014), the observation vector for the n^{th} car at time t becomes $\mathbf{o}_t^{(n)} = (x_t^{(n-1)}, y_t^{(n-1)}, \theta_t^{(n-1)}, v_t^{(n-1)}, a_t^{(n-1)}) \times (x_t^{(1)}, y_t^{(1)}, \theta_t^{(1)}, v_t^{(1)}, a_t^{(1)})$. The joint observation of N cars, i.e., $\mathbf{o}$, is denoted as $\mathbf{o} = (o^{(1)}, \ldots, o^{(N)})$. The joint observation space $\mathscr{O} = \mathscr{O}^{(1)} \times \mathscr{O}^{(2)} \times \cdots \times \mathscr{O}^{(N)}$.

  In a complex urban environment, assume that the ego AV can only observe the positions of its surrounding road users, then its observation space includes the positions of these agents.
- $\mathscr{G}$. The observation function. Denote $\mathscr{G}^{(n)}$ as the observation function mapping of the n^{th} car from the state vector to its observation vector. It draws an observation $o^{(n)} \in \mathscr{O}^{(n)}$ that is correlated with $\mathbf{s}$ according to the observation function $\mathscr{G}^{(n)} : \mathscr{S} \to \mathscr{O}^{(n)}$, i.e., $o^{(n)} \sim \mathscr{G}^{(n)}(o^{(n)}|\mathbf{s})$. In a fully observable environment, $\mathscr{G}^{(n)}$ is the identity function. The set of observation functions is $\mathscr{G} = \{\mathscr{G}^{(1)}, \mathscr{G}^{(2)}, \ldots, \mathscr{G}^{(N)}\}$.
- $\mathbf{a}, \mathscr{A}$. The action vector and action space. The action of the n^{th} car, denoted as $a^{(n)} \in \mathscr{A}^{(n)}$, can be a two-dimensional vector, including acceleration and steering angle. The joint action of N cars is $\mathbf{a} = (a^{(1)}, \ldots, a^{(N)})$. The joint action space is $\mathscr{A} = \mathscr{A}^{(1)} \times \mathscr{A}^{(2)} \times \cdots \times \mathscr{A}^{(N)}$.
- $f(\cdot)$. The state update function that updates the state of the traffic environment from time $t-1$ to t, which is

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{a}_t; \mathbf{s}_{t-1}, \mathbf{a}_{t-1}; \cdots). \tag{2.2}$$

The state update function can vary in different traffic environments. If N cars drive either in a platoon or an isolated environment (such as on a closed track), the state update of these cars is reduced to a deterministic mapping.

If AVs are driving on highways or urban streets surrounded by HVs or other road users, the traffic environment becomes stochastic, because generally the state is partially observable. Then the ego cars can only predict the state transition with certain probability. The stochastic environment, if Markovian, is commonly characterized by a state transition function, denoted as $\mathscr{P}(\cdot)$, which is a mapping $\mathscr{S}\times\mathscr{A}\rightarrow[0,1]$. The joint action at time t triggers a state transition according to function $P(\mathbf{s}_{t+1}|\mathbf{s}_t,\mathbf{a}_t)$. This state transition can be computed from a specific form (model-based) or a mixed traffic simulator (model-free).

The state update function can also be parameterized by factors including roadway characteristics (such as speed limit, lane width) and driver characteristics (aggressive or conservative), which can be calibrated using historical data or learned online using real-time data.

- $r(\cdot),\mathscr{R}$. The reward function and reward space. Along with the state transition, car n receives an immediate reward, i.e., $r^{(n)}\in\mathscr{R}^{(n)}:\mathscr{S}\times\mathscr{A}\times\mathscr{S}\rightarrow\mathbb{R}$. The reward may include traffic safety (e.g., not off-road, collision avoidance), efficiency (e.g., fast speed, short travel time to reach destinations, minimum delay), and emissions (e.g., minimum energy use per mile travel). One aims to maximize its discounted expected cumulative reward by deriving an optimal policy, which is the best response to the driving policies of other cars in the environment. The joint reward space of N cars is $\mathscr{R}=\mathscr{R}^{(1)}\times\mathscr{R}^{(2)}\times\cdots\times R^{(N)}$.

  Depending on if N cars are cooperative or competitive, reward functions are in different forms. Even when N cars are cooperative, centralized and distributed controllers result in different reward forms.
- $\mathscr{J}(\cdot)$. The total cost function.

  In a deterministic environment, assuming that the central planner knows the state of every vehicle and aims to optimize a total system performance, from the system perspective, the multi-AV cost function becomes:

$$\mathscr{J}_N(\mathbf{s}^{(1)},\mathbf{a}^{(1)};\cdots;\mathbf{s}^{(N)},\mathbf{a}^{(N)}). \tag{2.3}$$

One example is to formulate an optimal control problem for a platoon of collaborative AVs. Define $\mathscr{J}_{run},\mathscr{J}_{ter}$ as the total running cost and terminal cost, respectively. Then a centralized cost function is specified as:

$$\mathscr{J}_N=\min_{\mathbf{a}}\sum_{t=0}^{T-1}\mathscr{J}_{run}\left(\mathbf{s}_t^{(1)},\mathbf{a}_t^{(1)};\cdots;\mathbf{s}_t^{(N)},\mathbf{a}_t^{(N)}\right)+\mathscr{J}_{ter}(\mathbf{s}_T,\mathbf{a}_T). \tag{2.4}$$

When AVs are programmed to optimize its own objective and not cooperate with other AVs, the multi-AV control becomes a non-cooperative game. In a non-cooperative system, vehicles select theirs own controls to achieve individual goals, which may likely conflict with others' goals. The objective of the non-cooperative framework for a simultaneous game is:

$$\min_{\mathbf{a}^{(n)}}\mathscr{J}_N^{(n)}(\mathbf{s}^{(n)},\mathbf{a}^{(n)};\mathbf{s}^{(-n)},\mathbf{a}^{(-n)}),n=1,\cdots,N. \tag{2.5}$$

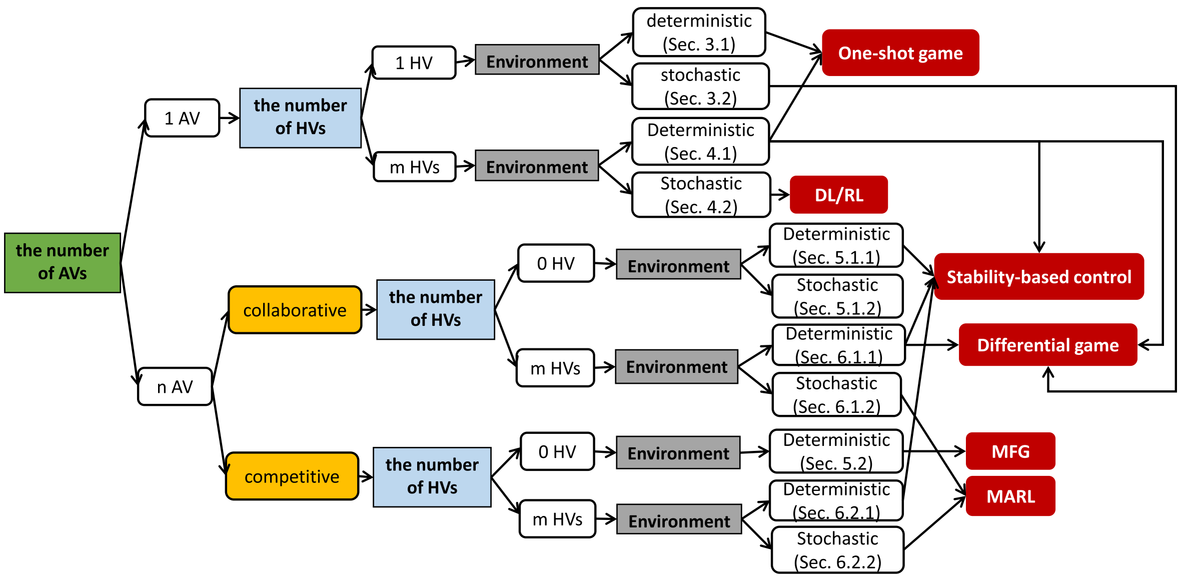


Fig. 2. Roadmap to models reviewed in this paper.

2.2. Research questions

Given the problem components defined above, we would like to pinpoint four major research questions, which will guide the flow of methodological review in the rest of the paper.

- Q1 What **scalable driving policies** are to control a large number of AVs in mixed traffic comprised of human drivers and uncontrollable AVs? (Sections)3–6
- Q2 How do we estimate human driver behaviors? (Sections)3–6
- Q3 How should the driving behavior of uncontrollable AVs be modeled in the environment? (Sections)3–6
- Q4 How are the interactions between human drivers and AVs characterized? (Sections)3–6

Remark 2.2. The first research question (Q1) is associated with how to establish the mapping of driving control policies. The second and third research questions (Q2,Q3) are associated with the state update dynamics, aiming at accurate modeling and prediction of surrounding road user behaviors. The forth research question (Q4) relates to how the selected driving control policies influence other road users and vice versa, which provides feedback to AV controls and also determines the state update dynamics.

2.3. Roadmap to navigate: Control dimensions

In the subsequent sections, we will give an overview of how the existing literature addresses these gaps. We categorize these studies based on how many AVs and HVs are involved. They include: one AV interacts with one HV (1 AV + 1 HV), one AV navigates in the HV-dominated traffic environment, i.e., one AV interacts with multiple HVs (1 AV + m HVs), multiple AVs interact with many HVs (n AVs + m HVs, $n \ll m$), multiple AVs interact with one HV (n AVs + 1 HV), and a pure AV market (n AVs).

Other than the number of AVs and HVs, we further categorize the multi-AV control problem based on two dimensions: the first dimension is whether these controllable AVs are cooperative or not, while the second dimension is whether the AV control takes into account uncertainty arising from the external environment.

In the subsequent Sections 3–6, we will categorize the literature using the number of involved AVs and HVs as the primary category, vehicle cooperation as the secondary category, and the environmental stochasticity as the third category.

Because of the broad scope of the relevant studies, to navigate readers to literature on AV controls in mixed traffic, Fig. 2 provides a roadmap pointing to each (sub) section. The colored rectangles represent the classification criteria and the red round-corner rectangles (as the end nodes) represent the methodologies used to control AVs.

Remark 2.3.

1. When there exists only one single AV that needs to be controlled, cooperation or competition across AVs does not exist. So this category will be skipped in Section 3 ("1 AV + 1 HV") and Section 4 ("1 AV + m HV").
2. In the pure AV market in Section 5 ("n AVs"), a majority of studies assume the environment is deterministic because all the AVs are controllable and fully observable (for CAVs), while stochasticity could originate from measurement errors or communication delay. In the mixed market covered in Sections 3, 4,6, the environment is usually random due to stochasticity of human driving behavior, but in certain scenarios like platooning or driving in a closed environment, a deterministic environment is also assumed.

3. 1 AV + 1 HV, 1 AV + 1 AV

In this section, we will categorize the models based on whether the traffic environment, constituted by one HV or AV, is deterministic or stochastic. In both contexts, game based controls are employed in the existing literature. Game theory is a natural approach to model the non-cooperative strategic interactions among AVs or between one AV and one HV, who are usually taken as intelligent agents aiming to optimize an individual objective function. In the game theoretic framework, cars are referred to as "agents" or "players".

3.1. A deterministic environment: One-shot game

The one-shot two-person game is applied to model two cars' strategic actions at one step. Driving (Yoo and Langari, 2012), merging (Liu et al., 2007; Yoo and Langari, 2013), lane-changing (Talebpour et al., 2015; Yu et al., 2018; Zhang et al., 2019c; Yoo and Langari, 2020), and unprotected left-turning behavior (Rahmati and Talebpour, 2017) is modeled as either a two-person non-zero-sum non-cooperative game (Liu et al., 2007; Talebpour et al., 2015), a Stackelberg game (Yoo and Langari, 2012; Yoo and Langari, 2013; Yu et al., 2018; Zhang et al., 2019c; Yoo and Langari, 2020) or a mixed-motive game (Kim and Langari, 2014). The outcome of these games can be a pure or mixed Nash equilibrium, based on the payoff bimatrix. The payoff of a given strategy accounts for traffic safety and efficiency, depending on current driving speed, relative positions, reaction and perception time, aggressiveness, and collision avoidance. When human driving behavior is modeled, human driver data is collected to estimate parameters of payoff functions, using bi-level optimization (Liu et al., 2007), simulated moments (Talebpour et al., 2015), and maximum likelihood (Rahmati and Talebpour, 2017). When one of the game players is an AV, utility or reward needs to be designed while accounting for aggressiveness of surrounding drivers (Yoo and Langari, 2020). Zhang et al. (2019c) further develops a game theoretic model predictive controller that solves a Stackelberg equilibrium with multiple interacting vehicles continuously.

3.2. A stochastic environment: Dynamic game

The one-shot game cannot model vehicles' dynamic driving actions. To solve for time-varying controls, dynamic optimal control (Wang et al., 2015), model predictive control (MPC) (Wang et al., 2016; Gong et al., 2016; Gong and Du, 2018) or rolling horizon control (Swaroop et al., 1994; Wang et al., 2014a; Wang et al., 2014b; Zhou et al., 2017b) have been formulated for AVs. When one agent solves an optimal control problem, while its interacting agent also does so with conflicting goals, a differential game forms (Wang et al., 2015). *Differential game* models dynamic behaviors of interacting agents with conflicting goals, where agents' optimal strategies are obtained from optimal control problems.

If the AV were to be able to predict the HV's strategy in the entire planning horizon, it then optimizes its own objective function that depends on both its own current and future strategies as well as the HV's current and future strategies to generate a continuous sequence of control strategies along this horizon and implement it. The same process holds for the HV. Due to the dynamic coupling, it is challenging to solve this equilibrium. Thus, the classical differential game is primarily focused on two players and becomes intractable for an equilibrium of more than two players.

To simplify, several techniques have been applied. Sadigh et al. (2016b), Lazar et al. (2018a) have simplified the original two-player differential game to a leader–follower game (or Stackelberg game) played at discretized time steps. In this game, the AV takes actions first. Then the HV observes actions taken by the AV and predicts the AV's future action based on the AV's historical actions, maximizes its own objective and calculates its own future actions for a short period of time. Then the AV maximizes its own objective using the HV's future actions and replans repeatedly using MPC at each iteration. In other words, in a leader–follower scheme, the AV directly solves an optimization problem based upon its prediction of human driver actions rather than human's actual strategies. The advantage of the Stackelberg game is that the AV can be designed beforehand to influence uncontrolled HVs via a carefully selected reward function (Sadigh et al., 2016b). The reward function contains two parts: one controls the AV's driving efficiency and safety, while the other determines the influence the AV would like to impose to neighboring HVs. Lazar et al. (2018a) extends this framework to a Stackelberg game between one AV and multiple HVs, but assumes that one AV only influences one HV and the actions of others HVs are fixed.

Fisac et al. (2019) further develops a hierarchical game-theoretic planning scheme, where the strategic planner solves a closed-loop dynamic game with approximate dynamics in a relatively long planning horizon (e.g., 5 s), while the tactical planner solves an open-loop trajectory optimization with high-fidelity vehicle dynamics over a shorter planning horizon (e.g., 0.5 s). On the strategic planner level, the AV and the HV still play a feedback Stackelberg dynamic game in which their driving actions are recursively solved through successive application of dynamic programming. The solved optimal Q-value obtained from the strategic level is then introduced to the objective function of the tactical level as a guiding terminal reward representing an optimal reward-to-go. On the tactical planner level, the trajectory of the AV is iteratively optimized using a nested optimization problem that estimates the human's best trajectory response to each candidate plan in the short-term planning horizon. The hierarchical game-theoretic model is tested on two scenarios with merging and overtaking maneuvers: one on a straight empty multi-lane highway with only two-vehicle interaction and one with the presence of a third vehicle (i.e., a truck with a slower moving speed).

In a stochastic environment when adversarial risks are present, adversarial learning game has been borrowed to model human-robotic interaction and train robust AV controllers. Assuming that the HV is an adversary attempting to falsify the AV's actions, Sadigh et al. (2019) first learns the HV's reward underlying its actions using maximum entropy inverse reinforcement learning and then computes sequential AV controls with nested optimization.

4. 1 AV + m HV

In this section, we will divide the models based on whether the traffic environment, constituted by a number of HVs, is deterministic or stochastic. In a deterministic traffic environment, stability based and game based controls are employed; In a stochastic traffic environment, deep learning and reinforcement learning are employed to train driving policies.

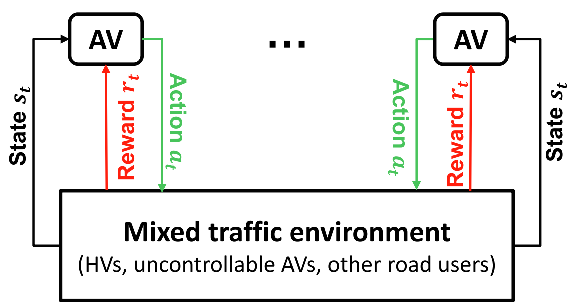


Fig. 3. Single- and multi-agent reinforcement learning framework.

4.1. A deterministic environment

4.1.1. Stability-based control

When the environment is deterministic, the traffic control community aims to understand how one AV can stabilize an HV platoon using linear (Cui et al., 2017; Wang, 2018) or nonlinear controllers (Jin and Orosz, 2014; Jin and Orosz, 2018), based on the *head-to-tail string stability* (i.e., stability from the first vehicle to the last vehicle in a platoon (Jin and Orosz, 2014)). Field experiments have also demonstrated the feasibility of using one AV to stabilize HVs (Stern et al., 2018; Jin and Orosz, 2018). Stability oriented method is applied more extensively for the scenario of a platoon of multiple CAVs, so more studies along this line will be introduced in Section 5 ("n AVs") and Section 6.1.1 ("n AV + m HV").

4.1.2. Game based control

Assume the HV-dominated environment is deterministic and every vehicle interacts among one another in a game-theoretic framework, a simultaneous one-shot game or a differential game is modeled between one AV and multiple HVs.

Zhang et al. (2019c), Coskun et al. (2019), Zhang et al. (2019b) develop a sequence of one-shot Stackelberg game for an AV to perform mandatory lane-change in the HV-dominated traffic stream. The goal is to train autonomous driving with the human-like performance in gap acceptance and lane-change. HVs are modeled either by IDM (Zhang et al., 2019c) or as a game player (Zhang et al., 2019b; Coskun et al., 2019). A hierarchical control is developed: The high-level controller estimates the surrounding HVs' aggressiveness and performs Stackelberg games with every HV in the neighborhood of the AV. The strategy (including time to change lane and driving accelerations) of a Stackelberg game that leads to the highest reward is chosen as the optimal driving strategy, which is fed into the low-level controller. MPC is used in the low-level controller to update the subject vehicle's position based on both current and future rewards. A human-in-the-loop simulator (Zhang et al., 2019b; Coskun et al., 2019) is built in MATLAB Simulink, with participants controlling HVs in a virtual multi-lane highway environment.

Schwarting et al. (2019) develops an autonomous control policy by solving an iterative best-response, with embedded levels of tacit negotiation. In a two-agent case, an iterative best-response can be written as $\mathbf{a}^{(AV)}(\mathbf{a}^{(HV)}(\mathbf{a}^{(AV)}(\cdots)))$. In an MVS, a system of interdependent optimization is reduced to a single-level optimization using KKT conditions. The resulting Nash equilibrium offers not only a control law for the AV but also predicted actions for other HVs. The innovation of this study is to include a term "Social Value Orientation (SVO)" into the reward function of HVs, representing HVs' driving aggressiveness. One can adjust its SVO value while interacting with another vehicle. The control law is validated in highway merging and unprotected left turn. Social preference learning can improve the AV's performance by 25%.

To address the challenge in modeling a large amount of HVs simutaneously, Liu and Tomizuka (2015, 2016) combine multiple HVs as one effective agent and assume a sequential game in which HVs lead and the AV play reactive strategies. By mapping a baseline control law to a set of safe control, an online algorithm is developed for the AV controller to incorporate human intentions as safety constraints.

Remark 4.1. For the scenario "1 AV + m HV", there exist only a few studies using game-theory based controls, partly due to high dimensionality of the coupled game system. In contrast to regarding each individual HV as an intelligent agent, a majority of studies treat all these HVs as the stochastic environment where the controlled AV is operated. Thus deep learning and reinforcement learning based AV controls are modeled, which will occupy the most space in the rest of the section.

4.2. A stochastic environment

4.2.1. Deep learning (DL) based control

Prior to the prevalence of reinforcement learning, (deep) supervised learning is a popular tool to train model-free and end-to-end (End2End) controllers for AVs, which directly maps sensory inputs to control commands. Behavioral cloning (BC) simplifies the AV policy learning as a supervised-learning problem, which usually performs well when driving data are sufficient or the driving task is for limited regions. Pomerleau (1989) introduce a multi-layer network learned from simulated road images to control a vehicle to follow real roads. The next milestone of AVs is to employ convolutional neural networks (CNNs) to efficiently process raw camera images, which helps AVs drive through an obstacle-filled road after training on similar scenarios (Muller et al., 2006). Later on, CNN-based AV controllers are widely studied, and recently work includes NVIDIA's PilotNet (Bojarski et al., 2016; Bojarski et al., 2017) to control AVs in real traffic situations, Rausch's deep CNN policy (Rausch et al., 2017) and DeepPicar (Bechtel et al., 2018) for steering angle control, Agile driving (Pan et al., 2018) for steering angle and velocity controlling in aggressive scenarios. Temporal dependencies of the driving data have been considered to improve the performance of AV control, and recently, long-short-term memory (LSTM) and its variants have been leveraged for End2End AV policy learning. Xu et al. (2017) proposes FCN-LSTM, a combination of a fully-convolutional network (FCN) and LSTM, which can predict a distribution of future vehicle ego–motion data. Eraqi et al. (2017) develops a convolutional LSTM (C-LSTM) for learning both visual and dynamic temporal dependencies of driving. Hecker et al. (2018a) introduces Drive360, which combines CNN, fully connected layers and LSTM to integrate information from multiple sensors to predict the driving maneuvers. Bansal et al. (2019) presents ChauffeurNet, a mid-to-mid driving policy learning framework, in which inputs are prepossessed before they are received by a recurrent neural network (RNN) to generate low-level controls.

4.2.2. Reinforcement learning based control

In the HV-dominated traffic, a single AV's driving policy selection can be treated as a sequential decision-making process in a partially or fully observable random environment. Learning driving policies are needed to predict vehicles' acceleration and steering angle using their environmental information as input.

Reinforcement learning (Sutton and Barto, 1998), which enables the intelligent agents to learn optimal policies driven by a reward, has made breakthroughs to achieve super-human-level performance in game playing, such as Atari Mnih et al. (2015), Go game (Silver et al., 2016), Poker (Brown and Sandholm, 2018; Brown and Sandholm, 2019), Dota 2 (OpenAI, 2018), and StarCraft II (Vinyals et al., 2019). Its application to autonomous driving has become a promising direction.

When RL is used to control AVs in a stochastic environment, the basic idea is demonstrated in Fig. 3. Let us first discuss the single-agent RL setting. One controllable AV (the host vehicle whose controller needs to be designed) perceives the state of the mixed traffic environment, comprised of HVs, other AVs that are not controllable, and road users. Based on some predefined reward function, it executes an action (such as acceleration or steering angle), which in turn transforms the state of the traffic environment. In return, the environment provides a reward to the AV. Based on the received reward and the new environment state, the AV further solves an optimal policy based on the reward function and selects an action. This process iterates till the AV finishes its entire control process. When there are multiple AVs that are all required to make decisions simultaneously, we need a multi-agent RL (MARL) framework, which will be discussed in Section 6.2.

MDP implicitly assumes that the agent can fully observe the state dynamics. In other words, after the agent applies an action, he knows the probability of the next state the system will move to. A majority of transportation studies assume connectivity among vehicles via V2V or V2I. Thanks to these communication technologies, every driver obtains full information of other drivers and the system state. Control and robotics researchers, on the other hand, make various assumptions on observability (Liu and Tomizuka, 2015; Liu and Tomizuka, 2016; Liu et al., 2018; Bouton et al., 2017; Bouton et al., 2018). One may observe others' positions, headings, and sometimes longitudinal and lateral velocity, but not accelerations. Even a driver observes the entire state of other drivers, she may not know the intention of those drivers. Accordingly, the AV has to maintain a belief state space over all possible states based on its observations. Therefore a partially observable Markov decision process (POMDP) model is widely used in modeling a single AV's motion planning, so is the single-AV control problem in the HV-dominated traffic.

When traffic dynamics is too complex to model with POMDP, model-free approaches such deep reinforcement learning (DRL) are widely employed. To implement model-free training of AVs, a simulator is needed that provides consequential information, such as reward and traffic state update, for AVs to learn from. Wu et al. (2017b) train an autonomous driving controller using DRL with a designated reward function that avoids crashes into other agents, and applied trust region policy optimization (TRPO) method to train a Gaussian Multilayer perceptron (MLP) policy in SUMO simulator for improving traffic efficiency. Lillicrap et al. (2015) applies a deep deterministic policy gradient (DDPG) RL algorithm to control a car in a simulation environment. Their work designed a reward, which provides a positive reward at each step for the velocity of the car projected along the track direction and a penalty of -1 for collisions. Sallab et al. (2017) develops an integrated deep Q-network (DQN), which integrates attention models to make use of glimpse and action networks to direct the CNN kernels for steering command in the TORCS simulator, which provides a positive/negative reward for on/off-lane situations. Perot et al. (2017) proposes an asynchronous advantage Actor-Critic (A3C) method for training a policy network for realistic games, such as World Rally Championship 6 and the TORCS simulator. This work has been enhanced by Jaritz et al. (2018) with an improved convergence and generalization. Both of these studies have designed the reward as a function of the distance to the road center and angle between the road's and car's heading. In additional to some common reward functions that punish collisions and lane departure, Landolfi and Dragan (2018) introduce an extra reward term, a socially cohesive reward, which learns the mean and variance of social features extracted from surrounding human drivers. This additional reward term enables AVs to learn socially acceptable policies and match its own behavior to the neighbor. This socially cohesive driving is tested in scenarios when unwritten driving rules dominate, such as the presence of an ambulance and when everyone is speeding.

Monte Carlo tree search (MCTS) is one of the most effective methods for solving decision making problems online (Browne et al., 2012) and has been applied to AVs in recent years. Paxton et al. (2017) integrate MCTS with hierarchical neural net control policies trained on Linear Temporal Logic (ITL) constraints for motion and path planning in complex road environments. They designed the reward function as a combination of cost terms upon current continuous states (e.g., location or speed), and a bonus terms based on completing immediate goals (e.g., stopping at the sign or existing a region), and a penalty term for constraint violations. Sunberg et al. (2017) use MCTS to infer the internal state of traffic participants for operating safe lane changes on a highway. Their reward design penalized the average time taken for the ego to reach the target lane and the number of hard braking maneuvers that any vehicle undertakes during the time for the ego vehicle to reach the target lane. Hoel et al. (2020) combine MCTS with DRL to achieve tactical highway driving, and in their work, a deep neural network is trained to guide MCTS to the relevant regions of the search tree, while MCTS is used to improve the training process of the neural network at the same time. The associated reward is a combination of cost terms concerning the number of lane changes and difference from the desired speeds, and a bonus terms for highway exit (i.e., goal achieved).

There is another direction of DRL-based AV control, using prior knowledge or classical theory-driven controllers to constrain the learning and behaving of neural network-based driving models. Bouton et al. (2017) impose a computational safety factor as a penalty in the reward function rather than a hard constraint (Bouton et al., 2017), and as a result, the driving policy solved from MDP cannot avoid accidents. Bouton et al. (2018) add a model checking step to enforce probabilistic guarantees of the trained driving policy on an RL agent, and they used simplified reward function to penalize the number of action steps and award goal accomplishment. Zhang et al. (2016) propose to combine the traditional MPC method with RL in the framework of guided policy search for controlling autonomous aerial vehicles, where a deep neural network policy is trained on data generated by MPC for training robustness and generalizable control. The RL training is based on a cost function that measures the distribution difference between the action generated from the policy model and the data generated from MPC. Chen et al. (2019b) develops a hierarchical control framework, where the higher-level controller employs MDP to solve a reference driving policy and the lower-level controller implements it accounting for safety concerns.

For more examples of using single-agent RL for AV controlling, we refer readers to recent surveys such as (Zhou and Laval, 2019; Grigorescu et al., 2020; Kiran et al., 2020).

4.3. Adversarial attack and cybersecurity of AI-based models

When moving toward the era of mixed-autonomy, it is not inessential to assume the existence of malicious HVs and uncontrollable AVs that negatively affect or even intentionally attack the developed benign AV and CAV systems. A safe and efficient mixed traffic ecosystem cannot stand without a thorough understanding on (C) AV-related adversaries and the development of effective defence methods against them. This section gives a review of the two most important (C) AV-related attacks, cyber and DL attacks, and how existing works are leveraged to mitigate these issues.

AVs are vulnerable to adversarial cyber and DL attacks, especially when we are moving towards CAV systems. Cybersecurity of CAV has been challenged by cyber-attacks, including but not limited to authenticity attacks (Douceur, 2002; Amoozadeh et al., 2015; He et al., 2015; Ucar et al., 2017; Lyamin et al., 2018), network layer attacks (Garip et al., 2015; Qayyum et al., 2020), and systems level attacks (Shoukry et al., 2013). Because these attacks predominantly target physical communication layers not CAV controllers, we will not details these methods in this paper. Interested readers can refer to Rajbahadur et al. (2018) for details.

As DL is becoming a backbone of AV systems, adversarial attacks can happen to both the physical world where AVs are operated and the digital world where AVs need to process information and perform computation. The attack algorithms and their variants have been applied including but not limited to object detection, sensor fusion, and deep RL. Object detection, in particular image recognition and segmentation directly relates to the ability of an AV to process the image information from cameras to recognize and detect traffic signs, traffic signals, and objects nearby. Successful attacks in this domains have been demonstrated recently. Song et al., who generate perturbed "STOP" sign using the robust physical perturbation algorithm to make the sign remained hidden from the state-of-the-art YOLO model Eykholt et al. (2018). Aung et al. investigate a scheme called FGSM (Goodfellow et al., 2014b) to generate adversarial traffic signs to fool a DNN-based traffic sign recognition model (Aung et al., 2017). Zhang et al. demonstrate how a simulator applies camouflage to the vehicle to minimize the detection score and fool the image-based object detector (Zhang et al., 2018a). Sitawarin et al. design ''toxic" signs such as malicious ads and logos that look benign to human observers but deceive the traffic sign recognition mechanism of AVs Sitawarin et al. (2018a), Sitawarin et al. (2018b).

Multi-Sensor Fusion (MSF) is widely considered as a safer perception system for AVs because it combines perception results from multiple sensors such as LiDARs (Light Detection And Ranging) and cameras to achieve overall higher accuracy and robustness. However, this system has recently been shown to be deceivable. Cao et al. propose an attack design, called MSF-ADV, to generate physical-world adversarial 3D objects that simultaneously fool both LiDAR- and camera-based perception Cao et al. (2020). In demonstration, they print the designed 3D object using a 3D printer and place it in the middle of a road. Such adversarial object is able to fool the DNN-based object detection models for both LiDAR and camera, and thus, becomes completely invisible to the victim AV with advanced MSF algorithms. As a results, the victim AV will ignore such object and directly crash into it.

Besides perceptions, the DRL-based AV controllers can also be attacked, ranging from input perturbations to model attack. A majority of DRL attacks are based on input perturbations. Behzadan and Munir (2017) designs input perturbations to manipulate policies during the training time of DQNs. Huang et al. apply the FGSM to attack feed-forward policies trained with DQNs, asynchronous advantage actor-critic (A3C), and trust region policy optimization (TRPO) (Huang et al., 2017). Lin et al. introduce the "enchanting" attack to lure the DNN policy to a maliciously designed state by generating a sequence of corresponding actions through a sequence of adversarial examples (Lin et al., 2017).

In addition to input perturbations, the research of model-based backdoor attacks against DRL begins to thrive. The backdoor attacks are implanted into the policies during the training time through minuscule data poisoning (e.g., as little as 0.025% of the training data) and in-band reward modification that does not affect the reward on normal inputs. And the data poisoning results in malicious behaviors which are dormant until a certain trigger emerges, making these attack very stealthy. The trigger design is flexible because it is not dataset-dependent. Related works include TrojDRL, which presents backdoor attacks on policies to perform imperceptibly similar to benign policies but deteriorate drastically when the attack is triggered in both targeted and untargeted settings Kiourti et al. (2019). Wang et al. (2020b) explores backdoor attacks on deep reinforcement learning-based traffic congestion control systems, and use sensor values as triggers to activate attack on a DRL-based AV controller, including malicious vehicle deceleration and acceleration behaviors to cause stop-and-go traffic waves to emerge (congestion attacks), or AV acceleration resulting in the AV crashing into the vehicle in front (insurance attack).

Associated with the investigation of DL adversarial attacks, the adversarial training paradigm has been proposed to develop adversarially robust DL solutions, i.e., adversarial defense methods. As firstly proposed by Goodfellow et al. (Goodfellow et al., 2014b), learning with adversarial examples can make deep neural networks robust against adversarial attacks. Adversarial examples are a kind of inputs formed by applying small but intentionally worst-case perturbations to examples from the dataset, such that the perturbed input results in the model outputting an incorrect answer with high confidence. The adversarial defense methods that have thrived in DL hold the potential for robust applications in (C) AV systems. In particular, the development of adversarially robust objection detection systems can prevent adversaries from getting the control of (C) AVs in an attempt to cause an accident (Qayyum et al., 2020). Sadigh et al. (2019) employs such a concept to model a game-theoretic interaction: the HV serves as the AV's adversarial and play an adversarial game with the AV. The goal is to find a sequence of human driving actions that could lead to the AV's unsafe behavior. In this game, the AV takes actions to maximize its cumulative reward, while the HV tries to falsify the AV's action by selecting driving actions that minimize the AV's reward function. Nevertheless, the design, implementation and experiment with adversarially robust models for AVs are still lacking, and promising research opportunities lie in this direction.

5. n AVs: A driverless world

Control of a single AV is far from sufficient to exploit the potentials of AVs in the era of mixed autonomy, when an increasing number of AVs are introduced to public roads. Systems with multiple AVs have attracted increasing interest in recent years. In this and the next sections, we go against the AV deployment timeline by first discussing the pure driverless world, denoted as "n AVs", followed by the mixed market (of n AVs + m HVs). The pure AV market precedes the mixed one, because the former can be generalized to the latter by adding a traffic background comprised of HVs.

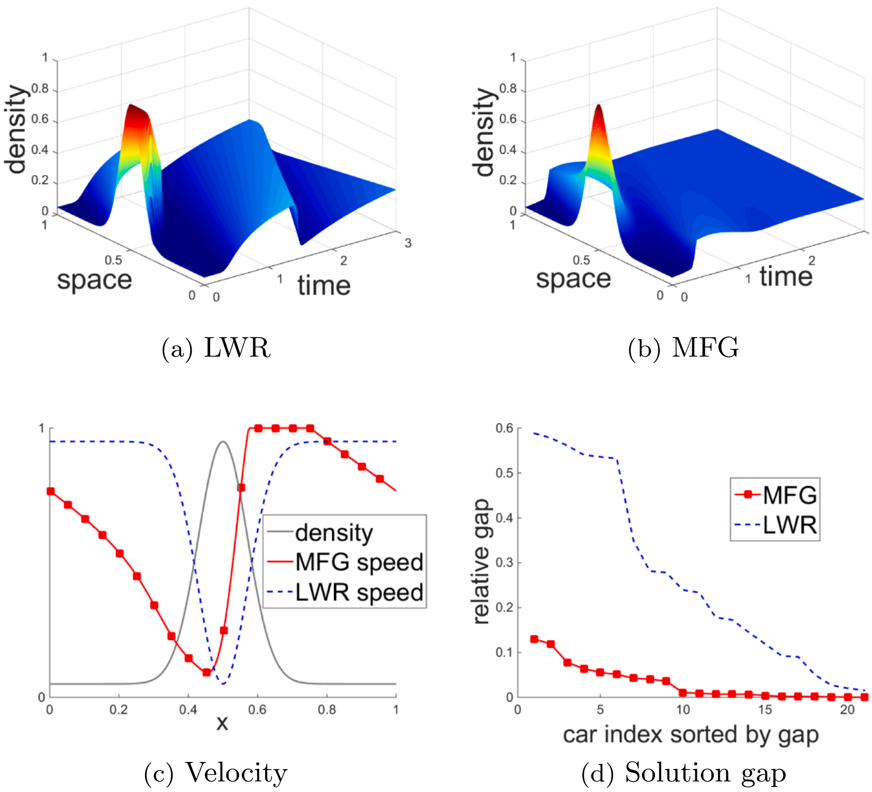
5.1. Cooperative control

5.1.1. A deterministic environment: Stability-based controls

A majority of research on control of multiple AVs falls within the category of cooperative coordination. In other words, AVs are assumed to communicate with one another for global traffic information and optimize a common goal of traffic flow improvement. Cooperative control has been widely studied in multi-robotic systems. Swarm intelligence (Bogue, 2008; Venayagamoorthy and Doctor, 2004), formation control (Chen and Wang, 2005), and consensus control (Zegers et al., 2017; Li et al., 2018c) have been widely used for a group of robots with a centralized goal to accomplish a task collaboratively, so is in multi-AV control (Wu et al., 2018; Lazar et al., 2018b).

In a cooperative MVS, the movement of vehicles is coordinated by a central controller or planner to achieve a common goal, such as to collectively stabilize traffic flow and smoothen traffic jam (Wang et al., 2016; Gong et al., 2016; Gong and Du, 2018), to optimize driving comfort (Wang et al., 2014b; Zhou et al., 2017b), or to improve fuel efficiency (Wang et al., 2014a; Yao et al., 2018). To achieve coordination, *full observability* and *full controllability* is required, meaning that all vehicles' states and controls are known to the central controller and every vehicle can be controlled in a centralized or distributed manner. The communication topology in a platoon of vehicles determines the degree of cooperation among CAVs (Li et al., 2014). For example, control protocols can be designed for a platoon of vehicles to reach an equilibrium state using a consensus based approach. Accordingly, the car-following coupling among vehicles are modeled as a consensus problem and a distributed nonlinear delay-dependent control algorithm is used to solve a safe velocity (Li et al., 2018d).

Assuming connectivity between predecessors and followers as well as between platoon leaders and followers, CACC contains two



(a) LWR (b) MFG (c) Velocity (d) Solution gap

Fig. 4. LWR (HV) v.s. MFG (AV).

control policies: constant spacing (Swaroop and Hedrick, 1996; Darbha and Rajagopal, 1999; Swaroop et al., 2001) and constant time headway (Ioannou and Chien, 1993; Rajamani and Shladover, 2001; Van Arem et al., 2006; Naus et al., 2010; VanderWerf et al., 2001; Zhou et al., 2017b; Arefizadeh and Talebpour, 2018; Stern et al., 2018). AVs longitudinal acceleration control can also be modeled using nonlinear CFMs, which is discussed in Section 1.4. All the aforementioned studies aim to develop a string stable car-following controller in order to smoothen traffic flow and prevent stop-and-go waves. But none of them considers control and physical safety constraints (Gong and Du, 2018). In other words, interactions among vehicles are not explicitly modeled (Li et al., 2018d). To explicitly model the physical interaction between vehicles, a growing body of literature formulates a platoon of AV longitudinal control as optimal control problems. The control policies based on linear spacing policies or non-linear CFMs are special cases of optimal control problems (Wang et al., 2014b).

Recent years have seen an exponential growth of literature on the design of cooperative driving strategies in either pure or mixed traffic environments, including merging at on-ramps (Zhou et al., 2016; Chen et al., 2020a; Sun et al., 2020; Karimi et al., 2020), lane-change (Ladino and Wang, 2020), driving at roundabouts (Zhao et al., 2018), platooning at signalized intersections (Yao and Li, 2020), and traffic organization on road networks (Wang et al., 2020a; Chen et al., 2020b; Lin et al., 2020). These studies are focused on higher-level trajectory optimization and motion planning with an ideal underlying assumption that CAV controllers can realize the solved trajectories precisely and perfectly regardless of environmental uncertainties or perturbations.

Centralized control requires the central controller to solve for an optimal control for each car at each time step. It is challenging to solve a centralized control of this type, because: (1) all vehicles' states and controls are coupled through objective functions and constraints; (2) A longer planning horizon requires prediction of future traffic dynamics, which may suffer from both curse of dimensionality and disturbances.

To resolve the first issue of state coupling, a distributed algorithm is usually designed and implemented on each vehicle (Wang et al., 2016; Gong et al., 2016; Gong and Du, 2018; Li et al., 2018d; Yao and Li, 2020). Consensus based approaches are also employed to design a control protocol for a platoon of vehicles to reach a consensus and a distributed nonlinear delay-dependent control algorithm is designed to solve a safe velocity (Li et al., 2018d). To resolve the second issue of prediction horizons, the original optimal control problem can be approximated as a one-step MPC and a distributed algorithm is developed. The MPC control is close to optimal control strategies if the planning horizon is short, but may deviate when the planning horizon is long. MPC is also employed as a higher level control model to compute reference planning trajectories (Wang et al., 2014a; Wang et al., 2014b; Gong et al., 2016; Gong and Du, 2018; Zhou et al., 2017b). Interested readers can refer to Li et al. (2014) for a comprehensive survey of stability-based controls.

5.1.2. *A stochastic environment*

The optimal control framework can be extended in several ways. When there are measurement errors or when there is only partial observability, a measurement equation is introduced into the state-space model (Wang et al., 2014a; Wang et al., 2014b; Zhou et al., 2017b). Considering stochastic communication delay arising from packet drops, decomposition is proposed for the stability analysis of a large system of CAVs (Qin and Orosz, 2017; Jin and Orosz, 2018; Jin et al., 2018). Formulating the optimal control problem in the space domain instead of the time domain adopted by most optimal control-based systems, Zhang et al. (2020) develops a robust CACC controller against packet loss and communication delay.

5.2. *Noncooperative control: a deterministic environment*

Compared to the cooperative control, the non-cooperative interactions among AVs are relatively understudied. The noncooperative control of multiple AVs is modeled as N-player game-theoretic models. To the best of our knowledge, all the work on multi-AV competitive control assumes a deterministic environment.

The first group of studies assume there is a small number of AVs to control in specific scenarios such as platooning, in other words, n is a finite number. Wang et al. (2015) formulates AVs discrete lane change and continuous acceleration selections as a differential game, where agents' optimal strategies are obtained from solving optimal control problems. The outcome of a differential game is a dynamic equilibrium. Computation of such a dynamic equilibrium involving N players is mathematically intractable when the number of coupled agents becomes large. To get around, Wang et al. (2015) decomposes the problem into a finite number of sub-problems and applies MPC to each vehicle. Dreves and Gerdts (2018) solves a generalized Nash equilibrium by summing up all vehicles objective functions, which is essentially a cooperative control. Because the game-based control suffer from scalability issues, all the aforementioned studies had to constrain their applications to a limited number of AVs. As a growing number of AVs are put on public roads, a scalable and computational efficient algorithm is needed for a large number of AV controllers.

Another school of research assumes a more generic traffic scenario, which is a large number of AVs interacting with one another on a transportation system, in other words, n goes to infinity. Mean field game (MFG) has shown to be a scalable model for the N-car differential game, as the AV population grows (Huang et al., 2019; Huang et al., 2020a; Huang et al., 2020b). MFG is a game-theoretic framework to model complex multi-agent dynamics arising from the interactions of a large population of rational utility-optimizing agents whose dynamical behaviors are characterized by optimal control problems (Lasry and Lions, 2007; Huang et al., 2006). By exploiting the "smoothing" effect of a large number of interacting individuals, MFG assumes that each agent only responds to and contributes to the density distribution of the whole population. It has become increasingly popular in finance (Guéant et al., 2011; Lachapelle et al., 2010), engineering (Djehiche et al., 2016), and pedestrian crowds (Lachapelle and Wolfram, 2011). In the longitudinal control of AVs, each car solves its optimal velocity backward in time, the aggregate effect of which is formulated by a Hamilton–Jacobi-Bellman (HJB) equation; while the mean field approximation derives the evolution of traffic density solved by a transport equation (with many other names like continuity equation, flow conservation equation) forward in time. To solve the mean field equilibrium, The distributed velocity controller derived from the MFE is shown to be an ϵ-equilibrium of the N-car differential game.

Huang et al. (2020a) has also established a connection between an MFG-based macroscopic continuum model and the existing traffic flow theory. The LWR model, which implicitly assumes that cars move according to hydrodynamics without modeling driving intent, is proved to be a myopic MFG with a specially designed objective function. In conclusion, MFG embodies classical traffic flow models with behavioral interpretation, thereby providing a flexible behavioral foundation and a promising direction to accommodate new traffic entities like AVs. Under the more intelligent objective function of AVs, the LWR velocity does not represent a socially optimal driving strategy as demonstrated by larger deviations from the actual equilibrium in Fig. 4(d). Fig. 4(a-b) illustrate that the MFG mitigates traffic oscillation faster than LWR. Fig. 4(c) reveals the rationale at one time instant. Around a jam area with symmetric traffic density, vehicles driven by MFG controllers tend to slow down farther upstream before joining the jam and immediately speed up after leaving the jam; in contrast to those driven by LWR controllers whose speed remains symmetric before and after the jam area. This is because LWR's velocity is determined only through traffic density at that location, while that of the MFG depends on traffic density of the entire horizon.

6. n AV + m HV: controllable AVs navigating the HV-dominated traffic

As mentioned in the previous section, in this section, we add HVs into the traffic environment where multiple AVs navigate. Likewise, we will use vehicle cooperation as the primary category and the environment stochasticity as the secondary category.

6.1. *Cooperative control*

When multi-AV control is cooperative, the existing literature covers both a deterministic and a stochastic environment.

6.1.1. *A deterministic environment: Multi-class microscopic or mesoscopic traffic models*

In Section 4.1.1, stability-oriented control is discussed for one AV to stabilize multiple HVs. When the number of AVs grows, the stability of a mixed platoon highly depends on the topology of AVs and HVs including the vehicle composition and the positions of each vehicle type. How to design a controller to stabilize a mixed traffic platoon remains relatively understudied, due to the scalability issue from the stability analysis of individual vehicles to that of a platoon with multiclass traffic compositions.

To avoid enumerating various topology of a mixed platoon, a majority of studies use a general concept of head-to-tail stability in which the stability of a platoon only depends on the total numbers of AVs and HVs, not their topology (Wu et al., 2018). Using simulations, Talebpour and Mahmassani (2016), Yao et al. (2019) implement CACC on CAVs and investigated the string stability of the mixed traffic system. Different controller parameters and the CAV's penetration rates are tested to illustrate their relations to the stability.

By decomposing the entire platoon into small subsystems, Zhou et al. (2020) introduces a more practical head-to-tail stability criterion for subsystems and analyzes the mixed traffic system with multiple CAVs and multiple HVs under the new stability criterion. Gong and Du (2018) solves a p-step MPC instead of a one-step MPC to mitigate the uncertainty of human driver trajectories.

As opposed to string stability actually pertaining to the transient behavior of a platoon, there are two more formal stability notions, namely, *Lyapunov stability* and *asymptotic stability* commonly used by the control community (Di Vaio et al., 2019; Zheng et al., 2017; Zhang et al., 2020). "Lyapunov stability" refers to stability that a small perturbation of an equilibrium stays near it, while "asymptotic stability" refers to when a small perturbation of an equilibrium converges to it. To a certain degree, these two stability notions are more generic than string stability, because they require to perturb the state of all the vehicles in a platoon, while string stability only requires to perturb the velocity of the platoon leader.

To analyze the stability of a multiclass platoon, *traffic flow stability* (i.e., the magnitude of the perturbations to the density and velocity profiles around traffic equilibrium states are controlled as time increases (Darbha and Rajagopal, 1999; Ngoduy, 2013b)) is used as the counterpart of string stability for microscopic models. Both mesoscopic and macroscopic models have been employed to characterize traffic flow stability. Ngoduy et al. (2009), Ngoduy (2013b), Ngoduy (2013a) employ the gas-kinetic theory to derive the macroscopic CACC traffic flow and characterize stability diagrams using linear and/or nonlinear analysis methods. Porfyri et al. (2015) model the macroscopic traffic flow of mixed ACC and CACC vehicles and analyze how the penetration rate of ACC vehicles influences

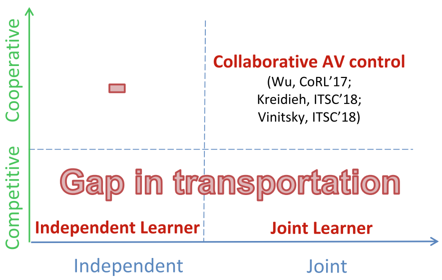


Fig. 5. Literature on MARL.

the traffic stability.

6.1.2. A stochastic environment

Wu et al. (2017a), Wu et al. (2017b), Wu et al. (2017d), Vinitsky et al. (2018) assume a fully observable system where the goal of multi-AV control is to optimize total system performances, such as velocity, energy consumption. A model-free MARL is employed. In other words, there is no need to define a state transition matrix explicitly. Instead, the state transition is computed from the simulation platform. A traffic simulator has been developed in SUMO to simulate HVs and uncontrolled AVs using IDM models. Given actions selected by controlled AVs, the simulator updates every car's position based on selected actions of controllable AVs and IDM models of uncontrollable vehicles. Then the centralized training and execution with TRPO policy gradient is implemented to solve for an optimal policy. Kreidieh et al. (2018a) trains AVs on a multi-lane ring road and implements transfer learning to execute the AV control on an open multi-lane highway. In a routing scenario on a road network, Lazar et al. (2019) trains routing policies for a fleet of centralized AVs to move from an origin to a destination using DRL, with individual human drivers moving according to cell transmission models.

6.2. Noncooperative control

6.2.1. A deterministic environment: Multi-class macroscopic traffic models

The macroscopic approach to address the scalability issues is the PDE approximation (Barooah et al., 2009; Zheng et al., 2016). This approach suggests to study the stability of continuum traffic flow models which are the limits of microscopic models. Building on MFG, Huang et al. (2019, 2020b) analyzes traffic stability for mixed traffic, assuming that HVs are modeled by ARZ and AVs are modeled by an MFG. Linear stability analysis demonstrates that the MFG traffic flow model behaves differently from traditional traffic flow models. The impact of AV's penetration rate and controller design on traffic stability are quantified on ring roads.

There are another body of literature that does not focus on stability of mixed platoons. Instead, they are more interested in the implications of multi-class traffic flow models on traffic performances. Accordingly, a multi-class LWR model has been used to capture the evolution of hybrid traffic dynamics, assuming AVs are powered by LWR models characterized by different fundamental diagrams from HVs, which can be treated as stable controllers (Levin and Boyles, 2016; Patel et al., 2016).

6.2.2. A stochastic environment

In a multi-AV system where human drivers exist and dominate the traffic environment, uncertainty arises from human driving behavior. Controllable AVs have to learn the environment while selecting optimal driving policies with a maximum reward. Built upon single-agent RL, multi-agent reinforcement learning (MARL) extends the control of single robot to multiple ones. In a multi-agent system with stochasticity and uncertainty, MARL becomes a natural tool for control of multiple AVs. MARL tasks can be broadly grouped into three categories, namely, fully cooperative, fully competitive, and a mix of the two, depending on different applications (Zhang et al., 2019a): (1) In the fully cooperative setting, agents collaborate with each other to optimize a common goal; (2) In the fully competitive setting, agents have competing goals, and the return of agents sums up to zero; (3) The mixed setting is more like a general-sum game where each agent cooperates with some agents while competes with others. For instance, in the video game *Pong*, an agent is expected to be either fully competitive if its goal is to beat its opponent or fully cooperative if its goal is to keep the ball in the game as long as possible (Tampuu et al., 2017). A progression from fully competitive to fully cooperative behavior of agents was also presented in Tampuu et al. (2017) by simply adjusting the reward.

The future AVs will be manufactured by different companies with different technical specifications. It will thus be challenging for AVs to collaborate with a common goal. We believe it is reasonable to picture that each AV is an independent and fully decentralized agent with its own goal (e.g., to send its occupant to her destination on a shortest path). Fig. 5 illustrates the classification of MARL based on if agents are collaborative (to optimize a common goal) or competitive (with competing goals) and if they share information with others (i.e., being independent or coordinate). There exists a void when multiple AVs compete for limited road resources regardless of whether AVs are independent or joint learners. The independent or joint learning algorithms will be detailed in "MALR algorithms" subsequently.

MARL algorithms. The noncooperative multi-AV control problem can be formulated as a stochastic game or Markov game (Littman, 1994), where each agent solves a POMDP. Without cooperation and with the presence of human drivers, it is challenging for AVs to sense and perceive the environment precisely. Therefore we assume a stochastic environment. Fig. 3 illustrates the multi-AV control framework. The main different from the single-AV control is the interaction among multiple AVs when they simultaneously explore the mixed traffic environment and select their individual optimal policies.

Depending on if AVs can exchange information and learn the environment information, driving policy learning can be categorized into joint or independent learners. Local observation leads to independent learners, while information sharing can change AVs' learning behavior to joint learners.

Independent Learners. If AVs only sense neighboring vehicles' information, each AV learns the environment and policies independently. Conceptually, most single-agent RL techniques can be directly applied to multi-agent scenarios for independent learners. Popular examples include DQN (Mnih et al., 2015), DDPG (Lillicrap et al., 2015) and soft actor-critic (Haarnoja et al., 2018). When the independent learner algorithms are applied to MVS, the implicit assumption is usually made that there is no interaction between two controllable AVs. In other words, at every time step, one AV only interacts with its surrounding HVs not other adaptive AVs.

Difficulties of independent learner algorithms arise from the following cases (Omidshafiei et al., 2017; Nguyen et al., 2018): (1) non-stationarity of Q-value estimation due to co-existence of other adaptive AVs; (2) invalid theoretical convergence in multi-AV scenarios because the Markovian property may not apply; (3) confusing domain stochasticity from both environments and other

AVs; and more importantly, (4) the curse of dimensionality, i.e., the search space in state and action is too large, making the learning intractable. Advanced MARL algorithms are developed to mitigate some of the aforementioned challenges. Decentralized hysteretic deep recurrent Q-networks (Dec-HDRQNs) utilizes different learning rates for different partially-observable domains (Omidshafiei et al., 2017). This approach exploits the robustness of hysteresis to non-stationarity and alter-exploration, in addition to the representational power and memory-based decision making of DRQNs. More recently, lenient DQN (Palmer et al., 2018) is proposed, with which lenient agents map state-action pairs to decaying temperature values that control the amount of leniency applied towards negative policy updates that are sampled from the experience replay. This introduces optimism in the value function update, and can facilitate cooperation in tabular fully-cooperative MARL problems.

A key challenge arises in MARL when independent agents have no knowledge of other agents, that is, the theoretical convergence guarantee is no longer applicable since the environment is no longer Markovian and stationary (Matignon et al., 2012). To fundamentally tackle this issue, one way is to exchange or share information among agents.

Joint Learners. If AVs receive global information regarding all others' state and action (e.g., via V2V/V2I), they can learn their optimal policies jointly. The performance of the agents could be better off through coordination. To the best of our knowledge, there only exist a small amount of studies (Wu et al., 2017b; Wu et al., 2017d; Wu et al., 2017a; Wu et al., 2018; Lazar et al., 2019) on multi-AV control using MARL. They assume that all controllable AVs share a common objective, which constitutes a fully observable cooperative MVS taking into account uncontrollable vehicles.

A joint learning framework suffers from the curse of dimensionality, as the agent size grows. Thus the centralized learning (i.e., based on global information) and decentralized execution (i.e., based on local observation) paradigm has become an increasingly popular paradigm for independent learners (Foerster et al., 2016; Lowe et al., 2017; Lin et al., 2018; Li et al., 2019). While training is stabilized conditioning on the information of other agents, scalability becomes a critical issue in MARL because the joint state space and joint action space grow exponentially with the number of agents. To mitigate the curse of dimensionality, mean field reinforcement learning (Yang et al., 2018; Shou and Di, 2020; Shou and Di, 2021) have become a popular technique, where the interactions within the population of agents are approximated by those between a single agent and the average effect from the overall population or neighboring agents. In this way, learning of individual agent's optimal policies depends on the population dynamics, which makes possible a scalable policy learning for achieving Nash equilibrium in multi-agent environments. Its potential to multi-AV control could be one direction to explore.

6.3. n AVs + m HV ($n \gg m$): A special case

There is little research on the AV-dominated world, partly because that it is highly likely that human drivers will adapt their driving behavior when surrounded with AVs. But it remains unclear how such behavior evolves. Different hypotheses could drive the evolution of human driving behavior toward opposite directions.

One hypothesis is that humans may gradually adapt their driving behavior in the presence of AVs and consequently develop moral hazards (Pedersen, 2001; Pedersen, 2003; Chatterjee and Davis, 2013; Chatterjee, 2016; Millard-Ball, 2016; Di et al., 2020). These speculations cannot be validated in the existing market with a too low penetration rate of AVs. Laboratory driving simulator using driving simulators could serve as a safe and effective alternative (Creech et al., 2019; Tilbury et al., 2020). In spite of the fact that participants could possibly exhibit unrealistic behaviors on a driving simulator, the value of these simulators should not be ignored for advancing our understanding of people's behavioral adaptation for a future scenario.

6.4. Revisiting Research Question Q1: AV control models

Summarizing the models described in Sections 3–6, we would like to reiterate the first research question raised in Section 2.1: *(Q1) What scalable driving policies are to control a large number of AVs in mixed traffic comprised of human drivers and uncontrollable AVs?* Which method should be used depending on the number of AVs, whether AVs cooperate or compete, and whether the traffic environment is deterministic or stochastic. Physics-based models such as stability-based control are more suitable for scenarios such as collaborate platoons, including "n AV", "n AV + 1 HV", and "n AV + m HV", and have been studied mostly for deterministic traffic environments. Game-theoretical models are more studied for two interacting vehicles such as "1 AV + 1 HV" or "1 AV + 1 AV". To scale it up to the scenario when a large number of AVs needs to be controlled, i.e., "n AV", mean-field approximation has been applied. AI-based models including deep learning and reinforcement learning are more robust for scenarios like "1 AV + m HV" and "n AV + m HV". In terms of scalability, both physics and AI-based models would suffer from curse of dimensionality when the number of AVs to be controlled increases exponentially. Thus a decentralized control scheme should be the ultimate solution. To be scalable for AI models, independent learning algorithms or mean-field approximation are options, but they are subject to stringent assumptions such as the Markovian property or homogeneous agents.

7. Data-driven policy learning

A major challenge in the study of AVs, different from other autonomous systems, is the highly dynamic, uncertain, complex environment in which it navigates. Unlike training robots in a controlled laboratory environment, training intelligent AVs requires them to interact continuously with the traffic environment to learn optimal driving policies. Such a traffic environment, primarily comprised of intelligent actors including human drivers and other uncontrollable AVs, needs to be learned from real data.

7.1. Human driving policy learning

Human movement trajectories are treated as hard safety constraints or boundaries for robots motion planning. To this end, accurate and precise models of human behavior are required to ensure safety–critical applications. Driving is a complex task. It is a sequential decision-making process with a complex mapping from the perception of neighboring traffic or the prediction of global traffic environment onto driver actions. Human driving behavior has long been studied in the transportation community. It has recently gained growing attentions from the control and robotics communities for its importance in designs of AVs that will drive alongside human drivers.

7.1.1. Dataset

There are aggregate traffic data and individual trajectory based data. Aggregate traffic data are collected from various sensors, including loop detectors (Rakha and Gao, 2010; Rakha and Crowther, 2002; Rakha and Crowther, 2003), surveillance cameras (Mao et al., 2018; Tang et al., 2017; Lu and Skabardonis, 2007), Bluetooth detection (Singer et al., 2013; Allström et al., 2014), roadside radar/LiDAR (Zhang et al., 2018b).

Emerging traffic sensors, including connected vehicles, smart phones, on-board cameras, and LiDARs, are expected to generate terabytes of streaming data daily (SAS, 2015). These new datasets would offer new opportunities to understand human driving behavior. Collecting real-time vehicle trajectory data, however, is costly and may infringe privacy, as it involves placing sensors inside individual vehicles (e.g., naturalistic driving devices continuously collecting vehicle movement information in the real traffic environment (Hecker et al., 2018a; Hecker et al., 2018b; Hammit et al., 2018; Flores et al., 2018; Zhang et al., 2018b; Zhu et al., 2018a)). Albeit lower cost, laboratory driving simulators (Sadigh et al., 2016b; Sadigh et al., 2019; Abbeel and Ng, 2011; Ziebart et al., 2008) allow only one driver to test at a time, unable to offer realistic experience of interacting with other vehicles on roads. To understand the emergent dynamics arising from human drivers requires information of all the vehicles dynamically moving in a traffic stream. By far there are only a few such public datasets.

Next Generation Simulation (NGSIM) is the mostly widely used human driver trajectory dataset. It provides all vehicle trajectories across a time span along some multi-lane highways. The shortcoming is that no camera images are recorded for each vehicle, which may limit the usage of image features for human driving policy learning.

Naturalistic data, collected while driving in the real traffic environment, provide an non-intrusive approach of personal driving data collection. The largest naturalistic dataset has been collected via the Strategic Highway Research Program (SHRP2) (NDS, 2018; McLaughlin and Hankey, 2015; Hankey et al., 2016). There were 3,400 participating vehicles instrumented with a data acquisition system recording speed, acceleration, latitude and longitude. Forward radar detects distance and speed relative to other vehicles. Four video views are also available. Such a dataset can train a driving policy using camera sensing information.

7.1.2. Physics-based model parameter calibration

Human driving behavior includes driving intent identification and prediction of internal states. Without communication among one another or via turning on signals, neither the intent nor internal states of neighboring vehicles are unknown and has to be estimated. We will first present the estimation of internal states in the car-following behavior, which is extensively studied in the transportation community, and then the prediction of driving intent.

CFMs have been extensively calibrated using a maximum likelihood approach (Hoogendoorn and Hoogendoorn, 2010), Bayesian estimation (van Hinsbergen et al., 2009, 2013; Rahman et al., 2015; Lee and Ozbay, 2009; Davis, 2017), fundamental diagram regression (Qu et al., 2015; Phegley et al., 2014), or heuristics (Ma and Abdulhai, 2002). Most of them are calibrated using a pair of leading and following vehicle trajectories. It loses the information of how perturbation in one vehicle may propagate to those far behind in the platoon, thus may not capture instability of traffic.

With a rising volume of data generated by vehicles and their sensors, the conventional traffic models cannot predict generalizable driving behaviors. Leveraging big data, researchers are able to leverage data-hungry machine learning methods to learn the policies underlying the diverse human driving behaviors. In the context of driving, states are observations of a driver's environment and actions are acceleration and steering angle. We have seen a growing body of literature characterizing driving behaviors using (deep) artificial neural networks (Khodayari et al., 2012; Panwai and Dia, 2007; Zhou et al., 2017a; Huang et al., 2018) and RL (Zhu et al., 2018b). These models aim to capture various phenomena arising from human drivers, including asymmetric behaviors, traffic oscillations.

Compared to the car-following behavior, lane-change is more challenging to estimate, partly because of intent identification. Driving intents, which are intended actions, can be represented by discrete categories, including driving straight with a constant speed or acceleration or deceleration (or lane-keeping), turning (or preparing to change lanes), and changing to its left or right lane (or lane-changing). The driving intent estimation problem is commonly modeled as a classification problem, which will be discussed in the next subsection. However, one school of researchers argue that human's unpredictability, randomness, and non-Markovian property makes it infeasible to learn true dynamics (Driggs-Campbell et al., 2017). Instead, task-specific Bayesian optimization (Bansal et al., 2017), stochastic reachable set (Driggs-Campbell et al., 2018), non-parametric driver model (Driggs-Campbell et al., 2017), and probabilistic approaches (Bouton et al., 2017) have been developed. Humans usually convey intent through motion, which plays a crucial role in social interactions (Becchio et al., 2012). Built upon such understanding, Driggs-Campbell and Bajcsy (2016) assume drivers tend to follow some nominal trajectory, given by the spatial empirical distributions on a cost map. Accordingly, the lane-change intent can be formulated as an optimal control problem (and can be reduced to an MPC control). The parameters of the control objective function are estimated using 10 subjects' 200 lane-change trajectories. Bansal et al. (2017) learns human dynamics via Bayesian optimization. The learned dynamic model is the one that achieves the best control performance for the task at hand but could be different from the true dynamic. Driggs-Campbell et al. (2017, 2018) solve a mixed integer linear program to estimate a stochastic reachable set that encapsulates the likely trajectories of human drivers intent and this model can generate trajectories that are similar to those performed by humans.

Physics-based models simplify the complex decision-making processes of human beings and may lack predictive powers due to its open-loop procedure of parameter estimation.

7.1.3. AI-based methods

Estimation of discrete human intent can be essentially formulated as a classification problem. Support vector machine (SVM) (Aoude et al., 2012), hidden Markov model (HMM) (Li et al., 2016), dynamic Bayesian Networks (Kasper et al., 2012; Bahram et al., 2016), Bayesian filtering (BF) (Li et al., 2016), random forest (Yang et al., 2019), and neural networks (Shou et al., 2020c) have been used for (online) classification of human intent. Features used for classification include driving-related features (e.g. longitudinal or lateral accelerations, longitudinal or lateral positions (Shou et al., 2020c), speed relative to traffic flow (Liu and Tomizuka, 2015; Liu and Tomizuka, 2016), steering angle, yaw rate (Li et al., 2016)), lane-related features (e.g., lane occupancy (Kasper et al., 2012), lane number, lane size, speed limit (Yang et al., 2019)), and gap-related features (e.g., distances between vehicles and merging points (Yang et al., 2019)). Which features can be obtained highly rely on sensor sources, including in-vehicle camera or LiDAR, inertial measurement unit (IMU), steering wheel angle and gas or brake pedal positions from CAN Bus, or even through V2V/V2I communication.

Once human intentions are known, the internal state of a vehicle, i.e., its future trajectory, is estimated with Gaussian mixture models (Wiest et al., 2012), dynamic Bayesian Networks (Gindele et al., 2010), or Kalman filter with parameter adaptation algorithm (Liu and Tomizuka, 2015; Liu and Tomizuka, 2016).

There is another body of literature that develops human driver models using game theory and RL (Li et al., 2018a; Albaba and Yildiz, 2019; Albaba and Yildiz, 2020). Li et al. (2018a) assumes that human drivers play a game based on hierarchical reasoning. A level-0 play ignores the interaction of other players, while a level-1 player assumes all other players are level-0 players. Similarly, a level-k player assumes that all other players act according to level-$(k-1)$ models. In other words, in a two-driver scenario, when the ego driver is level-1, the action of her opposite driver is assumed a level-0 driver whose actions are solved without accounting for the vehicle interaction. Then the ego driver selects her actions based on the fixed actions of the opposite driver. The simultaneous game is reduced to solving two optimal control problems sequentially. When the ego driver is level-2, the action of this ego driver depends on that of the opposite vehicle, which in turn depends on that of ego vehicle. This becomes an embedded game like proposed in Sadigh et al. (2016b). Albaba and Yildiz (2019) extends this work by both increasing the class of scenarios that can be modeled and verifying these models using real traffic data. Albaba and Yildiz (2020) improves the driver models by incorporating DRL and also verifies the results using real traffic data.

Table 3
Public AV related datasets (partly adapted from Zhou and Laval (2019)).

Data collection vehicle	Dataset	Purpose	Sensor setup	Location	Institute
HV	KITTI (Geiger et al., 2012)	3D object detection tracking	grayscale/color cameras, a rotating 3D laser scanner, GPS, IMU	Karlsruhe	Karlsruhe Institute of Technology
	KAIST (Choi et al., 2018)	object detection, drivable region detection, depth estimation	2 RGB & 1 thermal camera, 1 integrated GPS/IMU device	Seoul	Korea Advanced Institute of Science and Technology
	H3D (Patil et al., 2019)	3D detection, 3D multi-object tracking	GPS/IMU device, a LiDAR, 3 cameras	San Francisco	Honda Research Institute
	A2D2 (Geyer et al., 2019)	3D semantic segmentation, object detection	5 LiDARs, 5 surround cameras	Germany	Audi AG
	ApolloCar3D (Song et al., 2019)	3D car instance understanding	GPS, 2 laser scanners, 6 video cameras, a combined IMU/GNSS system, LiDARs	Various cities in China	Baidu
	nuScenes (Caesar et al., 2020)	3D detection, tracking	6 cameras, 5 radars and 1 LiDAR, IMU, GPS	Boston, Singapore	Aptiv Autonomous Mobility (Aptiv)
AV	A*3D (Pham et al., 2019)	3D object detection	2 Chameleon3 USB3 cameras, 1 Velodyne 64-beam 3D-LiDAR	Singapore	Agency for Science, Technology And Research (A*STAR)
	Argoverse (Chang et al., 2019)	3D tracking and motion forecasting	2 long-range LiDARs, 9 cameras for 360° coverage, GPS and other localization sensors	Pittsburgh, PA; Miami, PT	Argo AI and Carnegie Mellon Univ.
	Lyft L5 (Kesten et al., 2019)	perception systems, motion prediction	2 40-beam and 1 64-beam LiDARs, 360° cameras built in-house, a long-focal camera points upward	Palo Alto, CA	Lyft level 5 self-driving system
	Waymo Open (Waymo, 2019)	2/3D object detection, 2/3D tracking	1 mid-range LiDAR, 4 short-range LiDARs, 5 cameras (front and sides), IMUs	Various places in USA	Waymo self-driving cars

Vehicle behavior estimation and prediction is built upon vehicle detection and tracking that happens within one's perception system (Sivaraman and Trivedi, 2013). Vision-based or feature-based tracking is widely used to detect the presence of moving objects and associate vehicles between frames (Darms et al., 2008). These vehicle tracking techniques provide a foundation for end-to-end (or perception-to-control) training of autonomous driving policies (Amini et al., 2020).

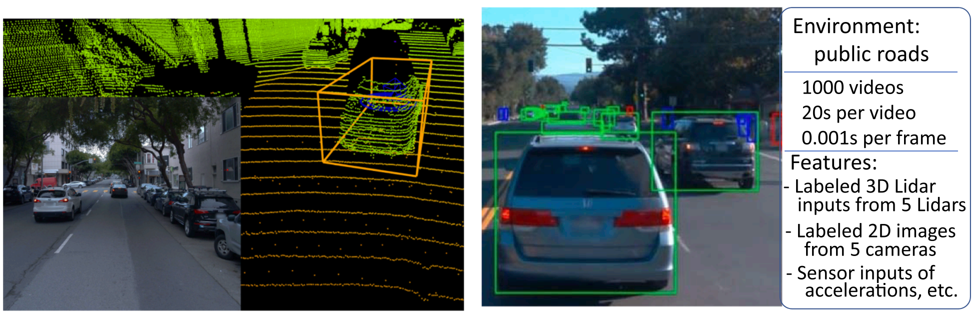
The mainstream research on human's driving policy learning is imitation learning, which will be primarily discussed subsequently.

Imitation learning. Imitation learning (IL) approaches learn the policy directly from expert demonstration data in order to behave similarly to an expert. Popular IL approaches include BC (Pomerleau, 1989; Bojarski et al., 2016; Syed and Schapire, 2008), inverse reinforcement learning (IRL) (Abbeel and Ng, 2004; Gonzalez et al., 2016; Sadigh et al., 2016a; Shou et al., 2020b), and generative adversarial imitation learning (Ho and Ermon, 2016).

In early attempts to model human driving behavior, BC formulates IL as a supervised learning problem and directly learns a mapping from states to actions using available datasets (Pomerleau, 1989). Compared to rule-based models, the advantage of these systems is that no assumptions are made about road conditions or driver behaviors. While BC approaches are conceptually sound Syed and Schapire (2008), they may fail in practice when there are states and conditions unrepresented in the dataset. As a results, even the post-trained policy model performs well on the observed states, small inaccuracies will compound resulting in cascading errors (Ross and Bagnell, 2010). In the case of driving behavior, for example, when the vehicle drifts from the center of the lane, a human driver should correct itself and move back to the centre. However, since this condition does not happen very often for human drivers, data on the correcting action is scarce, resulting in the cascading error problem, and the learned policy will continue to deviate from the center and drive off-road. Dataset aggregation (DAgger) is one popular technique to mitigate the propagation errors of BC by augmenting original training data with expert demonstration for missing states (Ross et al., 2011). Assuming human drivers follow hierarchical reasoning decision-making, Tian et al. (2019) employs DAgger to establish a mapping from the ego car's state, all others' state, and the ego car's reasoning level-k to the ego car's level-k action. To accommodate heterogeneity in human drivers, different HVs are assumed to follow different reasoning levels.

Instead of directly learning actions from observed states, IRL estimates one's underlying reward function that drives observed actions, thus avoiding the issue of missing states. Assuming that the expert follows an optimal policy with respect to an unknown reward function, IRL (Ng and Russell, 2000; Abbeel and Ng, 2004; Abbeel and Ng, 2011; Shou et al., 2020b) and its variants (Ziebart et al., 2008) have become increasingly popular to learn optimal sequential policies from expert demonstration. In general, IRL attempts to recover the reward function prior to finding the policy that behaves identically to the expert. Because the recovered reward function extends to unseen states, the corresponding policy can generalize much more efficiently and mitigate the cascading errors from which BC approaches suffer. For example, when driving on the highway, the vehicle knows to return to the centre of the lane when it is close to the side, because the reward function gives a high penalty in this situation. As to BC, due to scarce learning samples of driving at rare situations, such as driving at the side of the road, this would be a problem for BC to handle these situations. IRL has been used for modeling human driving behavior Gonzalez et al. (2016), Sadigh et al. (2016a). In particular, the reward function is specified as a linear combination of features (or a DNN) (Sadigh et al., 2019; Song et al., 2018; Bıyık and Sadigh, 2018; Abbeel and Ng, 2011). Sadigh et al. (2016b) employs a continuous-time version of IRL, which is the continuous inverse optimal control with locally optimal examples (Levine and Koltun, 2012). Schwarting et al. (2019) models heteorgenity in human drivers by introducing a social preference value into one's reward function. An online IRL learning algorithm is developed for the AV to learn such value while interacting with HVs. Despite the increasing potential in imitation learning, IRL approaches are typically computationally expensive toward recovery of the expert reward function (or cost function).

Instead of learning the expert cost function directly and learning the policy based on it, recent work has attempted to learn the expert behavior through direct policy optimization and skip the step of cost function recovery. These methods have been successfully applied to modeling human driving behavior (Ho et al., 2016). With the advent of the generative adversarial network (GAN) (Goodfellow et al., 2014a) and generative adversarial imitation learning (GAIL) (Ho and Ermon, 2016), new policy learning methods have become available, performing well on certain benchmarking tasks. GANs are based on a two-layer minimax game where one network acts as a discriminator to learn the difference between real and generated samples. The second network, i.e. the generator, is to generate fake samples to fool the discriminator. The goal is to find a Nash-equilibrium of the racing game between the generator and



(a) Camera and Lida data (b) Data description

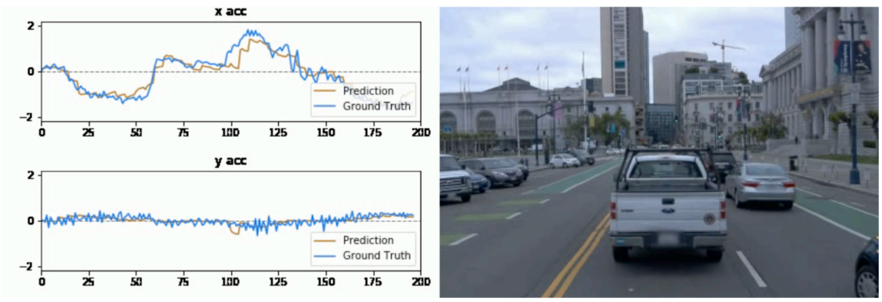
Fig. 6. Waymo self-driving car data (Li et al., 2017a).

discriminator. More recently, Wasserstain GAN (Arjovsky et al., 2017) is proposed to replace the standard KL divergence objective with Wasserstein distance, which solves the mode collapse issue in standard GAN. GANs can be extended to imitation learning domain by replacing the generator with the policy network, i.e., the action generator given states. The generator generates actions based on a learned policy, which is derived via the objective of fooling the discriminator. The discriminator distinguishes between the generated actions and expert actions given states. GAIL uses the GAN technique in combination with TRPO. TRPO updates the policy within a properly bounded region, and based on which, a monotonic improvement in policy over iterations is guaranteed (Schulman et al., 2015a). For more stable training, generalized advantage estimation (GAE) (Schulman et al., 2015b) is used to adjust variance-bias trade-off and reduce the variance in learning. TRPO combined with GAE is able to learn complicated high-dimensional control tasks (Schulman et al., 2015b). GAIL combined with recurrent policy learning (Wierstra et al., 2010; Heess et al., 2015), in particular, has been used for modeling human driving behavior, achieving advanced results (Kuefler et al., 2017). Later on, other algorithms combining GAIL with Wasserstein GAN (WGAIL), and gradient penalty (WGAIL-GP) (Gulrajani et al., 2017) are explored and show improved performances in some conditions compared o standard GAIL (Greveling, 2018).

GAIL has been used to imitate human driving behaviors (Kuefler et al., 2017; Bhattacharyya et al., 2018; Bhattacharyya et al., 2019). It however may not reflect realistic human driving behaviors. For instance, humans on a straight road would drive without maneuvering steering wheels, but the trained driving polices could alternate between small left and right wheel-turning actions,

(a) In the first frame of the video, this is a typical "car-following" scenario: the AV follows the truck steadily, and the initial value of the acceleration is around zero.

(b) In the 108th frame of the video, the truck turns left and at this moment, there is no detected front car and the AV accelerates. The ground-truth and the prediction acceleration curves climb simultaneously.



(c) In the last frame of the video, the declining acceleration curves shows that the AV brakes accordingly as the front car stops due to the traffic ahead.

Fig. 7. LSTM prediction on longitudinal and lateral accelerations (left) and video snapshot captured from one AV camera (right) (Three key frames are extracted from segment-10289507859301986274 from tar validation_0001.).

indicating instability of policies. Moreover, the newest algorithms, such as WGAIL-GP proposed in Greveling (2018), have not been thoroughly evaluated to form a well-built conclusion about the performance for modelling driving behavior, and are open to further practice in future study.

In summary, when only a small portion of states are visited in training datasets, BC suffers from cascading errors in prediction over that unseen states. IRL mitigates the cascading error issue by learning an expert's unknown reward function, because the inferred reward would provide a feedback to the learned actions generated from unseen states. However, IRL approaches are typically computationally expensive. Instead of learning reward functions, GAIL learns expert behavior through direct policy optimization (Ho et al., 2016). GAIL can extract a generalizable policy from limited driving scenarios compared to BC, and has a relatively faster learning speed compared to IRL. We believe GAIL could be one promising tool for human driving policy learning.

7.2. Autonomous driving models for uncontrollable AVs

Uncontrollable AVs refer to those AVs that interact with controllable AVs in the traffic environment but cannot be controlled, probably because they are manufactured from different companies and their driving algorithms are unknown to the host AV. Their driving behavior also needs to be learned by controllable AVs. Unfortunately, researchers' inaccessibility of AV data has greatly hindered such understanding. Due to manufacturers' proprietary protection, however, no documentation has revealed how the existing AVs are actually programmed to drive and interact with other road users on public roads. In this subsection, we strive to provide some insights into how researchers may leverage some public datasets collected for computer vision to model the driving behavior of existing AV fleets on public roads.

7.2.1. Dataset

Researchers should be very careful when they claim an AV dataset or when they need to seek some AV related data, because most public AV datasets are actually collected by HVs. We summarize a non-exhaustive list of AV datasets in Table 3. These data are collected by vehicles equipped with a variety of sensors, such as radar, LiDAR, GPS, cameras, and inertial measurement units (IMU). These sensor data, if collected from HVs, are solely used to train computer vision algorithms for object detection, segmentation, 3D tracking, pedestrian detection, and Simultaneous Localization and Mapping (SLAM). Once trained, these computer vision algorithms are mounted to AVs for testing. Fortunately, there exist several public datasets collected directly from AVs, that were pre-trained by academic institutes or AV technology companies. It would be a good strategy for academic researchers to make use of these public AV-collected datasets to learn uncontrollable AV models for simulation, which might behave similarly to existing AV fleets.

To train driving behavioral models of AVs (i.e., end-to-end driving policies), we need not only data from sensors mounted for computer vision, but also driving data directly collected from AVs' motion sensors. Thus, some of AV-collected datasets in Table 3 might not be ideal for AV policy training in their raw format. For example, A*3D and Argoverse did not provide acceleration, which requires inference using other sensor information, such as GPS. Fortunately, Waymo and Lyft datasets provide complete acceleration records, based on which AV policy training can be made.

To the best of our knowledge, Waymo/Lyft data (Waymo, 2019; Kesten et al., 2019) are the only two public datasets on how Level-5 AVs drive and interact with other road users on public roads. Both sets are composed of sensor data collected from accelerometer (i.e., IUM), camera, and LiDAR. They were originally released for the purpose of object detection and tracking algorithm development. These datasets also offer valuable insights into the AV driver models and the interaction between AVs and the environment. Such datasets are however distinct from conventional traffic data that our transportation community is used to handle and thus novel methods are required.

We discuss Waymo Open dataset as an example. As shown in Fig. 6a, camera images and LiDAR point clouds were collected when a Waymo car was driving on public roads. 2D/3D bounding box labels were included in camera/LiDAR data in each time frame. Fig. 6b illustrates the statistics of the dataset. It contains 1,000 driving videos, each with a duration of approximately 20 s. There are a total of

Table 4
Simulators for training AV policies.

Type of AV Sim.	Simulator	Features	Reference
Game-playing (vehicle sensor, environment and surrounding vehicles are fixed)	TORCS	speed, position on the road, distance to proceeding car, 3D image	Wymann et al. (2000)
	World Rally Championship 6 (WRC 6)	84x84 front view image, speed	Perot et al. (2017), Jaritz et al. (2018)
	Grand Theft Auto V (GTA V)	images from various viewpoints of vehicle	Richter et al. (2016), Johnson-Roberson et al. (2017), Richter et al. (2017)
Customizable (can configure vehicle sensor systems, environment specification, controlling of surrounding vehicles)	SUMO, VISSIM	2D top-down view of road, speed, location on roads, state info. of designated surrounding veh.	Institute of Transportation Systems (2018), PTV (2020), Ye et al. (2019)
	CARLA	GPS, speed, acceleratin, collision sensor, LiDAR, 3D images, and other sensor suites	Chen et al. (2019b), Codevilla et al. (2018)
	FLOW	Same as SUMO's features	Wu et al. (2017c), Jang et al. (2019)

200 million image frames. Built-in accelerometer sensors recorded accelerations taken by a car in each frame, making possible the retrieval of driving policies (that maps surrounding conditions to action of steering and acceleration) programmed in Waymo cars.

7.2.2. AV driving model for Waymo cars

Leveraging Waymo's sensor data, Gu et al. (2020) apply BC to learn generalizable autonomous driving polices for two reasons: First, AVs are assumed to follow the same driving policies in the same traffic environment, which is different from human drivers who behave highly heterogeneously. Second, Waymo datasets cover a wide range of traffic scenarios, including on highways or urban streets, at intersections with traffic lights or stop signs, Car-following scenarios were selected from a vast amount of Waymo video data to validate the algorithm performance. An LSTM-based learning model is trained, which takes sensor inputs from accelerometer and camera of the past ten frames and predicts acceleration for the next frame. Fig. 7 illustrates three scenarios in one video: the ego car follows a truck, the truck leaves, and another leading car decelerates. This model could be a basis to build a mixed traffic environment that captures the interactions between AVs and their sounding environment.

7.2.3. AV simulators in mixed traffic

It is crucial to validate efficiency and safety of designed AV controllers. High-risk and high-cost of real AV test urgently requires the development of a mixed traffic simulation environment for virtual testing of AVs. We want to stress that there are tons of "AV simulators" out there but their settings and purposes differ significantly. Researchers should be aware of the types of traffic simulators in the market and select the one tailored to their own purposes.

We divide the existing AV simulators into two primary types based on their purposes: one for traffic performance assessment implemented with preprogrammed physics-based CAV driving models, and the other for AV driving policy training and evaluation by creating a static or interactive environment. The former encode the already calibrated driving models (such as IDM) of each agent to simulate outcomes without further updating these driving models, while the latter require continuous interactions of the trained AV system with the environment. The transportation community is focused on the first category aiming to evaluate the impact of AVs on traffic congestion or emission, while the robotic community has widely used the second category for RL based AV policy training.

The first category of simulators include VISSIM (PTV, 2020) and AIMSUN (AIMSUN, 2020), which have accommodated AV driving modules. VISSIM, compatible to vehicle dynamics simulators including CarMaker, is able to simulate the full spectrum of vehicle automation from Level 1 to 5. Aimsun Auto can be integrated with sensor testing tools and vehicle dynamics simulation tools, such as Simcenter PreScan, These simulators are good to answer systematic questions likes the impact of various AV market penetration rates on traffic performances and tipping points, which would help operators and policy makers better assess the impact of AVs on safety, mobility, and sustainability. But they suffer from two issues: First, vehicle motion is simulated using physics-based traffic models, which limits their potential to include AI-based controls trained with high-dimensional features. Second, these simulators cannot update driving models online. Driving parameters have to be calibrated offline before running online simulation. This may prevent CAVs from adapting to the imminent traffic environment.

The second category of simulators can be further divided into two types: game-playing and customizable (see Table 4). Game-playing AV simulators are task-oriented with specified tasks for AV players to accomplish. It is not allowed to change the environment and surrounding vehicles because they are pre-programmed and fixed. TORCS (Wymann et al., 2000) is a racing simulator, which provides real-time observations like speed, position on roads, distance to proceeding car, and image. This simulator has been used for AV training in the lane-keeping and racing scenarios (Chen et al., 2015; Sallab et al., 2017; Yang et al., 2017). Another popular racing simulator currently used for learning a racing AV policy is World Rally Championship 6 (WRC 6) (Perot et al., 2017; Jaritz et al., 2018). WRC 6 provide front view image and speed information for players to control steering, brake, and gas. Compared to TORCS, WRC 6 has a more realistic physics engine. In addition to racing simulator, AV communities recently extend their work to action-adventure games, such as Grand Theft Auto V (GTA V), in which multiple vehicle-related missions need to be completed (Richter et al., 2016; Johnson-Roberson et al., 2017; Richter et al., 2017).

Customizable AV simulators provide APIs for users to design their environments, surrounding vehicles, and sensor suites at will. Once setting up the initial scenarios, the built-in engine will simulate the involved vehicles and traffic scenarios of interest. Various AV training tasks, such as object detection, lane keeping and collision avoidance, can be made on the platforms provided by customizable simulators. SUMO (Institute of Transportation Systems, 2018) is a typical example, with which users can design road networks, surrounding vehicles' policies and sensor systems. SUMO can provide state information, such as speed, location on a road, 2D top-down image of the road, and the state of other designated vehicles in the simulation. CARLA (Chen et al., 2019b; Codevilla et al., 2018) has a 3D engine, and can provide more realistic traffic simulation, though the computation is heavier. CARLA also has a flexible setup of sensor suites and signals, including GPS, collision, LiDAR, 3D images and etc. The 3D road traffic can be configured in the simulator. FLOW (Wu et al., 2017c) is another customizable simulator, which utilizes SUMO based engine. FLOW incorporates RL libraries, such as rllab and RLlib, and thus, it has convenient interface for RL developers to train and evaluate their AV policies in the simulation provided by FLOW. FLOW has recently been used to train muti-AV policies in a mixed traffic environment (Wu et al., 2017c; Jang et al., 2019).

There is another type of AV simulators that is gaining growing attention, which is the scaled-down laboratory testbed. It requires to apply the similitude theory to ensure the consistency between the scaled-down and the real-size dimensions (Di et al., 2019).

7.3. Revisiting Research Question Q2-Q3: Data-driven models for background traffic

Summarizing the models described in Sections 7.1,2,3,4,5,6,7.2, we would like to reiterate the second and the third research questions raised in Section 2.1: *(Q2) How do we estimate human driver behaviors? (Q3) How should the driving behavior of uncontrollable AVs be modeled in the environment?* Which methods to use to estimate driving behavior depend on the available data type and research goal. In contrast to the physics-based models that have been extensively calibrated using aggregate traffic data, AI-based methods are more capable of processing high-dimensional high-volume data collected by multimodal sensors, including spatio-temporal mobile trajectories, in-vehicle camera images, and LiDAR point clouds. They also tend to provide more accurate predictions of complex dynamics of driving environments. For AV driver models, we only present one model trained using Waymo data because of limited availability of Level-5 AV data from proprietary protection. Despite a high volume of sensor data supplied by Waymo, the data lacks behavioral information of AVs moving in traffic streams and thus is insufficient to characterize a comprehensive spectrum of AV driving behaviors. With little driving behavioral information, how to model Level-5 AVs as background traffic is challenging. With the increasing number of Level-1 or 2 production AVs, CAN (Controller Area Network) bus data could provide resourceful information of Level-1 or 2 AV driving models. “Data is the new oil.” The advance of data-driven behavioral learning requires partnership between the industry, agencies, and academia.

The present focus of researchers is on human driver modeling, not AV modeling, due to a too low penetration rate of AVs. However, we should vision the future when AVs are likely to encounter other AVs manufactured or programmed by different companies. When that time arrives, the understanding of how other AVs drive as background traffic would become crucial to autonomous driving.

8. AV-HV Interaction

Te characterization of AV-HV interaction is crucial to the development of AV controllers, because it not only provides feedback to AV controllers and forms physical constraints. More importantly, it dictates the social acceptance of AVs, especially when AVs are deployed at a large scale in the mixed traffic environment. Ideally how one AV interacts with HVs and other AVs should be learned from data. But the lack of such data along with an extremely low penetration rate of AVs makes this task infeasible. All the existing studies resort to theoretical modeling. We will first point out what needs to be learned if we have data and then turn our attention to modeling approaches.

1. **HV-HV:** How human drivers interact with one another has been extensively studied in the existing literature. Its characterization uses both physics-based and AI-based methods (detailed in Section 7.1).
2. **HV-AV:** How human drivers react to the presence of AVs has two directions: (1) Most studies assume that HVs drive the same way as they do in the pure HV traffic environment. In other words, even when they encounter AVs, they cannot identify AVs and interact as if AVs were HVs. (2) If HVs have the capability of identifying AVs, AVs are essentially another vehicle class and HVs may likely interact differently. On one hand, the interaction of heterogeneous vehicle classes can provide comparative studies (Ossen and Hoogendoorn, 2011). On the other hand, AVs are fundamentally different from other vehicle types propelled by human drivers and may transform humans’ car-following behavior significantly. Unfortunately, due to lack of behavioral data for HV-AV scenarios, this arena is understudied.

   A limited number of studies have all pointed out that humans’ behavioral reaction to AVs highly depends on their trust to the technology. Färber (2016) studies the importance of communications among road users, such as eye contact, gestures, or anticipatory behavior, on the safety of human drivers. Thus, the challenge arises when humans attempt to communicate but cannot get feedback from AVs. As a result, humans may not predict what will happen and behave cautiously. Dekker (2019) raises a concern that lack of local traffic culture, such as when to honk rather than light signals or yield modestly, may cause people’s distrust, and increase the risk of conflicts between human drivers and AVs. Unfortunately, most current AV training processes mainly focus on learning a general and culture-blind AV policy instead of learning how to drive like a local driver. Zhao et al. (2020) conducts a sequence of field experiments with ten recruited drivers for HV-AV scenarios. Using headways, gaps, and speed deviation, the participants are grouped into three types: AV-believers, AV-skeptics, and AV- insensitives. In other words, one’s car-following reaction to AVs highly depends on her trusts on AV technologies.

Table 5
Vehicle interaction types.

	Interaction Type		Reference	Community
Micro	Local	pairwise car-following	Talebpour and Mahmassani (2016), Cui et al. (2017), Wu et al. (2018)	Transportation
		influence by design	Sadigh et al. (2016b), Lazar et al. (2018a)	Robotics
		adversarial game	Sadigh et al. (2019)	Robotics
	Global	k-ahead-vehicle term	Jin and Orosz (2014), Qin and Orosz (2017), Jin and Orosz (2018), Jin et al. (2018)	Control
		interaction link	Li et al. (2014), Li et al. (2017a), Li et al. (2017b), Li et al. (2018d)	Transportation & control
		physical constraints	Wang et al. (2014a), Wang et al. (2014b), Gong et al. (2016), Gong and Du (2018), Zhou et al. (2017b)	Transportation
Macro	Equ.	multiclass	Levin and Boyles (2016), Patel et al. (2016), Kockelman (2017), Melson et al. (2018)	Transportation
	Non-equ.	gas-kinetic	Ngoduy et al. (2009), Ngoduy (2013b), Ngoduy (2013a)	Transportation

Table 6
Physics-based mixed traffic model summary.

Interaction scenario	Coop. or comp.	Model	AV controller	Goal	HV driving model	HV data & estimation	Traffic scenarios	Simulation	Algorithm	Reference
n AVs	Coop.	Linear controller		string stability	CACC	-	CF	numerical, field experiments	-	Schakel et al. (2010), Naus et al. (2010), Ploeg et al. (2011), Milanés et al. (2014), Milanés and Shladover (2014), Cui et al. (2017)
	Coop.	linearly constrained linear quadratic Gaussian, optimal control	Longitudinal vehicle dynamics, (serial distributed) MPC	robust local and string stability (with uncertainties)	-	-	CF	-	Sequential distributed algorithm	Zhou et al. (2017b), Zhou and Ahn (2019), Zhou et al. (2019)
	Coop.	optimal control	Longitudinal vehicle dynamics	system performance (e.g., efficiency, safety, capacity, emission)	-	-	merging, LG, roundabouts, intersections, networks	-	(bi-level) trajectory optimization	Zhou et al. (2016), Chen et al. (2020a), Sun et al. (2020), Karimi et al. (2020), Ladino and Wang (2020), Zhao et al. (2018), Yao and Li (2020), Wang et al. (2020a), Chen et al. (2020b), Lin et al. (2020)
n AVs + 1 HV	Coop.	CCC	*n*-car-ahead OVM	optimal velocity, close to uniform flow	spacing and speed feedback	sweeping least square	CF	numerical, field experiments	recursive method for LQ with distributed delay	Qin and Orosz (2013), Jin and Orosz (2014), Qin and Orosz (2017), Jin and Orosz (2018), Jin et al. (2018)
	Coop.	Optimal control	fully observable one- or p-step MPC	transient traffic smoothness, asymptotic stability	Newell	Online curve matching NGSIM data	CF	-	Dual-based distributed algorithms	Gong et al. (2016), Gong and Du (2018)
n AVs + m HV	Coop.	Linear quadratic regulator optimization	closed form	Equilibrium spacing, speed difference, acceleration rates	(Chained) asymmetric behavior model	-	CF	A 10-vehicle platoon	-	Chen et al. (2019a)
	Coop.	distributed frequency-domain-based	hierarchical control	String stability for mixed platoons	Linearized general form	CF	NGSIM	Two mixed vehicular platoons led by two HVs sampled from NGSIM in MATLAB	H-infinity control	Zhou et al. (2020)

(Abbreviation: coop. – cooperative, comp. – competitive. CF – Car-following; LC – Lane-change.).

Table 7
AI-based mixed traffic model summary.

Interaction scenario	Coop. Comp.	Model	AV controller	Reward	HV driving model	HV data & estimation	Traffic scenarios	Simulation	Control solution algorithm	Reference
1 AV + 1 HV	Comp.	two–person non-zero sum game	game	spacing, safety	game	NGSIM	LC, unprotected left-turn	numerical	simulated moments, MLE	Talebpour et al. (2015), Yu et al. (2018), Zhang et al. (2019c), Yoo and Langari (2020)
	Comp.	Two-person game	mixed-motive or Stackelberg game	spacing, safety, LC feasibility	PD controllers	NHTSA 100-care naturalistic driving safety dataset	CF, LC, merge	multi-lane highway in MATLAB Simulink and dSPACE	Bilevel evolutionary algorithm (BLEAQ)	Yoo and Langari (2012), Yoo and Langari (2013), Kim and Langari (2014), Yu et al. (2018), Coskun et al. (2019)
	Coop.	Stackelberg or hierarchical game	MDP	efficiency and safety, the AV's influence on HVs, lane-keeping	Continuous inverse optimal control	Simulated driving trajectories	LC, overtaking, merge	two-lane highway	Feedback Stackelberg dynamic program	Sadigh et al. (2016b), Sadigh et al. (2019), Fisac et al. (2019)
1 AV + m HVs	-	Hierarchical control	Stackelberg game	spacing, safety, LC feasibility	IDM or HIL	aggressiveness estimation	Mandatory LC	multi-lane highway in MATLAB Simulink	MPC	Zhang et al. (2019c), Coskun et al. (2019), Zhang et al. (2019b)
	-	DL	End-to-End controller based on BC	Similarity to experts' behavior in data	HV behaviors recorded in data	collected by a data collection car driven by a human	Driving on real roads or pre-designed obstacle-filled roads	No simulation involved	CNN, LSTM, and their combinations and variants, such as FCN-LSTM, C-LSTM	Pomerleau (1989), Muller et al. (2006), Bojarski et al. (2016), Bojarski et al. (2017), Rausch et al. (2017), Bechtel et al. (2018), Pan et al. (2018), Xu et al. (2017), Eraqi et al. (2017), Hecker et al. (2018a), Bansal et al. (2019)
	-	MDP	End-to-end controller based on DRL	cost related to collisions, location on the road, angel between vehicle and road headings, difference from desired speed, et al.	IDM, PD controllers or other pre-programmed HV in the gaming simulator	simulated	CF, LC and racing scenarios	SUMO, TORCS, World Rally Championship 6 (WRC 6)	Deep policy networks, DQNs using training methods, such as TRPO, DDPG, advantage Artor-critic, A3C	Lillicrap et al. (2015), Zhang et al. (2016), Sallab et al. (2017), Perot et al. (2017), Jaritz et al. (2018), Landolfi and Dragan (2018)
	-	(PO) MDP	MCTS + DRL	collision avoidance, keep on the road, task completion bonus and etc.	IDM, PD controllers or pre-programmed HV in the simulator	simulated	tactical LC, path planning	self-developed simulators, such as POMDPs.jl, SIMULATE function	MCTS guided by Deep policy network and value estimation	Paxton et al. (2017), Sunberg et al. (2017), Hoel et al. (2020)
	-	MDP with probabilistic guarantees	probabilistic specification expressed with	make a left turn safely and efficiently	IDM, pedestrians	-	Unsignalized intersections	-	DQN with prioritized experience replay	Bouton et al. (2017), Bouton et al. (2018)

(*continued on next page*)

Table 7 (*continued*)

Interaction scenario	Coop. Comp.	Model	AV controller	Reward	HV driving model	HV data & estimation	Traffic scenarios	Simulation	Control solution algorithm	Reference
			linear temporal logic, modified ϵ-greedy exploration policy		move at a constant speed					
n AVs	Comp., Coop.	differential game	(Multi-anticipative) ACC, rolling horizon control	safety, equilibrium, control, travel efficiency, route, lane preference, lane switch	IDM, Helly CFM		CF, LC	one-lane highway	Pontryagin's Minimum Principle	Wang et al. (2015), Wang et al. (2016)
n AVs + m HVs	Comp.	Multi-population differential game	MFG	efficiency, safety, kinetic energy	ARZ	-	CF	Numerical	Multigrid preconditioned Newton's finite difference	Huang et al. (2019), Huang et al. (2020a), Huang et al. (2020b)
	Coop.	Model-free fully observable MDP	DNN	total travel time	CTM	-	routing	simulated	PPO	Lazar et al. (2019)
	Coop.	Model-free fully observable DRL	GRU NN	close to desirable system-level velocity, collision penalty	IDM	-	CF, LC, merge	FLOW (based on SUMO)	centralized training and execution with TRPO policy gradient in state equivalence class representation, transfer learning	Wu et al. (2017b), Wu et al. (2017a), Wu et al. (2017d), Wu et al. (2018), Kreidieh et al. (2018b), Wu (2018)
	Comp.	Reactive game	Stackelberg, decision tree policies, DQN	collision, on-road, distance-to-object, safe separation, lane position, speed, effort	DAgger BO (hierarchical reasoning)	simulated or NGSIM	LC, unsignalized intersection	multi-lane highways, grid network with roundabouts in TORCS	receding-horizon optimization	Li et al. (2018a), Tian et al. (2018), Tian et al. (2019), Albaba and Yildiz (2019), Albaba and Yildiz (2020)

(Abbreviation: coop. – cooperative, comp. – competitive. CF – Car-following; LC – Lane-change.).

3. **AV-HV:** For those uncontrollable AVs, how they interact with the HV-dominated traffic depends on the autonomous driving algorithms programmed into these AVs. The existing AV data, such as Waymo open data, contains rich information of how one AV interacts with its surrounding environment. Our work (Gu et al., 2020) proposes an LSTM model to understand how an AV follows HVs.

   For those AVs that are controllable, effective mutual interaction would help AVs to effectively communicate and exchange messages with HVs and other road users. To design AVs' interaction interface is an emerging field in marriage of human factors and human–machine interaction (Vinkhuyzen and Cefkin, 2016; Müller et al., 2016; Wolf, 2016; Zhang et al., 2017).
4. **AV-AV:** For those uncontrollable AVs, with a low market penetration rate, it is less likely for an AV to encounter another AV. Thus, at this point it is difficulty to model how two AVs interact in the HV-dominated traffic environment using data-driven approaches.

   For those AVs that are controllable, AVs should be capable of communicating with one another via V2V/V2I or cloud-based technologies.

In the modeling aspect, there does not exist any formal definition of what constitutes "interactions" between AVs and HVs at a microscopic level. Here we provide an abstract definition of vehicular interactions.

Definition 8.1. **Vehicular interactions.** Broadly speaking, interactions model how one AV follows a leading vehicle, how it changes lanes, merges, diverges, dodges pedestrians and so on. Formally speaking, interactions indicate how the presence of other vehicles influences driving strategies of AVs and vice versa. The vehicular interaction can be modeled through joint states, physical constraints, coupled rewards or objective functions. It can be categorized into *local* and *global* interactions, replying on information technologies.

In a platoon of CAVs, vehicles interact either locally (e.g., the immediate leader and the follower) or globally. The local pairwise interaction between the immediate leader and the follower is captured in all CFMs, where the speed difference and headway with the immediate leader influences one's acceleration (Talebpour and Mahmassani, 2016; Cui et al., 2017; Wu et al., 2018). When V2V communication links are introduced in a platoon, additional interaction terms that reflect the speed and headway influence of far upstream leading vehicles are accounted for in CFMs by adding accelerations of k-ahead-vehicle (Jin and Orosz, 2014; Qin and Orosz, 2017; Jin and Orosz, 2018; Jin et al., 2018) or interaction links (Li et al., 2017a; Li et al., 2017b; Li et al., 2018d). These models do not consider physical collisions. To fix it, physical distance constraints must be imposed. Accordingly, Wang et al. (2014a), Wang et al. (2014b), Gong et al. (2016), Gong and Du (2018), Zhou et al. (2017b) encode these hard constraints into optimal control problems.

Researchers from control and robotics communities primarily focus on local interaction and formalize it using different tools. Sadigh et al. (2016b) assumes that AVs actions can influence HVs immediately through carefully selected reward functions. Lazar et al. (2018a) further illustrates that an "interaction-aware" AV can maximize road capacity leveraging such interaction.

Transportation researchers are more interested in the impact of microscopic AV-HV interactions on macroscopic traffic flow patterns (Chen et al., 2019a) and its implication for traffic controls in the presence of AVs (Levin and Boyles, 2016). A majority of studies use simulations, due to the complex project from micro to macro scales. Analytical studies are primarily focused on meso- and macroscopic scales. Among a large amount of studies on the multiclass LWR for the interaction between multiple types of traffic flows, Levin and Boyles (2016), Patel et al. (2016), Kockelman (2017), Melson et al. (2018) have applied it to AV-HV mixed traffic and proposed networked traffic controls. To capture the effect of communication and information sharing on traffic flow, Ngoduy et al. (2009, 2013b,a) propose a multiclass non-equilibrium gas-kinetic theory based model. Huang et al. (2019, 2020b) couple the CAV traffic flow driven by MFG and the HV flow characterized by ARZ assuming that both vehicle types observe total traffic density but CAVs can also observe the HV flow density. These models, however, may lack detailed interpretations of how two types of vehicles interact on a microscopic level. We urgently need a micro–macro analytical framework to offer insights into how microscopic interactions are designed for desirable traffic flow patterns.

We summarize the above vehicular interaction types in Table 5.

8.1. Revisiting Research Question Q4: Interaction characterization

Let us come back to the forth research question raised in Section 2.1: *(Q4) How are the interactions between human drivers and AVs characterized?* The answer is, it should really be learned from data. It is unfortunate that there does not exist sufficient data to validate how AVs will interact with a variety of road users in different traffic environments. Thus in this section we primarily focus on the modeling aspects and scientific hypotheses that employ both physics-based and AI-based models. Zhao et al. (2020) examplifies how to measure various interactions of Level-1 ACC vehicles using laboratory experiments. Modeling the interaction involved with Level-5 AVs remains unclear, unless extensive experiments are performed by AV companies on their fleets. However, as pointed by Kalra and Paddock (2016), it may take 400 years for the AV test to reach a comparable level of safety as human drivers. It is even more challenging to predict how many miles needed to make a reasonable hypothesis of interactions between AVs and other road users.

8.2. Revisiting all four research questions

When human road users are prevalent, it is important for AV controllers to understand driving intentions of road users, predict their moving trajectories, and model the impact of road users' driving actions and the consequences of the driving policies selected by AVs on others. To realistically model how humans drive and interact with AVs that have been or will be introduced to public roads, we need to apply adaptive policies learned from humans using real-time data, which is the promise of AI-based models compared to physics-based models. In summary, to better answer (Q1), we should in parallel develop a foundational understanding of (Q2-Q4) in the era of mixed autonomy.

9. Model summary

In this section, we summarize all the physics-based and AI-based models that have been introduced in the previous sections in Tables 6 and 7, respectively. While summarizing each model, we try to incorporate how each model type addresses the four research questions (Q1)-(Q4) raised in Section 2.1. The first two columns define the problem setting: The first column "Interaction scenario" lists the number of AVs and HVs; The second column "Cooperative or competitive" (*abbr.* as "Coop. or comp.") indicates if AVs cooperate or compete. Columns (3–5) define the model type and aim to answer our research question (Q1): The third column "Model" reveals the specific type of models used for vehicle interactions; The forth column "AV controller" represents the model used to depict AV driving behavior; The fifth column "Goal" (in Table 6) or "Reward" (in Table 7) states the goal or reward for AV controllers to achieve. Columns (6–7) indicate how HV traffic is modeled or estimated, aiming to answer our research question (Q2): The sixth column "HV driving model" represents the model used to depict human driving behavior who interact with AVs; The seventh column "HV data & estimation" indicates what dataset is used to calibrate HV driving model. Columns (8–9) aim to characterize the vehicular interactions and answer our research question (Q4): The eighth column "Traffic scenarios" demonstrates the driving tasks and traffic environments; The ninth column "Simulation" indicates how the AV control method is validated and what simulator is used. The tenth column "Control solution algorithm" lists the algorithm used to solve the AV control problem. The last column "Reference" lists the relevant papers in that category.

10. Conclusions and open questions

We will discuss open questions that are unanswered in the existing literature and provide several promising research directions.

10.1. Scalable multi-AV controls for social optima

There are few successful applications for MARL in autonomous driving, especially in complex multi-AV driving scenarios. Most previous research focuses on either using centralized but computationally-heavy MARL approaches for cooperative policy to achieve long-term traffic efficiency (Vinitsky et al., 2018), or applying decentralized parameter-sharing but non-cooperative techniques for collision-free driving for multiple AVs (Bhattacharyya et al., 2018). In addition, MARL is a fast evolving research area, but its application to multi-autonomous driving has lagged behind. Most researchers are still using basic deep RL algorithms such as deep Q network, which is not able to solve some complex problems with more than one AV. As having been discussed above, much more powerful MARL algorithms were developed in recent years but few of them have been applied to multi-AV tasks and traffic domain. In the sense, research is highly in demand on extending existing MARL algorithms or developing brand-new MARL for multi-AV and mixed AV-HV scenarios.

Nevertheless, AV algorithmic designers tend to program AVs for individual welfare, such as protecting occupants or selecting a fastest route selfishly, with no incentive for improved traffic performance. City planners have to regulate the behavior of AVs or their designers for social good. Such competing goals pose difficulty in upper level control imposed by planners, which has not been explored in the existing literature. A socially optimal control scheme needs to be devised for city planners to guide the autonomous driving technology toward social optima.

10.2. Human driving policy learning

Only when we begin to study AVs, have we learned that as humans, we know little about our own driving behavior.

10.2.1. Experimental design in data selection and model validation

Little research discusses the optimal experiment design for training and validation of a robust traffic model. The major issue with experimental design is the distributional shift in the stage of data selection and model validation, in particular, the distribution of training and test data as well as the setup of training and test environments.

The interactions of HVs on roads could generate emergent dynamics, i.e., traffic jam in the form of stop-and-go wave or oscillation. A driving model calibrated by one dataset may not be capable of predicting the emergent dynamics arising from another dataset. In other words, the predictive power of a driving model heavily depends on its training and test datasets. Field experiments are, of course, not just costly but highly risky to perform, which results in sparse human driver data with sample biases. Thus optimal experimental design can help collect representative training and validation datasets. What experiments to perform, what data to collect, and what data to use for behavioral learning and policy training are all unresolved questions.

The trained AV driving policies can also subject to distribution shift when the underlying distribution of the data at run is different from that in training. This problem is more prominent when offline RL algorithms are employed to train driving policies of AVs, such as centralized training and decentralized execution. Even worse, there may exist extremely dangerous traffic edge cases in the test environment that may not appear in the training environment. To address this challenge, Henaff et al. (2019) introduces an "uncertainty cost" that leverages the emerging area of uncertainty estimation for deep networks, which could be one promising direction to address the distributional shift when there is abundant observation data but letting agents to interact with the environment is prohibitively costly or risky.

10.2.2. Heterogeneity

The mixed traffic model need to account for heterogeneity of human drivers (e.g., different capabilities and risk profiles, human driving errors) and AVs (e.g. acceleration and braking capacity, as well as manufacturer's choice of risk tolerance). In particular, humans are highly heterogeneous, due to personal taste and preference, randomness or aggressiveness, and driving experience. With the same environment and information, different drivers may maneuver their cars differently. A robust model has to be able to accommodate these deviations and predict a distribution of actions that is consistent with real-world observations.

10.3. Multi-scale human–machine ecosystem modeling

The overarching goal of researchers is to understand the new traffic pattern comprised of large numbers of AVs and HVs and the systematic impact of AVs on traffic safety and efficiency. Such an understanding of the macroscopic traffic behavior should be rooted in both the microscopic behavior of AVs (Chen et al., 2019a) and evolution of the driving behavior of HVs over time (Di et al., 2020). This topic can be positioned to a broader context which is the *collective behavior of hybrid human–machine* (Rahwan et al., 2019). Thus, bridging traffic models on both the micro- and micro-scale using a multi-scale scheme needs to be understood.

A majority of existing studies on AVs are primarily focused on highways. An urban traffic environment consists traffic entities including cars, traffic lights, pedestrians, (motor) cyclists, scooters, and other road users. This multimodal mixed traffic environment will further complicate the control of AVs driving alongside various road users.

10.4. Safety guarantee and interpetability

Two long-standing issues with AI-based models are: (1) how to guarantee safety in safety–critical applications, and (2) how to interpret the resulting policies learned from deep learning models. These challenges have attracted growing attentions in the AI field.

Safety guarantee is crucial to safety–critical applications like AVs. However, AI-based models, RL in particular, could fundamentally fail to guarantee safety. Some existing studies incorporate safety as a reward component in RL (Bouton et al., 2017). But it is essentially a soft instead of hard constraint, so model checking (Bouton et al., 2018; Chen et al., 2019b) is needed as an extra step to avoid collisions. (Cheng et al., 2019) develops a controller architecture that synthesizes a model-free RL controller and a model-based controller utilizing control barrier functions, which is shown to guarantee safety with high probability.

Interpretability of an AI model lies on whether the model provides an explanation for the process that leads the model from its inputs to outputs (Xie et al., 2020). Without a satisfactory interpretability, an AI model's recommendation or decision-making can hardly be trusted to be safe. One category of studies, namely, intrinsic methods, enhance the interpretability of deep neural networks (DNN) by adding explainable components to the DNN structure or designing interpretable components in the learning loss. This type of methods is termed as intrinsic because it tries to make the DNN inherently and structurally interpretable in the design phase Ras et al. (2018). An increasingly popular direction in this category is a hybrid paradigm that integrates both model-driven and data-driven components, which is physics-informed deep learning (PIDL). The existing traffic models would provide some prior knowledge and help constrain the admissible solutions of AI approaches. Since its inception (Raissi, 2018; Raissi and Karniadakis, 2018), PIDL has become an increasingly popular tool in scientific and engineering areas (Yang and Perdikaris, 2019; Raissi et al., 2019; Fang and Zhan, 2020). PIDL has also shown its predictive robustness with smaller datasets in both traffic state estimation Shi et al. (2021a, 2021b) and longitudinal acceleration prediction Mo et al. (2021). The application of PIDL to data-driven solution of traffic dynamical equations or system identification of driving behavior, however, is a largely unexploited area. We believe PIDL models hold the potential not only to extract information from data, but also discover physical principles and structures underlying the data, which would help further generate and design complex and novel autonomous systems that comply with physical rules.

10.5. A Pathway to Artificial General Intelligence (AGI)

There is still a long way to reach AGI, which is the ultimate intelligence of machines, AVs need to reach human-level AIG, with the capabilities of reasoning, knowledge representation, planning, learning, communicating in natural language, and integration of all these skills towards common goals (Hodson, 2020). This not only requires bridging gaps with AI tools, but also a convergence of engineering, cognitive science, and social science. If achieved, it is not only a breakthrough to AV controls, but also to humanity.

In summary, with a rapid growing AV fleet on public roads, it is crucial to develop analytical tools for mixed traffic, which will help traffic engineers better understand the impact of AVs on transportation system performances, for the AV industry to develop a scalable autonomous driving control algorithm, and ultimately, for city planners, policymakers, and lawmakers to manage AVs for social good.

CRediT authorship contribution statement

Xuan Di: Visualization, Writing - review & editing, Supervision. **Rongye Shi:** Visualization, Writing.

Acknowledgments

The authors would like to thank Data Science Institute from Columbia University for providing a seed grant for this research. This work is also partially supported by the National Science Foundation CAREER under award number CMMI-1943998 and Amazon AWS Machine Learning Research Award Gift (#US3085926). The first author would also like to thank Prof. Ketan Savla from University of

Southern California for his inspiring comments on various notions of stability used in the transportation and control communities. The authors would also like to thank the three reviewers for their valuable comments to improve readability of this paper.

References

Abbeel, P., Ng, A.Y., 2004. Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the Twenty-first International Conference on Machine Learning (ICML), p. 1.

Abbeel, P., Ng, A.Y., 2011. Inverse reinforcement learning. In: Encyclopedia of Machine Learning. Springer, pp. 554–558.

AIMSUN, 2020. Simulate a driverless future. https://www.aimsun.com/aimsun-auto/, [Online; accessed 12.31.2020].

Albaba, B.M., Yildiz, Y., 2019. Modeling cyber-physical human systems via an interplay between reinforcement learning and game theory. Ann. Rev. Control 48, 1–21.

Albaba, B.M., Yildiz, Y., 2020. Driver modeling through deep reinforcement learning and behavioral game theory. arXiv preprint arXiv:2003.11071.

Allström, A., Bayen, A.M., Fransson, M., Gundlegård, D., Patire, A.D., Rydergren, C., Sandin, M., 2014. Calibration framework based on bluetooth sensors for traffic state estimation using a velocity based cell transmission model. Transpor. Res. Procedia 3, 972–981.

Altan, O.D., Wu, G., Barth, M.J., Boriboonsomsin, K., Stark, J.A., 2017. Glidepath: Eco-friendly automated approach and departure at signalized intersections. IEEE Trans. Intell. Vehicles 2 (4), 266–277.

Amini, A., Gilitschenski, I., Phillips, J., Moseyko, J., Banerjee, R., Karaman, S., Rus, D., 2020. Learning robust control policies for end-to-end autonomous driving from data-driven simulation. IEEE Robot. Automat. Lett. 5 (2), 1143–1150.

Amoozadeh, M., Raghuramu, A., Chuah, C.-N., Ghosal, D., Zhang, H.M., Rowe, J., Levitt, K., 2015. Security vulnerabilities of connected vehicle streams and their impact on cooperative driving. IEEE Commun. Mag. 53 (6), 126–132.

Aoude, G.S., Desaraju, V.R., Stephens, L.H., How, J.P., 2012. Driver behavior classification at intersections and validation on large naturalistic data set. IEEE Trans. Intell. Transp. Syst. 13 (2), 724–736.

Arefizadeh, S., Talebpour, A., 2018. A platooning strategy for automated vehicles in the presence of speed limit fluctuations. Transp. Res. Rec. 2672 (20), 154–161.

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. In: International Conference on Machine Learning (ICML), pp. 214–223.

Aung, A.M., Fadila, Y., Gondokaryono, R., Gonzalez, L., 2017. Building robust deep neural networks for road sign detection. arXiv preprint arXiv:1712.09327.

Aw, A., Rascle, M., 2000. Resurrection of second order models of traffic flow. SIAM J. Appl. Mathe. 60 (3), 916–938.

Bahram, M., Hubmann, C., Lawitzky, A., Aeberhard, M., Wollherr, D., 2016. A combined model-and learning-based framework for interaction-aware maneuver prediction. IEEE Trans. Intell. Transp. Syst. 17 (6), 1538–1550.

Bansal, M., Krizhevsky, A., Ogale, A., 2019. Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. In: Robotics: Science and Systems.

Bansal, S., Calandra, R., Xiao, T., Levine, S., Tomlin, C.J., 2017. Goal-driven dynamics learning via bayesian optimization. arXiv preprint arXiv:1703.09260.

Barooah, P., Mehta, P.G., Hespanha, J.P., 2009. Mistuning-based control design to improve closed-loop stability margin of vehicular platoons. IEEE Trans. Autom. Control 54 (9), 2100–2113.

Becchio, C., Manera, V., Sartori, L., Cavallo, A., Castiello, U., 2012. Grasping intentions: from thought experiments to empirical evidence. Front. Human Neurosci. 6, 117.

Bechtel, M.G., McEllhiney, E., Kim, M., Yun, H., 2018. DeepPicar: A low-cost deep neural network-based autonomous car. In: IEEE 24th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), pp. 11–21.

Behzadan, V., Munir, A., 2017. Vulnerability of deep reinforcement learning to policy induction attacks. In: International Conference on Machine Learning and Data Mining in Pattern Recognition.

Bevly, D., Cao, X., Gordon, M., Ozbilgin, G., Kari, D., Nelson, B., Woodruff, J., Barth, M., Murray, C., Kurt, A., et al., 2016. Lane change and merge maneuvers for connected and automated vehicles: A survey. IEEE Trans. Intell. Vehicles 1 (1), 105–120.

Bhattacharyya, R.P., Phillips, D.J., Liu, C., Gupta, J.K., Driggs-Campbell, K., Kochenderfer, M.J., 2019. Simulating emergent properties of human driving behavior using multi-agent reward augmented imitation learning. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, pp. 789–795.

Bhattacharyya, R.P., Phillips, D.J., Wulfe, B., Morton, J., Kuefler, A., Kochenderfer, M.J., 2018. Multi-agent imitation learning for driving simulation. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 1534–1539.

Bıyık, E., Sadigh, D., 2018. Batch active preference-based learning of reward functions. arXiv preprint arXiv:1810.04303.

Bogue, R., 2008. Swarm intelligence and robotics. Ind. Robot: Int. J. 35 (6), 488–495.

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

Bojarski, M., Yeres, P., Choromanska, A., Choromanski, K., Firner, B., Jackel, L., Muller, U., 2017. Explaining how a deep neural network trained with end-to-end learning steers a car. arXiv preprint arXiv:1704.07911.

Bouton, M., Cosgun, A., Kochenderfer, M.J., 2017. Belief state planning for autonomously navigating urban intersections. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 825–830.

Bouton, M., Karlsson, J., Nakhaei, A., Fujimura, K., Kochenderfer, M.J., Tumova, J., 2018. Reinforcement learning with probabilistic guarantees for autonomous driving. In: Workshop on Safety Risk and Uncertainty in Reinforcement Learning, https://drive.google.com/open?id=1d2tl4f6GQgH1SERveTmPAR42bMVwHZAZ.

Brown, N., Sandholm, T., Jan. 2018. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. Science 359 (6374), 418–424.

Brown, N., Sandholm, T., Aug. 2019. Superhuman AI for multiplayer poker. Science 365 (6456), 885–890.

Browne, C.B., Powley, E., Whitehouse, D., Lucas, S.M., Cowling, P.I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., Colton, S., 2012. A survey of monte carlo tree search methods. IEEE Trans. Comput. Intell. AI Games 4 (1), 1–43.

Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O., 2020. nuScenes: A multimodal dataset for autonomous driving. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11621–11631.

Cao, Y., Wang, N., Xiao, C., Yang, D., Fang, J., Yang, R., Chen, Q.A., Liu, M., Li, B., 2020. Demonstration: 3d adversarial object against msf-based perception in autonomous driving. In: Proceedings of the 3rd Conference on Machine Learning and Systems.

Chang, M.-F., Lambert, J., Sangkloy, P., Singh, J., Bak, S., Hartnett, A., Wang, D., Carr, P., Lucey, S., Ramanan, D., Hays, J., 2019. Argoverse: 3D tracking and forecasting with rich maps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8748–8757.

Chatterjee, I., 2016. Understanding driver contributions to rear-end crashes on congested freeways and their implications for future safety measures. Ph.D. thesis. University of Minnesota.

Chatterjee, I., Davis, G., 2013. Evolutionary game theoretic approach to rear-end events on congested freeway. Transport. Res. Rec.: J. Transport. Res. Board 2386, 121–127.

Chen, C., Seff, A., Kornhauser, A., Xiao, J., 2015. Deepdriving: Learning affordance for direct perception in autonomous driving. In: IEEE International Conference on Computer Vision, pp. 2722–2730.

Chen, D., Srivastava, A., Ahn, S., Li, T., 2019a. Traffic dynamics under speed disturbance in mixed traffic with automated and non-automated vehicles. Transport. Res. Part C: Emerg. Technol.

Chen, J., Yuan, B., Tomizuka, M., 2019b. Deep imitation learning for autonomous driving in generic urban scenarios with enhanced safety. In: arXiv preprint arXiv:1903.00640.

Chen, J., Yuan, B., Tomizuka, M., 2019c. Model-free deep reinforcement learning for urban autonomous driving. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, pp. 2765–2771.

Chen, N., van Arem, B., Alkim, T., Wang, M., 2020a. A hierarchical model-based optimization control approach for cooperative merging by connected automated vehicles. IEEE Trans. Intell. Transp. Syst.

Chen, X., Li, M., Lin, X., Yin, Y., He, F., 2020b. Rhythmic control of automated traffic–part i: Concept and properties at isolated intersections. arXiv preprint arXiv: 2010.04322.

Chen, Y.Q., Wang, Z., 2005. Formation control: a review and a new consideration. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 3181–3186.

Chen, Z., He, F., Yin, Y., Du, Y., 2017. Optimal design of autonomous vehicle zones in transportation networks. Transport. Res. Part B: Methodol. 99, 44–61.

Chen, Z., He, F., Zhang, L., Yin, Y., 2016. Optimal deployment of autonomous vehicle lanes with endogenous market penetration. Transport. Res. Part C: Emerg. Technol. 72, 143–156.

Cheng, R., Orosz, G., Murray, R.M., Burdick, J.W., 2019. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. Proceedings of the AAAI Conference on Artificial Intelligence 33, 3387–3395.

Choi, Y., Kim, N., Hwang, S., Park, K., Yoon, J.S., An, K., Kweon, I.S., 2018. KAIST multi-spectral day/night data set for autonomous and assisted driving. IEEE Trans. Intell. Transp. Syst. 19 (3), 934–948.

Codevilla, F., Miiller, M., López, A., Koltun, V., Dosovitskiy, A., 2018. End-to-end driving via conditional imitation learning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 1–9.

Colombaroni, C., Fusco, G., 2014. Artificial neural network models for car following: experimental analysis and calibration issues. J. Intell. Transport. Syst. 18 (1), 5–16.

Coskun, S., Zhang, Q., Langari, R., 2019. Receding horizon markov game autonomous driving strategy. In: 2019 American Control Conference (ACC). IEEE, pp. 1367–1374.

Creech, C., Tilbury, D., Yang, J., Pradhan, A., Tsui, K., Robert, L., et al., 2019. Pedestrian trust in automated vehicles: Role of traffic signal and av driving behavior. Front. Robot. AI (Forthcoming).

Cui, S., Seibold, B., Stern, R., Work, D.B., 2017. Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 1336–1341.

Darbha, S., Rajagopal, K., 1999. Intelligent cruise control systems and traffic flow stability. Transport. Res. Part C: Emerg. Technol. 7 (6), 329–352.

Darms, M., Rybski, P., Urmson, C., 2008. Classification and tracking of dynamic objects with multiple sensors for autonomous driving in urban environments. In: 2008 IEEE Intelligent Vehicles Symposium. IEEE, pp. 1197–1202.

Davis, G.A., 2017. Bayesian estimation of drivers' gap selections and reaction times in left-turning crashes from event data recorder pre-crash data. Tech. rep., SAE Technical Paper.

Dekker, J., 2019. The problem with autonomous cars that no ones talking about. https://www.fastcompany.com/90392131/the-problem-with-autonomous-cars-that-no-ones-talking-about.

Delis, A.I., Nikolos, I.K., Papageorgiou, M., 2016. Simulation of the penetration rate effects of ACC and CACC on macroscopic traffic dynamics. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 336–341.

Di, X., Chen, X., Talley, E., 2020. Liability design for autonomous vehicles and human-driven vehicles: A hierarchical game-theoretic approach. Transp. Res. Part C 118 (102710).

Di, X., Liu, H.X., Zhu, S., Levinson, D.M., 2017. Indifference bands for boundedly rational route switching. Transportation 44 (5), 1169–1194.

Di, X., Zhao, Y., Huang, S.H., Liu, H.X., 2019. A similitude theory for modeling traffic flow dynamics. IEEE Trans. Intell. Transport. Syst. 20 (3), 900–911.

Di Vaio, M., Fiengo, G., Petrillo, A., Salvi, A., Santini, S., Tufo, M., 2019. Cooperative shock waves mitigation in mixed traffic flow environment. IEEE Trans. Intell. Transp. Syst. 20 (12), 4339–4353.

Djehiche, B., Tcheukam, A., Tembine, H., 2016. Mean-field-type games in engineering. arXiv preprint arXiv:1605.03281.

Douceur, J.R., 2002. The sybil attack. In: International Workshop on Peer-to-peer Systems. Springer, pp. 251–260.

Dresner, K.M., Stone, P., 2007. Sharing the road: Autonomous vehicles meet human drivers. IJCAI. 7, 1263–1268.

Dreves, A., Gerdts, M., 2018. A generalized Nash equilibrium approach for optimal control problems of autonomous cars. Opt. Control Appl. Methods 39 (1), 326–342.

Driggs-Campbell, K., Bajcsy, R., 2016. Communicating intent on the road through human-inspired control schemes. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 3042–3047.

Driggs-Campbell, K., Dong, R., Bajcsy, R., 2018. Robust, informative human-in-the-loop predictions via empirical reachable sets. IEEE Trans. Intell. Vehicles.

Driggs-Campbell, K., Govindarajan, V., Bajcsy, R., 2017. Integrating intuitive driver models in autonomous planning for interactive maneuvers. IEEE Trans. Intell. Transp. Syst. 18 (12), 3461–3472.

Eraqi, H.M., Moustafa, M.N., Honer, J., 2017. End-to-end deep learning for steering autonomous vehicles considering temporal dependencies. In: Machine Learning for Intelligent Transportation Systems Workshop in the 31st Conference on Neural Information Processing Systems.

Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Tramèr, F., Prakash, A., Kohno, T., Song, D., 2018. Physical adversarial examples for object detectors. In: Proceedings of the 12th USENIX Conference on Offensive Technologies.

Fang, Z., Zhan, J., 2020. Physics-informed neural network framework for partial differential equations on 3D surfaces: Time independent problems. IEEE Access.

Fernandez Fisac, J., 2019. Game-theoretic safety assurance for human-centered robotic systems. Ph.D. thesis. UC Berkeley.

Fisac, J.F., Bronstein, E., Stefansson, E., Sadigh, D., Sastry, S.S., Dragan, A.D., 2019. Hierarchical game-theoretic planning for autonomous vehicles. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, pp. 9590–9596.

Flores, C., Merdrignac, P., de Charette, R., Navas, F., Milanés, V., Nashashibi, F., 2018. A cooperative car-following/emergency braking system with prediction-based pedestrian avoidance capabilities. IEEE Trans. Intell. Transp. Syst. 99, 1–10.

Foerster, J.N., Assael, Y.M., de Freitas, N., Whiteson, S., 2016. Learning to Communicate with Deep Multi-agent Reinforcement Learning. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16. Curran Associates Inc., USA, pp. 2145–2153, event-place: Barcelona, Spain.

Färber, B., 2016. Communication and communication problems between autonomous vehicles and human drivers. In: Autonomous Driving. Springer, pp. 125–144.

Garip, M.T., Gursoy, M.E., Reiher, P., Gerla, M., 2015. Congestion attacks to autonomous cars using vehicular botnets. In: NDSS Workshop on Security of Emerging Networking Technologies (SENT), San Diego, CA.

Ge, J.I., Orosz, G., 2014. Dynamics of connected vehicle systems with delayed acceleration feedback. Transp. Res. Part C 46, 46–64.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the KITTI vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3354–3361.

Geyer, J., Kassahun, Y., Mahmudi, M., Ricou, X., Durgesh, R., Chung, A., Hauswald, L., Pham, V., Muhlegg, M., Dorn, S., Fernandez, T., 2019. A2D2: Aev autonomous driving dataset. https://www.a2d2.audi.

Gindele, T., Brechtel, S., Dillmann, R., 2010. A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments. In: 13th International IEEE Conference on Intelligent Transportation Systems. IEEE, pp. 1625–1631.

Gipps, P.G., 1981. A behavioural car-following model for computer simulation. Transport. Res. Part B: Methodol. 15 (2), 105–111.

Gong, S., Du, L., 2018. Cooperative platoon control for a mixed traffic flow including human drive vehicles and connected and autonomous vehicles. Transport. Res. Part B: Methodol. 116, 25–61.

Gong, S., Shen, J., Du, L., 2016. Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon. Transport. Res. Part B: Methodol. 94, 314–334.

Gonzalez, D.S., Dibangoye, J.S., Laugier, C., 2016. High-speed highway scene prediction based on driver models learned from demonstrations. In: Proceedings of the 19th IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 149–155.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014a. Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680.

Goodfellow, I.J., Shlens, J., Szegedy, C., 2014b. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

Greenshields, B., Channing, W., Miller, H., et al., 1935. A study of traffic capacity. In: Highway research board proceedings. Vol. 1935. National Research Council (USA), Highway Research Board.

Greveling, D.P., 2018. Modelling human driving behaviour using Generative Adversarial Networks. MS Thesis. University of Groningen.
Grigorescu, S., Trasnea, B., Cocias, T., Macesanu, G., 2020. A survey of deep learning techniques for autonomous driving. J. Field Robot. 37 (3), 362–386.
Gu, Z., Li, Z., Di, X., Shi, R., 2020. An LSTM-based autonomous driving model using a Waymo open dataset. Appl. Sci. 10 (6), 2046.
Guéant, O., Lasry, J.-M., Lions, P.-L., 2011. Mean field games and applications. In: Paris-Princeton Lectures on Mathematical Finance 2010. Springer, pp. 205–266.
Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courvill, A.C., 2017. Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems (NIPS), pp. 5767–5777.
Haarnoja, T., Zhou, A., Abbeel, P., Levine, S., 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In: International Conference on Machine Learning, pp. 1856–1865.
Hammit, B.E., Ghasemzadeh, A., James, R.M., Ahmed, M.M., Young, R.K., 2018. Evaluation of weather-related freeway car-following behavior using the shrp2 naturalistic driving study database. Transport. Res. Part F: Traffic Psychol. Behav. 59, 244–259.
Hankey, J.M., Perez, M.A., McClafferty, J.A., 2016. Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets. Tech. rep., Virginia Tech Transportation Institute.
Hao, P., Wu, G., Boriboonsomsin, K., Barth, M.J., 2018. Eco-approach and departure (ead) application for actuated signals in real-world traffic. IEEE Trans. Intell. Transp. Syst. 20 (1), 30–40.
He, D., Zeadally, S., Xu, B., Huang, X., 2015. An efficient identity-based conditional privacy-preserving authentication scheme for vehicular ad hoc networks. IEEE Trans. Inf. Forensics Secur. 10 (12), 2681–2691.
Hecker, S., Dai, D., Van Gool, L., 2018a. End-to-end learning of driving models with surround-view cameras and route planners. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 435–453.
Hecker, S., Dai, D., Van Gool, L., 2018b. Learning driving models with a surround-view camera system and a route planner. ArXiv e-prints.
Heess, N., Hunt, J.J., Lillicrap, T.P., Silver, D., 2015. Memory-based control with recurrent neural networks. In: arXiv preprint arXiv:1512.04455.
Henaff, M., Canziani, A., LeCun, Y., 2019. Model-predictive policy learning with uncertainty regularization for driving in dense traffic. arXiv preprint arXiv: 1901.02705.
Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A.M., 2010. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. Transport. Res. Part C: Emerg. Technol. 18 (4), 568–583.
Ho, J., Ermon, S., 2016. Generative adversarial imitation learning. In: Advances in Neural Information Processing Systems (NIPS). pp. 4565–4573.
Ho, J., Gupta, J., Ermon, S., 2016. Model-free imitation learning with policy optimization. In: Proceedings of International Conference on Machine Learning (ICML), pp. 2760–2769.
Hodson, H., 2020. Deepmind and google: the battle to control artificial intelligence. https://www.1843magazine.com/features/deepmind-and-google-the-battle-to-control-artificial-intelligence/, [Online; accessed 12.31.2020].
Hoel, C.-J., Driggs-Campbell, K., Wolff, K., Laine, L., Kochenderfer, M.J., 2020. Combining planning and deep reinforcement learning in tactical decision making for autonomous driving. IEEE Trans. Intell. Vehicles 5 (2), 294–305.
Hoogendoorn, S., 1999. Multiclass continuum modelling of multilane traffic flow. trail thesis series.
Hoogendoorn, S., HL Bovy, P., 2003. Simulation of pedestrian flows by optimal control and differential games. Opt. Control Appl. Methods 24 (3), 153–172.
Hoogendoorn, S.P., Bovy, P.H., 2001. Generic gas-kinetic traffic systems modeling with applications to vehicular traffic flow. Transport. Res. Part B: Methodol. 35 (4), 317–336.
Hoogendoorn, S.P., Hoogendoorn, R., 2010. Generic calibration framework for joint estimation of car-following models by using microscopic data. Transp. Res. Rec. 2188 (1), 37–45.
Huang, K., Di, X., Du, Q., Chen, X., 2019. Stabilizing traffic via autonomous vehicles: A continuum mean field game approach. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, pp. 3269–3274.
Huang, K., Di, X., Du, Q., Chen, X., 2020a. A game-theoretic framework for autonomous vehicles velocity control: Bridging microscopic differential games and macroscopic mean field games. Disc. Contin. Dynam. Syst. - Series B 25 (12), 4869–4903.
Huang, K., Di, X., Du, Q., Chen, X., 2020b. Scalable traffic stability analysis in mixed-autonomy using continuum models. Transport. Res. Part C: Emerg. Technol. 111, 616–630.
Huang, M., Malhamé, R.P., Caines, P.E., et al., 2006. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. Commun. Informat. Syst. 6 (3), 221–252.
Huang, S.H., Papernot, N., Goodfellow, I.J., Duan, Y., Abbeel, P., 2017. Adversarial attacks on neural network policies. In: International Conference on Learning Representations Workshop Track Proceedings.
Huang, X., Sun, J., Sun, J., 2018. A car-following model considering asymmetric driving behavior based on long short-term memory neural networks. Transport. Res. Part C: Emerg. Technol. 95, 346–362.
Institute of Transportation Systems, D., 2018. Sumo (simulation of urban mobility). https://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/, [Online; accessed 12.31.2020].
International Standard J3016, S., 2016. SAE international.
Ioannou, P.A., Chien, C.-C., 1993. Autonomous intelligent cruise control. IEEE Trans. Vehicular Technol. 42 (4), 657–672.
Jang, K., Vinitsky, E., Chalaki, B., Remer, B., Beaver, L., Malikopoulos, A.A., Bayen, A., 2019. Simulation to scaled city: zero-shot policy transfer for traffic control via autonomous vehicles. In: Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems, pp. 291–300.
Jaritz, M., Charette, R.D., Toromanoff, M., Perot, E., Nashashibi, F., 2018. End-to-end race driving with deep reinforcement learning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 2070–2075.
Jin, I.G., Avedisov, S.S., He, C.R., Qin, W.B., Sadeghpour, M., Orosz, G., 2018. Experimental validation of connected automated vehicle design among human-driven vehicles. Transport. Res. Part C: Emerg. Technol. 91, 335–352.
Jin, I.G., Orosz, G., 2014. Dynamics of connected vehicle systems with delayed acceleration feedback. Transport. Res. Part C: Emerg. Technol. 46, 46–64.
Jin, I.G., Orosz, G., 2018. Connected cruise control among human-driven vehicles: Experiment-based parameter estimation and optimal control design. Transport. Res. Part C: Emerg. Technol. 95, 445–459.
Johnson-Roberson, M., Barto, C., Mehta, R., Sridhar, S.N., Rosaen, K., Vasudevan, R., 2017. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?. In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 746–753.
Kalra, N., Paddock, S.M., 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? Transport. Res. Part A: Policy Practice 94, 182–193.
Karimi, M., Roncoli, C., Alecsandru, C., Papageorgiou, M., 2020. Cooperative merging control via trajectory optimization in mixed vehicular traffic. Transport. Res. Part C: Emerg. Technol. 116, 102663.
Kasai, M., Shibagaki, S., Terabe, S., 2013. Application of hierarchical bayesian estimation to calibrating a car-following model with time-varying parameters. In: 2013 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 870–875.
Kasper, D., Weidl, G., Dang, T., Breuel, G., Tamke, A., Wedel, A., Rosenstiel, W., 2012. Object-oriented bayesian networks for detection of lane change maneuvers. IEEE Intell. Transp. Syst. Mag. 4 (3), 19–31.
Katrakazas, C., Quddus, M., Chen, W.-H., Deka, L., 2015. Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. Transport. Res. Part C: Emerg. Technol. 60, 416–442.
Kesten, R., Usman, M., Houston, J., Pandya, T., Nadhamuni, K., Ferreira, A., Yuan, M., Low, B., Jain, A., Ondruska, P., Omari, S., Shah, S., Kulkarni, A., Kazakova, A., Tao, C., Platinsky, L., Jiang, W., Shet, V., 2019. Lyft Level 5 AV dataset. https://self-driving.lyft.com/level5/data.
Kesting, A., Treiber, M., Helbing, D., 2010. Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. Philosoph. Trans. Roy. Soc. London A: Mathe., Phys. Eng. Sci. 368 (1928), 4585–4605.

Khodayari, A., Ghaffari, A., Kazemi, R., Braunstingl, R., 2012. A modified car-following model based on a neural network model of the human driver effects. IEEE Trans. Syst., Man, Cybernet.-Part A: Syst. Humans 42 (6), 1440–1449.

Kim, C., Langari, R., 2014. Game theory based autonomous vehicles operation. Int. J. Veh. Des. 65 (4), 360–383.

Kiourti, P., Wardega, K., Jha, S., Li, W., 2019. TrojDRL: Trojan attacks on deep reinforcement learning agents. arXiv preprint arXiv:1903.06638.

Kiran, B.R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A.A.A., Yogamani, S., Pérez, P., 2020. Deep reinforcement learning for autonomous driving: A survey. arXiv preprint arXiv:2002.00444.

Kockelman, K., 2017. An Assessment of Autonomous Vehicles: Traffic Impacts and Infrastructure Needs–final Report. Center for Transportation Research, The University of Texas at Austin.

Kreidieh, A., Wu, C., Bayen, A.M., 2018a. Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning. In: 21st IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 1475–1480.

Kreidieh, A.R., Wu, C., Bayen, A.M., 2018b. Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1475–1480.

Kuefler, A., Morton, J., Wheeler, T., Kochenderfer, M., 2017. Imitating driver behavior with generative adversarial networks. In: Proceedings of 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 204–211.

Kumagai, T., Akamatsu, M., 2006. Prediction of human driving behavior using dynamic bayesian networks. IEICE Trans. Informat. Syst. 89 (2), 857–860.

Kumar, P., Perrollaz, M., Lefevre, S., Laugier, C., 2013. Learning-based approach for online lane change intention prediction. In: Proceedings of 2013 IEEE Intelligent Vehicles Symposium (IV), pp. 797–802.

Lachapelle, A., Salomon, J., Turinici, G., 2010. Computation of mean field equilibria in economics. Mathe. Models Methods Appl. Sci. 20 (04), 567–588.

Lachapelle, A., Wolfram, M.-T., 2011. On a mean field game approach modeling congestion and aversion in pedestrian crowds. Transport. Res. Part B: Methodol. 45 (10), 1572–1589.

Ladino, A., Wang, M., 2020. A dynamic game formulation for cooperative lane change strategies at highway merges. In: IFAC World Congress 2020.

Landolfi, N.C., Dragan, A.D., 2018. Social cohesion in autonomous driving. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 8118–8125.

Lasry, J.-M., Lions, P.-L., 2007. Mean field games. Japanese J. Mathe. 2 (1), 229–260.

Lazar, D.A., Bıyık, E., Sadigh, D., Pedarsani, R., 2019. Learning how to dynamically route autonomous vehicles on shared roads. arXiv preprint arXiv:1909.03664.

Lazar, D.A., Chandrasekher, K., Pedarsani, R., Sadigh, D., 2018a. Maximizing road capacity using cars that influence people. arXiv preprint arXiv:1807.04414.

Lazar, D.A., Pedarsani, R., Chandrasekher, K., Sadigh, D., 2018b. Maximizing road capacity using cars that influence people. In: 2018 IEEE Conference on Decision and Control (CDC), pp. 1801–1808.

Lee, J.-B., Ozbay, K., 2009. New calibration methodology for microscopic traffic simulation using enhanced simultaneous perturbation stochastic approximation approach. Transp. Res. Rec. 2124 (1), 233–240.

Levin, M.W., Boyles, S.D., 2015. Intersection auctions and reservation-based control in dynamic traffic assignment. Transport. Res. Rec.: J. Transport. Res. Board 2497, 35–44.

Levin, M.W., Boyles, S.D., 2016. A multiclass cell transmission model for shared human and autonomous vehicle roads. Transport. Res. Part C: Emerg. Technol. 62, 103–116.

Levine, S., Koltun, V., 2012. Continuous inverse optimal control with locally optimal examples. arXiv preprint arXiv:1206.4617.

Li, K., Wang, X., Xu, Y., Wang, J., 2016. Lane changing intention recognition based on speech recognition models. Transport. Res. Part C: Emerg. Technol. 69, 497–514.

Li, L., Wen, D., Yao, D., 2014. A survey of traffic control with vehicular communications.

Li, M., Qin, Z., Jiao, Y., Yang, Y., Wang, J., Wang, C., Wu, G., Ye, J., 2019. Efficient Ridesharing Order Dispatching with Mean Field Multi-Agent Reinforcement Learning. In: The World Wide Web Conference. WWW '19. ACM, New York, NY, USA, pp. 983–994, event-place: San Francisco, CA, USA.

Li, N., Oyler, D.W., Zhang, M., Yildiz, Y., Kolmanovsky, I., Girard, A.R., 2018a. Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems. IEEE Trans. Control Syst. Technol. 26 (5), 1782–1797.

Li, S.E., Zheng, Y., Li, K., Wang, L.-Y., Zhang, H., 2017a. Platoon control of connected vehicles from a networked control perspective: Literature review, component modeling, and controller synthesis. IEEE Trans. Veh. Technol.

Li, S.E., Zheng, Y., Li, K., Wu, Y., Hedrick, J.K., Gao, F., Zhang, H., 2017b. Dynamical modeling and distributed control of connected and automated vehicles: Challenges and opportunities. IEEE Intell. Transp. Syst. Mag. 9 (3), 46–58.

Li, X., Ghiasi, A., Xu, Z., Qu, X., 2018b. A piecewise trajectory optimization model for connected automated vehicles: Exact optimization algorithm and queue propagation analysis. Transport. Res. Part B: Methodol. 118, 429–456.

Li, Y., Tang, C., Li, K., He, X., Peeta, S., Wang, Y., 2018c. Consensus-based cooperative control for multi-platoon under the connected vehicles environment. IEEE Trans. Intell. Transp. Syst. 20 (6), 2220–2229.

Li, Y., Tang, C., Peeta, S., Wang, Y., 2018d. Nonlinear consensus-based connected vehicle platoon control incorporating car-following interactions and heterogeneous time delays. IEEE Trans. Intell. Transp. Syst.

Liao, S., Zhou, L., Di, X., Yuan, B., Xiong, J., 2018. Large-scale short-term urban taxi demand forecasting using deep learning. In: Proceedings of the 23rd Asia and South Pacific Design Automation Conference. IEEE Press, pp. 428–433.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves ii. a theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. A 229 (1178), 317–345.

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., 2015. Continuous control with deep reinforcement learning. In: arXiv preprint arXiv:1509.02971.

Lin, K., Zhao, R., Xu, Z., Zhou, J., 2018. Efficient Large-Scale Fleet Management via Multi-Agent Deep Reinforcement Learning. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18. ACM, New York, NY, USA, pp. 1774–1783.

Lin, X., Li, M., Shen, Z.-J. M., Yin, Y., He, F., 2020. Rhythmic control of automated traffic–part ii: Grid network rhythm and online routing. arXiv preprint arXiv: 2010.05416.

Lin, Y.-C., Hong, Z.-W., Liao, Y.-H., Shih, M.-L., Liu, M.-Y., Sun, M., 2017. Tactics of adversarial attack on deep reinforcement learning agents. arXiv preprint arXiv: 1703.06748.

Littman, M.L., 1994. Markov Games As a Framework for Multi-agent Reinforcement Learning. In: Proceedings of the Eleventh International Conference on International Conference on Machine Learning. ICML'94. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 157–163, event-place: New Brunswick, NJ, USA.

Liu, C., Lin, C.-W., Shiraishi, S., Tomizuka, M., 2018. Improving efficiency of autonomous vehicles by v2v communication. In: 2018 Annual American Control Conference (ACC). IEEE, pp. 4778–4783.

Liu, C., Tomizuka, M., 2015. Safe exploration: Addressing various uncertainty levels in human robot interactions. In: ACC. pp. 465–470.

Liu, C., Tomizuka, M., 2016. Enabling safe freeway driving for automated vehicles. In: American Control Conference (ACC), 2016. IEEE, pp. 3461–3467.

Liu, H.X., Xin, W., Adam, Z., Ban, J., 2007. A game theoretical approach for modelling merging and yielding behaviour at freeway on-ramp sections. Transport. Traffic Theory 3, 197–211.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., Mordatch, I., 2017. Multi-agent Actor-critic for Mixed Cooperative-competitive Environments. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Curran Associates Inc., USA, pp. 6382–6393, event-place: Long Beach, California, USA.

Lu, X.-Y., Skabardonis, A., 2007. Freeway traffic shockwave analysis: exploring the ngsim trajectory data. In: 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Lyamin, N., Kleyko, D., Delooz, Q., Vinel, A., 2018. Ai-based malicious network traffic detection in vanets. IEEE Network 32 (6), 15–21.

Ma, J., Li, X., Shladover, S.E., Rakha, H.A., Lu, X.-Y., Jagannathan, R., Dailey, D.J., 2016. Freeway speed harmonization. IEEE Trans. Intelligent Vehicles 1 (1), 78–89.

Ma, T., Abdulhai, B., 2002. Genetic algorithm-based optimization approach and generic tool for calibrating traffic microscopic simulation parameters. Transport. Res. Rec.: J. Transport. Res. Board 1800, 6–15.
Malikopoulos, A.A., Hong, S., Park, B.B., Lee, J., Ryu, S., 2018. Optimal control for speed harmonization of automated vehicles. IEEE Trans. Intell. Transp. Syst. 99, 1–13.
Mao, T., Zhang, W., He, H., Lin, Y., Kale, V., Stein, A., Kostic, Z., 2018. Aic 2018 report: Traffic surveillance research. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 85–92.
Matignon, L., Laurent, G.J., Fort-Piat, N.L., 2012. Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems. Knowledge Eng. Rev. 27 (1), 1–31.
McLaughlin, S.B., Hankey, J.M., 2015. Naturalistic Driving Study: Linking the Study Data to the Roadway Information Database. No. SHRP 2 Report S2–S31-RW-3.
Meinrenken, C.J., Shou, Z., Di, X., 2020. Using gps-data to determine optimum electric vehicle ranges: A michigan case study. Transport. Res. Part D: Transport Environ. 78, 102203.
Melson, C.L., Levin, M.W., Hammit, B.E., Boyles, S.D., 2018. Dynamic traffic assignment of cooperative adaptive cruise control. Transport. Res. Part C: Emerg. Technol. 90, 114–133.
Milanés, V., Shladover, S.E., 2014. Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. Transport. Res. Part C: Emerg. Technol. 48, 285–300.
Milanés, V., Shladover, S.E., Spring, J., Nowakowski, C., Kawazoe, H., Nakamura, M., 2014. Cooperative adaptive cruise control in real traffic situations. IEEE Trans. Intell. Transp. Syst. 15 (1), 296–305.
Millard-Ball, A., 2016. Pedestrians, autonomous vehicles, and cities. J. Plann. Educ. Res. 0739456X16675674.
Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533.
Mo, Z., Di, X., Shi, R., 2021. A physics-informed deep learning paradigm for car-following models. (under review) https://arxiv.org/abs/2012.13376.
Mueller, A., 2019. Modern robotics: Mechanics, planning, and control [bookshelf]. IEEE Control Syst. Mag. 39 (6), 100–102.
Mukadam, M., Cosgun, A., Nakhaei, A., Fujimura, K., 2017. Tactical decision making for lane changing with deep reinforcement learning.
Müller, L., Risto, M., Emmenegger, C., 2016. The social behavior of autonomous vehicles. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, pp. 686–689.
Muller, U., Ben, J., Cosatto, E., Flepp, B., Cun, Y., 2006. Off-road obstacle avoidance through end-to-end learning. In: Advances in Neural Information Processing Systems (NIPS), pp. 739–746.
Naus, G.J., Vugts, R.P., Ploeg, J., van de Molengraft, M.J., Steinbuch, M., 2010. String-stable cacc design and experimental validation: A frequency-domain approach. IEEE Trans. Vehicular Technol. 59 (9), 4268–4279.
NDS, I.-S., 2018. insight.shrp2nds.us/home, [Online; accessed 12.31.2020].
Newell, G.F., 1961. Nonlinear effects in the dynamics of car following. Oper. Res. 9 (2), 209–229.
Ng, A.Y., Russell, S.J., 2000. Algorithms for inverse reinforcement learning. In: Icml. vol. 1. p. 2.
Ngoduy, D., 2013a. Analytical studies on the instabilities of heterogeneous intelligent traffic flow. Commun. Nonlinear Sci. Numer. Simul. 18 (10), 2699–2706.
Ngoduy, D., 2013b. Instability of cooperative adaptive cruise control traffic flow: A macroscopic approach. Commun. Nonlinear Sci. Numer. Simul. 18 (10), 2838–2851.
Ngoduy, D., Hoogendoorn, S., Liu, R., 2009. Continuum modeling of cooperative traffic flow dynamics. Phys. A 388 (13), 2705–2716.
Nguyen, T., Nguyen, N., Nahavandi, S., 2018. Deep reinforcement learning for multi-agent systems: a review of challenges. In: arXiv preprint arXiv:1812.11794.
Omidshafiei, S., Pazis, J., Amato, C., How, J., Vian, J., 2017. Deep decentralized multi-task multi-agent reinforcement learning under partial observability. In: The 34th International Conference on Machine Learning, pp. 2681–2690.
OpenAI, 2018. Openai five. https://blog.openai.com/openai-five/.
Orosz, G., Wilson, R.E., Stépán, G., 2010. Traffic jams: dynamics and control.
Ossen, S., Hoogendoorn, S.P., 2011. Heterogeneity in car-following behavior: Theory and empirics. Transport. Res. Part C: Emerg. Technol. 19 (2), 182–195.
Palmer, G., Tuyls, K., Bloembergen, D., Savani, R., 2018. Lenient multi-agent deep reinforcement learning. In: The 17th International Conference on Autonomous Agents and MultiAgent Systems, pp. 443–451.
Pan, Y., Cheng, C.-A., Saigol, K., Lee, K., Yan, X., Theodorou, E., Boots, B., 2018. Agile autonomous driving using end-to-end deep imitation learning. In: Robotics: Science and Systems.
Panwai, S., Dia, H., 2007. Neural agent car-following models. IEEE Trans. Intell. Transp. Syst. 8 (1), 60–70.
Patel, R., Levin, M.W., Boyles, S.D., 2016. Effects of autonomous vehicle behavior on arterial and freeway networks. Transport. Res. Rec.: J. Transport. Res. Board (2561).
Patil, A., Malla, S., Gang, H., Chen, Y.-T., 2019. The H3D dataset for full-surround 3D multi-object detection and tracking in crowded urban scenes. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 9552–9557.
Paxton, C., Raman, V., Hager, G.D., Kobilarov, M., 2017. Combining neural networks and tree search for task and motion planning in challenging environments. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 6059–6066.
Payne, H.J., 1971. Model of freeway traffic and control. Mathe. Model Public Syst. 51–61.
Pedersen, P.A., 2001. A game theoretical approach to road safety. Tech. rep., Department of Economics Discussion Paper, University of Kent.
Pedersen, P.A., 2003. Moral hazard in traffic games. J. Transport Econ. Policy (JTEP) 37 (1), 47–68.
Perot, E., Jaritz, M., Toromanoff, M., de Charette, R., 2017. End-to-end driving in a realistic racing game with deep reinforcement learning. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 474–475.
Pham, Q.-H., Sevestre, P., Pahwa, R.S., Zhan, H., Pang, C.H., Chen, Y., Mustafa, A., Chandrasekhar, V., Lin, J., 2019. A*3D Dataset: Towards autonomous driving in challenging environments. arXiv preprint arXiv:1909.07541.
Phegley, B., Gomes, G., Horowitz, R., 2014. Fundamental diagram calibration: A stochastic approach to linear fitting. In: Proceedings of the 93rd Annual Meeting of the Transportation Research Board, Washington, DC.
Ploeg, J., Scheepers, B.T., Van Nunen, E., Van de Wouw, N., Nijmeijer, H., 2011. Design and experimental evaluation of cooperative adaptive cruise control. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 260–265.
Pomerleau, D.A., 1989. ALVINN: An autonomous land vehicle in a neural network. In: Advances in Neural Information Processing Systems (NIPS), pp. 305–313.
Porfyri, K., Nikolos, I., Delis, A., Papageorgiou, M., 2015. Stability analysis of a macroscopic traffic flow model for adaptive cruise control systems. In: ASME 2015 International Mechanical Engineering Congress and Exposition. American Society of Mechanical Engineers, pp. V012T15A002–V012T15A002.
PTV, 2020. Virtual testing of autonomous vehicles with PTV Vissim. https://www.ptvgroup.com/en/solutions/products/ptv-vissim/areas-of-application/autonomous-vehicles-and-new-mobility/, [Online; accessed 12.31.2020].
Qayyum, A., Usama, M., Qadir, J., Al-Fuqaha, A., 2020. Securing connected & autonomous vehicles: Challenges posed by adversarial machine learning and the way forward. IEEE Commun. Surv. Tutorials 22 (2), 998–1026.
Qin, W.B., Orosz, G., 2013. Digital effects and delays in connected vehicles: linear stability and simulations. In: ASME 2013 Dynamic Systems and Control Conference. American Society of Mechanical Engineers pp. V002T30A001–V002T30A001.
Qin, W.B., Orosz, G., 2017. Scalable stability analysis on large connected vehicle systems subject to stochastic communication delays. Transport. Res. Part C: Emerg. Technol. 83, 39–60.
Qu, X., Wang, S., Zhang, J., 2015. On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models. Transport. Res. Part B: Methodol. 73, 91–102.

Rahman, M., Chowdhury, M., Khan, T., Bhavsar, P., 2015. Improving the efficacy of car-following models with a new stochastic parameter estimation and calibration method. IEEE Trans. Intell. Transp. Syst. 16 (5), 2687–2699.
Rahmati, Y., Talebpour, A., 2017. Towards a collaborative connected, automated driving environment: A game theory based decision framework for unprotected left turn maneuvers. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 1316–1321.
Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J.W., Christakis, N.A., Couzin, I.D., Jackson, M.O., et al., 2019. Machine behaviour. Nature 568 (7753), 477–486.
Raissi, M., 2018. Deep hidden physics models: Deep learning of nonlinear partial differential equations. J. Machine Learn. Res. 19 (1), 932–955.
Raissi, M., Karniadakis, G.E., 2018. Hidden physics models: Machine learning of nonlinear partial differential equations. J. Comput. Phys. 357, 125–141.
Raissi, M., Wang, Z., Triantafyllou, M.S., Karniadakis, G.E., 2019. Deep learning of vortex-induced vibrations. J. Fluid Mech. 861, 119–137.
Rajamani, R., Shladover, S.E., 2001. An experimental comparative study of autonomous and co-operative vehicle-follower control systems. Transport. Res. Part C: Emerg. Technol. 9 (1), 15–31.
Rajbahadur, G.K., Malton, A.J., Walenstein, A., Hassan, A.E., 2018. A survey of anomaly detection for connected vehicle cybersecurity and safety. In: 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 421–426.
Rakha, H., Crowther, B., 2002. Comparison of greenshields, pipes, and van aerde car-following and traffic stream models. Transp. Res. Rec. 1802 (1), 248–262.
Rakha, H., Crowther, B., 2003. Comparison and calibration of fresim and integration steady-state car-following behavior. Transport. Res. Part A: Policy Practice 37 (1), 1–27.
Rakha, H.A., Gao, Y., et al., 2010. Calibration of steady-state car-following models using macroscopic loop detector data.
Ramamoorthy, A., Yampolskiy, R., 2018. Beyond mad? the race for artificial general intelligence. ITU J. 1, 1–8.
Ras, G., van Gerven, M., Haselager, P., 2018. Explanation methods in deep learning: Users, values, concerns and challenges. Explainable Interpretable Models Comput. Vision Machine Learn. 19–36.
Rausch, V., Hansen, A., Solowjow, E., Liu, C., Kreuzer, E., Hedrick, J.K., 2017. Learning a deep neural net policy for end-to-end control of autonomous vehicles. In: American Control Conference (ACC), pp. 4914–4919.
Richards, P.I., 1956. Shock waves on the highway. Oper. Res. 4 (1), 42–51.
Richter, S.R., Hayder, Z., Koltun, V., 2017. Playing for benchmarks. In: IEEE International Conference on Computer Vision, pp. 2213–2222.
Richter, S.R., Vineet, V., Roth, S., Koltun, V., 2016. Playing for data: Ground truth from computer games. In: European Conference on Computer Vision, pp. 102–118.
Rios-Torres, J., Malikopoulos, A.A., 2016. A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps. IEEE Trans. Intell. Transp. Syst. 18 (5), 1066–1077.
Ross, S., Bagnell, D., 2010. Efficient reductions for imitation learning. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 661–668.
Ross, S., Gordon, G., Bagnell, D., 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 627–635.
Sadigh, D., Sastry, S., Seshia, S.A., Dragan, A.D., 2016a. High-speed highway scene prediction based on driver models learned from demonstrations. Robot.: Sci. Syst. 2.
Sadigh, D., Sastry, S., Seshia, S.A., Dragan, A.D., 2016b. Planning for autonomous cars that leverage effects on human actions. In: Robotics: Science and Systems.
Sadigh, D., Sastry, S.S., Seshia, S.A., 2019. Verifying robustness of human-aware autonomous cars. IFAC-PapersOnLine 51 (34), 131–138.
Sallab, A.E., Abdou, M., Perot, E., Yogamani, S., 2017. Deep reinforcement learning framework for autonomous driving. In: arXiv preprint arXiv:1704.02532.
SAS, 2015. The connected vehicle: Big data, big opportunities. https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/connected-vehicle-107832.pdf, [Online; accessed 12.31.2020].
Schakel, W.J., Van Arem, B., Netten, B.D., 2010. Effects of cooperative adaptive cruise control on traffic flow stability. In: 2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 759–764.
Schulman, J., Levine, S., Abbeel, P., Jordan, M., Moritz, P., 2015a. Trust region policy optimization. In: International Conference on Machine Learning (ICML), pp. 1889–1897.
Schulman, J., Moritz, P., Levine, S., Jordan, M., Abbeel, P., 2015b. High-dimensional continuous control using generalized advantage estimation. In: arXiv preprint arXiv:1506.02438.
Schwarting, W., Pierson, A., Alonso-Mora, J., Karaman, S., Rus, D., 2019. Social behavior for autonomous vehicles. Proc. Nat. Acad. Sci. 116 (50), 24972–24978.
Shi, R., Mo, Z., Di, X., 2021a. Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models. In: 35th Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence. AAAI. In press.
Shi, R., Mo, Z., Huang, K., Di, X., Du, Q., 2021b. Physics-informed deep learning for traffic state estimation. (under review) https://arxiv.org/abs/2101.06580.
Shladover, S.E., Nowakowski, C., Lu, X.-Y., Ferlis, R., 2015. Cooperative adaptive cruise control: Definitions and operating concepts. Transport. Res. Rec.: J. Transport. Res. Board 2489, 145–152.
Shou, Z.Y., Cao, Z.H., Di, X., 2020. Similarity analysis of spatial-temporal travel patterns for travel mode prediction using Twitter data. In: The 23rd IEEE International Conference on Intelligent Transportation Systems (ITSC). https://doi.org/10.1109/ITSC45102.2020.9294709.
Shou, Z., Di, X., 2018. Similarity analysis of frequent sequential activity pattern mining. Transport. Res. Part C: Emerg. Technol. 96, 122–143.
Shou, Z., Di, X., 2020. Reward design for driver repositioning using multi-agent reinforcement learning. Transp. Res. Part C 119 (102738).
Shou, Z., Di, X., 2021. Multi-agent reinforcement learning for dynamic routing games: A unified paradigm. (under review) https://arxiv.org/abs/2011.10915.
Shou, Z., Di, X., Ye, J., Zhu, H., Zhang, H., Hampshire, R., 2020b. Optimal passenger-seeking policies on e-hailing platforms using markov decision process and imitation learning. Transport. Res. Part C: Emerg. Technol. 111, 91–113.
Shou, Z., Wang, Z., Han, K., Liu, Y., Tiwari, P., Di, X., 2020c. Long-term prediction of lane change maneuver through a multilayer perceptron. In: 2020 IEEE Intelligent Vehicles Symposium (IEEE IV 2020).
Shoukry, Y., Martin, P., Tabuada, P., Srivastava, M., 2013. Non-invasive spoofing attacks for anti-lock braking systems. In: International Conference on Cryptographic Hardware and Embedded Systems. Springer, pp. 55–72.
Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D., 2016. Mastering the game of go with deep neural networks and tree search. Nature 529 (7587), 484.
Singer, J.P., Atkinson, J., Myers, M., Robinson, A.E., Krueger, J., et al., 2013. Travel time on arterials and rural highways: state-of-the-practice synthesis on rural data collection technology. Tech. rep., United States. Federal Highway Administration.
Sitawarin, C., Bhagoji, A.N., Mosenia, A., Mittal, P., Chiang, M., 2018a. Darts: Deceiving autonomous cars with toxic signs. arXiv preprint arXiv:1802.06430.
Sitawarin, C., Bhagoji, A.N., Mosenia, A., Mittal, P., Chiang, M., 2018b. Rogue signs: Deceiving traffic sign recognition with malicious ads and logos. arXiv preprint arXiv:1801.02780.
Sivaraman, S., Trivedi, M.M., 2013. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. IEEE Trans. Intell. Transport. Syst. 14 (4), 1773–1795.
Song, J., Ren, H., Sadigh, D., Ermon, S., 2018. Multi-agent generative adversarial imitation learning.
Song, X., Wang, P., Zhou, D., Zhu, R., Guan, C., Dai, Y., Su, H., Li, H., Yang, R., 2019. Apollocar3D: A large 3D car instance understanding benchmark for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5452–5462.
Stern, R.E., Cui, S., Delle Monache, M.L., Bhadani, R., Bunting, M., Churchill, M., Hamilton, N., Pohlmann, H., Wu, F., Piccoli, B., et al., 2018. Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. Transport. Res. Part C: Emerg. Technol. 89, 205–221.
Sun, Z., Huang, T., Zhang, P., 2020. Cooperative decision-making for mixed traffic: A ramp merging example. Transport. Res. Part C: Emerg. Technol. 120, 102764.
Sunberg, Z.N., Ho, C.J., Kochenderfer, M.J., 2017. The value of inferring the internal state of traffic participants for autonomous freeway driving. In: 2017 American Control Conference (ACC), pp. 3004–3301.

Sutton, R.S., Barto, A.G., 1998. Introduction to Reinforcement Learning, 1st ed. MIT Press, Cambridge, MA, USA.
Swaroop, D., Hedrick, J.K., 1996. String stability of interconnected systems. IEEE Trans. Automatic Control 41 (3), 349–357.
Swaroop, D., Hedrick, J.K., Chien, C., Ioannou, P., 1994. A comparision of spacing and headway control laws for automatically controlled vehicles1. Vehicle Syst. Dyn. 23 (1), 597–625.
Swaroop, D., Hedrick, J.K., Choi, S.B., 2001. Direct adaptive longitudinal control of vehicle platoons. IEEE Trans. Veh. Technol. 50 (1), 150–161.
Syed, U., Schapire, R.E., 2008. A game-theoretic approach to apprenticeship learning. In: Advances in Neural Information Processing Systems (NIPS), pp. 1449–1456.
Talebpour, A., Mahmassani, H.S., 2016. Influence of connected and autonomous vehicles on traffic flow stability and throughput. Transport. Res. Part C: Emerg. Technol. 71, 143–163.
Talebpour, A., Mahmassani, H.S., Hamdar, S.H., 2015. Modeling lane-changing behavior in a connected environment: A game theory approach. Transport. Res. Procedia 7, 420–440.
Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., Aru, J., Vicente, R., 2017. Multiagent cooperation and competition with deep reinforcement learning. PLOS ONE 12 (4), e0172395.
Tanaka, M., 2013. Development of various artificial neural network car-following models with converted data sets by a self-organization neural network. J. Eastern Asia Soc. Transport. Stud. 10, 1614–1630.
Tang, Y., Zhang, C., Gu, R., Li, P., Yang, B., 2017. Vehicle detection and recognition for intelligent traffic surveillance system. Multimedia Tools Appl. 76 (4), 5817–5832.
Tennenholtz, M., 2002. Game theory and artificial intelligence. In: Foundations and Applications of Multi-agent Systems. Springer, pp. 49–58.
Tesla, 2018. Tesla. forums.tesla.com/forum/forums/tesla%E2%80%99s-fully-self-driving-timeline, [Online; accessed 12.31.2020].
Tian, R., Li, N., Kolmanovsky, I., Yildiz, Y., Girard, A., 2019. Game-theoretic modeling of traffic in unsignalized intersection network for autonomous vehicle control verification and validation. arXiv preprint arXiv:1910.07141.
Tian, R., Li, S., Li, N., Kolmanovsky, I., Girard, A., Yildiz, Y., 2018. Adaptive game-theoretic decision making for autonomous vehicle control at roundabouts. In: 2018 IEEE Conference on Decision and Control (CDC). IEEE, pp. 321–326.
Tilbury, D., Yang, J., Pradhan, A., Robert, L., et al., 2020. Analysis and prediction of pedestrian crosswalk behavior during automated vehicle interactions. In: Proceedings of the International Conference on Robotics and Automation.
Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Phys. Rev. E 62 (2), 1805.
Ucar, S., Ergen, S.C., Ozkasap, O., 2017. Data-driven abnormal behavior detection for autonomous platoon. In: 2017 IEEE Vehicular Networking Conference (VNC). IEEE, pp. 69–72.
USDOT, 2019. Connected vehicle. https://its.dot.gov/research_areas/WhitePaper_connected_vehicle.htm, [Online; accessed 12.31.2020].
USDOT, 2020. Connected vehicle pilot deployment program. https://www.its.dot.gov/pilots/, [Online; accessed 12.31.2020].
Van Arem, B., Van Driel, C.J., Visser, R., 2006. The impact of cooperative adaptive cruise control on traffic-flow characteristics. IEEE Trans. Intell. Transp. Syst. 7 (4), 429–436.
van Hinsbergen, C.P.I., van Lint, H.W., Hoogendoorn, S.P., van Zuylen, H.J., 2009. Bayesian calibration of car-following models. IFAC Proceedings Volumes 42 (15), 91–97.
van Wageningen-Kessels, F., Van Lint, H., Vuik, K., Hoogendoorn, S., 2015. Genealogy of traffic flow models. EURO J. Transport. Logist. 4 (4), 445–473.
VanderWerf, J., Shladover, S., Kourjanskaia, N., Miller, M., Krishnan, H., 2001. Modeling effects of driver control assistance systems on traffic. Transport. Res. Rec.: J. Transport. Res. Board 1748, 167–174.
Venayagamoorthy, G., Doctor, S., 2004. Unmanned vehicle navigation using swarm intelligence. In: International Conference on Intelligent Sensing and Information Processing, 2004. Proceedings of. IEEE, pp. 249–253.
Vinitsky, E., Kreidieh, A., Le Flem, L., Kheterpal, N., Jang, K., Wu, C., Wu, F., Liaw, R., Liang, E., Bayen, A.M., 2018. Benchmarks for reinforcement learning in mixed-autonomy traffic. In: Conference on Robot Learning, pp. 399–409.
Vinkhuyzen, E., Cefkin, M., 2016. Developing socially acceptable autonomous vehicles. In: Ethnographic Praxis in Industry Conference Proceedings. Vol. 2016. Wiley Online Library, pp. 522–534.
Vinyals, O., Babuschkin, I., Czarnecki, W.M., Mathieu, M., Dudzik, A., Chung, J., Choi, D.H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J.P., Jaderberg, M., Vezhnevets, A.S., Leblond, R., Pohlen, T., Dalibard, V., Budden, D., Sulsky, Y., Molloy, J., Paine, T.L., Gulcehre, C., Wang, Z., Pfaff, T., Wu, Y., Ring, R., Yogatama, D., Wünsch, D., McKinney, K., Smith, O., Schaul, T., Lillicrap, T., Kavukcuoglu, K., Hassabis, D., Apps, C., Silver, D., 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature 575 (7782), 350–354.
Wang, M., 2018. Infrastructure assisted adaptive driving to stabilise heterogeneous vehicle strings. Transport. Res. Part C: Emerg. Technol. 91, 276–295.
Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2014a. Rolling horizon control framework for driver assistance systems. part i: Mathematical formulation and non-cooperative systems. Transport. Res. Part C: Emerg. Technol. 40, 271–289.
Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2014b. Rolling horizon control framework for driver assistance systems. part ii: Cooperative sensing and cooperative control. Transport. Res. Part C: Emerg. Technol. 40, 290–311.
Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2016. Cooperative car-following control: Distributed algorithm and impact on moving jam features. IEEE Trans. Intell. Transp. Syst. 17 (5), 1459–1471.
Wang, M., Hoogendoorn, S.P., Daamen, W., van Arem, B., Happee, R., 2015. Game theoretic approach for predictive lane-changing and car-following control. Transport. Res. Part C: Emerg. Technol. 58, 73–92.
Wang, X., Jiang, R., Li, L., Lin, Y., Zheng, X., Wang, F.-Y., 2018. Capturing car-following behaviors by deep learning. IEEE Trans. Intell. Transp. Syst. 19 (3), 910–920.
Wang, Y., Cai, P., Lu, G., 2020a. Cooperative autonomous traffic organization method for connected automated vehicles in multi-intersection road networks. Transport. Res. Part C: Emerg. Technol. 111, 458–476.
Wang, Y., Sarkar, E., Maniatakos, M., Jabari, S.E., 2020b. Stop-and-go: Exploring backdoor attacks on deep reinforcement learning-based traffic congestion control systems. arXiv preprint arXiv:2003.07859.
Waymo, 2019. Waymo Open Dataset: An autonomous driving dataset. https://waymo.com/open/data/.
Waymo, 2020. Waymo safety report: On the road to fully self-driving.
Wei, Y., Avcı, C., Liu, J., Belezamo, B., Aydın, N., Li, P.T., Zhou, X., 2017. Dynamic programming-based multi-vehicle longitudinal trajectory optimization with simplified car following models. Transport. Res. Part B: Methodol. 106, 102–129.
Wei, Z., Wang, C., Hao, P., Barth, M.J., 2019. Vision-based lane-changing behavior detection using deep residual neural network. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, pp. 3108–3113.
Wierstra, D., Forster, A., Peters, J., Schmidhuber, J., 2010. Recurrent policy gradients. Logic J. IGPL 18 (5), 620–634.
Wiest, J., Höffken, M., Kreßel, U., Dietmayer, K., 2012. Probabilistic trajectory prediction with gaussian mixture models. In: 2012 IEEE Intelligent Vehicles Symposium. IEEE, pp. 141–146.
Wolf, I., 2016. The interaction between humans and autonomous agents. In: Autonomous Driving. Springer, pp. 103–124.
Woo, H., Ji, Y., Kono, H., Tamura, Y., Kuroda, Y., Sugano, T., Yamamoto, Y., Yamashita, A., Asama, H., 2017. Lane-change detection based on vehicle-trajectory prediction. IEEE Robot. Automat. Lett. 2 (2), 1109–1116.
Wu, C., 2018. Learning and optimization for mixed autonomy systems-a mobility context. Ph.D. thesis. UC Berkeley.
Wu, C., Bayen, A.M., Mehta, A., 2018. Stabilizing traffic with autonomous vehicles. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 1–7.
Wu, C., Kreidieh, A., Parvate, K., Vinitsky, E., Bayen, A.M., 2017a. Flow: Architecture and benchmarking for reinforcement learning in traffic control. arXiv preprint arXiv:1710.05465.
Wu, C., Kreidieh, A., Vinitsky, E., Bayen, A.M., 2017b. Emergent behaviors in mixed-autonomy traffic. In: Conference on Robot Learning, pp. 398–407.

Wu, C., Kreidieh, A.R., Parvate, K., Vinitsky, E., Bayen, A.M., 2017c. Flow: A modular learning framework for autonomy in traffic. In: arXiv preprint arXiv: 1710.05465.
Wu, C., Parvate, K., Kheterpal, N., Dickstein, L., Mehta, A., Vinitsky, E., Bayen, A.M., 2017d. Framework for control and deep reinforcement learning in traffic. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1–8.
Wymann, B., Espie, E., Guionneau, C., Dimitrakakis, C., Coulom, R., Sumner, A., 2000. TORCS, the open racing car simulator. Software available at http://torcs.sourceforge.net.
Xie, N., Ras, G., van Gerven, M., Doran, D., 2020. Explainable deep learning: A field guide for the uninitiated. arXiv preprint arXiv:2004.14545.
Xu, H., Gao, Y., Yu, F., Darrell, T., 2017. End-to-end learning of driving models from large-scale video datasets. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2174–2182.
Yang, S., Wang, W., Jiang, Y., Wu, J., Zhang, S., Deng, W., 2019. What contributes to driving behavior prediction at unsignalized intersections? Transport. Res. Part C: Emerg. Technol. 108, 100–114.
Yang, S., Wang, W., Liu, C., Deng, W., Hedrick, J.K., 2017. Feature analysis and selection for training an end-to-end autonomous vehicle controller using deep learning approach. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 1033–1038.
Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., Wang, J., 2018. Lenient multi-agent deep reinforcement learning. In: The International Conference on Machine Learning, pp. 5567–5576.
Yang, Y., Perdikaris, P., 2019. Adversarial uncertainty quantification in physics-informed neural networks. J. Comput. Phys. 394, 136–152.
Yao, H., Cui, J., Li, X., Wang, Y., An, S., 2018. A trajectory smoothing method at signalized intersection based on individualized variable speed limits with location optimization. Transport. Res. Part D: Transport Environ. 62, 456–473.
Yao, H., Li, X., 2020. Decentralized control of connected automated vehicle trajectories in mixed traffic at an isolated signalized intersection. Transport. Res. Part C: Emerg. Technol. 121, 102846.
Yao, Z., Hu, R., Wang, Y., Jiang, Y., Ran, B., Chen, Y., 2019. Stability analysis and the fundamental diagram for mixed connected automated and human-driven vehicles. Phys. A 533, 121931.
Ye, Y., Zhang, X., Sun, J., 2019. Automated vehicle's behavior decision making using deep reinforcement learning and high-fidelity simulation environment. Transport. Res. Part C: Emerg. Technol. 107, 155–170.
Yoo, J., Langari, R., 2020. A game-theoretic model of human driving and application to discretionary lane-changes. arXiv preprint arXiv:2003.09783.
Yoo, J.H., Langari, R., 2012. Stackelberg game based model of highway driving. In: ASME 2012 5th Annual Dynamic Systems and Control Conference joint with the JSME 2012 11th Motion and Vibration Conference. American Society of Mechanical Engineers, pp. 499–508.
Yoo, J.H., Langari, R., 2013. A stackelberg game theoretic driver model for merging. In: ASME 2013 Dynamic Systems and Control Conference. American Society of Mechanical Engineers pp. V002T30A003–V002T30A003.
Yu, H., Tseng, H.E., Langari, R., 2018. A human-like game theory-based controller for automatic lane changing. Transport. Res. Part C: Emerg. Technol. 88, 140–158.
Zegers, J.C., Semsar-Kazerooni, E., Ploeg, J., van de Wouw, N., Nijmeijer, H., 2017. Consensus control for vehicular platooning with velocity constraints. IEEE Trans. Control Syst. Technol. 26 (5), 1592–1605.
Zhang, J., Vinkhuyzen, E., Cefkin, M., 2017. Evaluation of an autonomous vehicle external communication system concept: a survey study. In: International Conference on Applied Human Factors and Ergonomics. Springer, pp. 650–661.
Zhang, K., Yang, Z., Başar, T., 2019a. Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms. arXiv preprint arXiv:1911.10635.
Zhang, Q., Filev, D., Tseng, H.E., Szwabowski, S., Langari, R., 2019b. A game theoretic four-stage model predictive controller for highway driving. In: 2019 American Control Conference (ACC). IEEE, pp. 1375–1381.
Zhang, Q., Langari, R., Tseng, H.E., Filev, D.P., Szwabowski, S., Coskun, S., 2019c. A game theoretic model predictive controller with aggressiveness estimation for mandatory lane change. IEEE Trans. Intell. Vehicles.
Zhang, T., Kahn, G., Levine, S., Abbeel, P., 2016. Learning deep control policies for autonomous aerial vehicles with MPC-guided policy search. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 528–535.
Zhang, Y., Bai, Y., Wang, M., Hu, J., 2020. Cooperative adaptive cruise control with robustness against communication delay: An approach in the space domain. IEEE Trans. Intell. Transp. Syst.
Zhang, Y., Foroosh, H., David, P., Gong, B., 2018a. CAMOU: Learning physical vehicle camouflages to adversarially attack detectors in the wild. In: International Conference on Learning Representations.
Zhang, Z.Y., Zheng, J., Wang, X., Fan, X., 2018b. Background filtering and vehicle detection with roadside lidar based on point association. In: 2018 37th Chinese Control Conference (CCC). IEEE, pp. 7938–7943.
Zhao, L., Malikopoulos, A., Rios-Torres, J., 2018. Optimal control of connected and automated vehicles at roundabouts: An investigation in a mixed-traffic environment. IFAC-PapersOnLine 51 (9), 73–78.
Zhao, X., Wang, Z., Xu, Z., Wang, Y., Li, X., Qu, X., 2020. Field experiments on longitudinal characteristics of human driver behavior following an autonomous vehicle. Transport. Res. Part C: Emerg. Technol. 114, 205–224.
Zheng, Y., Li, S.E., Li, K., Ren, W., 2017. Platooning of connected vehicles with undirected topologies: Robustness analysis and distributed h-infinity controller synthesis. IEEE Trans. Intell. Transp. Syst. 19 (5), 1353–1364.
Zheng, Y., Li, S.E., Li, K., Wang, L.-Y., 2016. Stability margin improvement of vehicular platoon considering undirected topology and asymmetric control. IEEE Trans. Control Syst. Technol. 24 (4), 1253–1265.
Zhou, H., Laval, J., 2019. Longitudinal motion planning for autonomous vehicles and its impact on congestion: A survey. arXiv preprint arXiv:1910.06070.
Zhou, M., Qu, X., Jin, S., 2016. On the impact of cooperative autonomous vehicles in improving freeway merging: a modified intelligent driver model-based approach. IEEE Trans. Intell. Transp. Syst. 18 (6), 1422–1428.
Zhou, M., Qu, X., Li, X., 2017a. A recurrent neural network based microscopic car following model to predict traffic oscillation. Transport. Res. Part C: Emerg. Technol. 84, 245–264.
Zhou, Y., Ahn, S., 2019. Robust local and string stability for a decentralized car following control strategy for connected automated vehicles. Transport. Res. Part B: Methodol. 125, 175–196.
Zhou, Y., Ahn, S., Chitturi, M., Noyce, D.A., 2017b. Rolling horizon stochastic optimal control strategy for acc and cacc under uncertainty. Transport. Res. Part C: Emerg. Technol. 83, 61–76.
Zhou, Y., Ahn, S., Wang, M., Hoogendoorn, S., 2020. Stabilizing mixed vehicular platoons with connected automated vehicles: An h-infinity approach. Transport. Res. Part B: Methodol. 132, 152–170.
Zhou, Y., Wang, M., Ahn, S., 2019. Distributed model predictive control approach for cooperative car-following with guaranteed local and string stability. Transport. Res. Part B: Methodol. 128, 69–86.
Zhu, M., Wang, X., Tarko, A., et al., 2018a. Modeling car-following behavior on urban expressways in shanghai: A naturalistic driving study. Transport. Res. Part C: Emerg. Technol. 93, 425–445.
Zhu, M., Wang, X., Wang, Y., 2018b. Human-like autonomous car-following planning by deep reinforcement learning. Tech. Rep.
Ziebart, B.D., Maas, A.L., Bagnell, J.A., Dey, A.K., 2008. Maximum entropy inverse reinforcement learning. In: AAAI. Vol. 8. Chicago, IL, USA, pp. 1433–1438.