



---

## A Review on Deep Reinforcement Learning for Autonomous Driving

Zheen Kamil Ali<sup>1</sup>, Adnan Mohsin Abdulazeez<sup>2</sup>

zheen.ali@dpu.edu.krd, adnan.mohsin@dpu.edu.krd

<sup>1</sup>ITM Dept., Duhok Technical College, Duhok Polytechnic University, Iraq

<sup>2</sup>IT Dept., Duhok Technical College, Duhok Polytechnic University, Iraq

---

### Article Information

Submitted : 16 May 2024

Reviewed: 22 May 2024

Accepted : 15 Jun 2024

---

### Keywords

Reinforcement learning,  
Autonomous Driving,  
Motion planning, Deep  
Reinforcement learning  
algorithms, Driving  
Technology

---

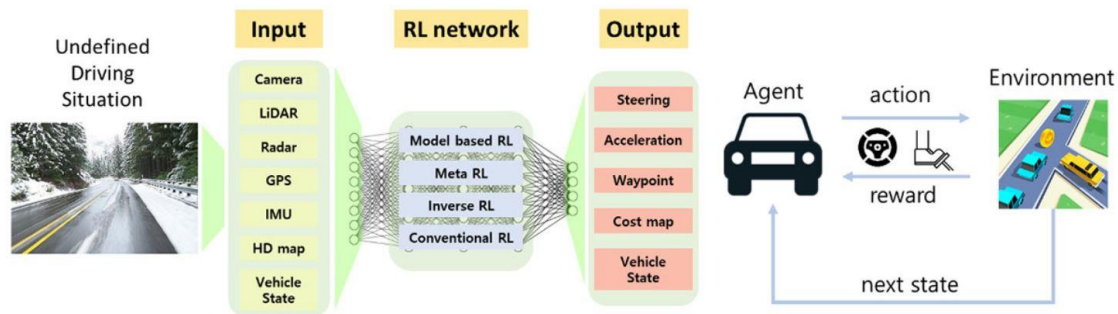
### Abstract

Autonomous driving technology has gained significant attention, offering opportunities to modernize transportation systems worldwide. Deep reinforcement learning (DRL) has emerged as a robust approach to design smart driving policies for intricate and changeable environments. This paper provides a detailed investigation of state-of-the-art DRL methodologies that are effectively applied to autonomous driving. It begins by providing a clear explanation of the fundamental concepts of deep learning and reinforced learning, highlighting their application for control of self-driving vehicles. Consequently, the paper presents an overview of various DRL algorithms, including Deep Q-Networks (DQN), Deep Deterministic Policy Gradients (DDPG), and Actor-Critic methods, describing their structures, training approaches, and applications in autonomous driving situations. Recent advancements in DRL research, such as domain adaptation, imitation learning, and meta-learning, have also been addressed in the study, with an investigation of their potential implications for autonomous driving. Via a thorough assessment of current literature, key trends, challenges, and research directions have been identified for exploiting DRL in autonomous car development. This review intends to provide a comprehensive understanding of the current and future possibilities of DRL for self-driving vehicles to researchers, practitioners, and enthusiasts.

## A. Introduction

Autonomous driving technology is a significant milestone in the progress of transportation systems across the globe. As it guarantees better safety, productivity, and availability, it has excited the interest of researchers, industry personnel, policymakers, and the general public. The core of this revolution in technology relies on the state-of-the-art method of DRL, which enables vehicles to steer and make choices in intricate and ever-changing settings [1][2][3][4].

The fusion of deep neural networks [5] with reinforcement learning concepts has advanced autonomous driving technology to unparalleled heights of refinement. By contrast to conventional rule-based systems that depend on manually constructed algorithms and heuristic principles, DRL empowers the vehicles with the ability to develop optimal driving techniques via trial and error, just like human drivers hone their capacities via practice [6][7][8]. Through leveraging data-supported decision-making [9][10][11], DRL has the potential of opening up novel realms in autonomy, permitting the vehicles to respond to varying traffic situations, unanticipated occurrences, and adjustments in road conditions [12][13].



**Figure 1.** The structure of autonomous driving based on reinforcement learning from end to end.

The purpose of this review paper is to present a thorough analysis of current cutting-edge DRL techniques that are used in autonomous driving. By delving deeply into the concepts, recent developments, advanced algorithms, and upcoming research trends, we aim to provide readers with a comprehensive understanding of the benefits and challenges involved in utilizing DRL for the progress of autonomous vehicle technology.

Our aim is to allow researchers, practitioners, and enthusiasts to gain a deeper understanding of the significance of DRL in shaping the future of autonomous driving. We hope to encourage them to explore new ways to further the progress of autonomous vehicle technology by providing a synthesis of insights from existing literature along with critical analysis. Ultimately, we seek to pave the way for safer, more efficient, and sustainable transportation systems through this review paper.

The subsequent sections of this paper are organized as follows: Section 2 explains the fundamentals of Reinforcement Learning. Section 3 provides an overview of DRL Algorithms. Section 4 covers Applications of Reinforcement

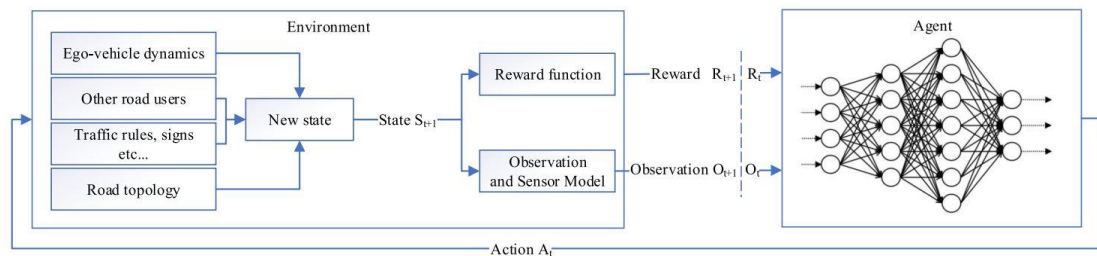
Learning in Autonomous Driving, while Section 5 delves into Recent Advancements in DRL.

## B. Fundamentals of Reinforcement Learning

Reinforcement Learning is a type of machine learning that focuses on how an agent can learn to achieve a goal by interacting with its environment [14]. Through trial and error, the agent takes actions in the environment and receives either rewards or penalties as feedback. The purpose of the agent is to learn a policy that will enable it to acquire the highest amount of cumulative reward over time.

At the core of RL are several fundamental concepts:

- **Agent:** The entity in charge of decision-making and taking actions within the environment.
- **Environment:** The external system with which the agent interacts
- **State:** A representation of the current situation or configuration of the environment.
- **Action:** The choices made by the agent that influence the state of the environment.
- **Reward:** Numeric feedback provided by the environment to indicate the desirability of a particular action taken by the agent.
- 



**Figure 2.** The (POMDP) model applied to autonomous driving using deep reinforcement learning [15]

Markov Decision Processes (MDPs) offer a numerical methodology to emulate RL problems. An MDP includes states, actions, transition probabilities, and rewards. An MDP [16] can be described by a tuple  $(S, A, P, R, \gamma)$ , where  $S$  denotes the set of states,  $A$  represents the set of actions, and  $P$  is the function governing state transitions, determining the probability of transitioning to the next state,  $s' \in S$ , when action  $a \in A$  is chosen in response to observing state  $s \in S$ . In cases involving continuous actions or state spaces, the mathematical formulation becomes more intricate. Additionally,  $R$  denotes the reward function mapping from state-action-state tuples to real numbers ( $R: S \times A \times S \rightarrow \mathbb{R}$ ), and  $\gamma \in [0, 1]$  represents the discount rate.

The objective of the agent is to acquire a policy,  $\pi: S \times A \rightarrow [0, 1]$ , that relates states to actions in such a manner that the expected cumulative reward is maximized.

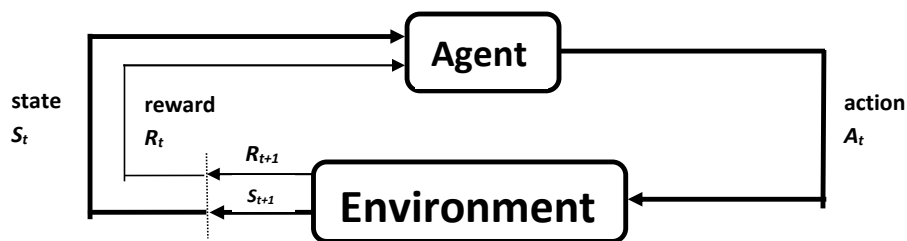
1. For the state-value function  $V^*(s)$ :
 
$$V^*(s) = \max_{\pi} E[\sum_k \gamma^k r_{t+k} | st = s] \quad (1)$$

2. For the action-value function  $Q^*(s, a)$ :
 
$$Q^*(s, a) = \max_{\pi} E[\sum_k \gamma^k r_{t+k} | st = s, at = a] \quad (2)$$

Here:

- $V^*(s)$  represents the optimal value function for state  $s$ ,
- $Q^*(s, a)$  represents the optimal action-value function for state-action pair  $(s, a)$ ,
- $\pi_\theta$  represents the policy parameterized by  $\theta$ ,
- $E[\cdot]$  represents the expected value operator,
- $\gamma$  is the discount factor,
- $r_{t+k}$  represents the reward at time  $t + k$ ,
- $s_t$  represents the state at time  $t$ ,
- $a_t$  represents the action at time  $t$ ,
- $S$  is the state space, and
- $A$  is the action space.

In this context, the state  $s$  corresponds to the input from car sensors, such as cameras and distance sensors. The steering angle is denoted by a continuous parameter  $a \in \mathbb{R}$ , which represents an action. The transition probability  $P$  is the probability of the car's state and action interacting. Based on the car's current state  $s$  and the action  $a$ , the reward function  $R$  provides feedback to the algorithm. The discount rate  $\gamma$  measures the significance of long-term rewards relative to immediate rewards, ranging from considering only immediate rewards ( $\gamma = 0$ ) to considering both immediate and long-term rewards equally ( $\gamma = 1$ ). The interaction between the agent and the environment is illustrated in this diagram [16].



let's consider a simplified example with two states: "Stop" and "Go." The autonomous vehicle can take two actions: "Brake" or "Accelerate." The transition probabilities and immediate rewards associated with each action are as follows:

1. Stop State:

- If the vehicle brakes, it remains in the stop state with a high probability (e.g., 0.9) and may transition to the go state with a lower probability (e.g., 0.1).
- Braking in the stop state incurs a negative reward due to the delay in reaching the destination (e.g., -10).

2. Go State:

- If the vehicle accelerates, it remains in the go state with a high probability (e.g., 0.8) and may transition back to the stop state with a lower probability (e.g., 0.2).
- Accelerating in the go state incurs a positive reward as it progresses towards the destination (e.g., +5).

Two frequently employed methods for resolving RL problems are Dynamic Programming and Q-Learning. In Dynamic Programming, algorithms like policy iteration and value iteration compute optimal policies by frequently revising value functions. In contrast, Q-Learning is an RL algorithm that is free of any underlying model and learns the value of state-action pairs from practice. It uses the Bellman equation to modify Q-values.

Having an understanding of these elementary concepts is crucial in comprehending the fundamental principles related to the implementation of RL techniques in self-driving systems.

### C. Deep Reinforcement Learning Algorithms

DRL methods [17] are essential for autonomous vehicles as they allow them to learn complicated driving behaviors by actively engaging with their surroundings. This portion of the paper presents a summary of the primary DRL algorithms implemented in autonomous driving R&D, complete with mathematical illustrations to explain their principles.

#### 1. Deep Q-Networks (DQN)

DQN is a method introduced in 2015 by Mnih et. al. [18][19] that utilizes deep neural networks to approximate the Q-function. The Q-function determines the expected cumulative reward for a given action in a particular state and is represented mathematically, the Q-function is represented as:

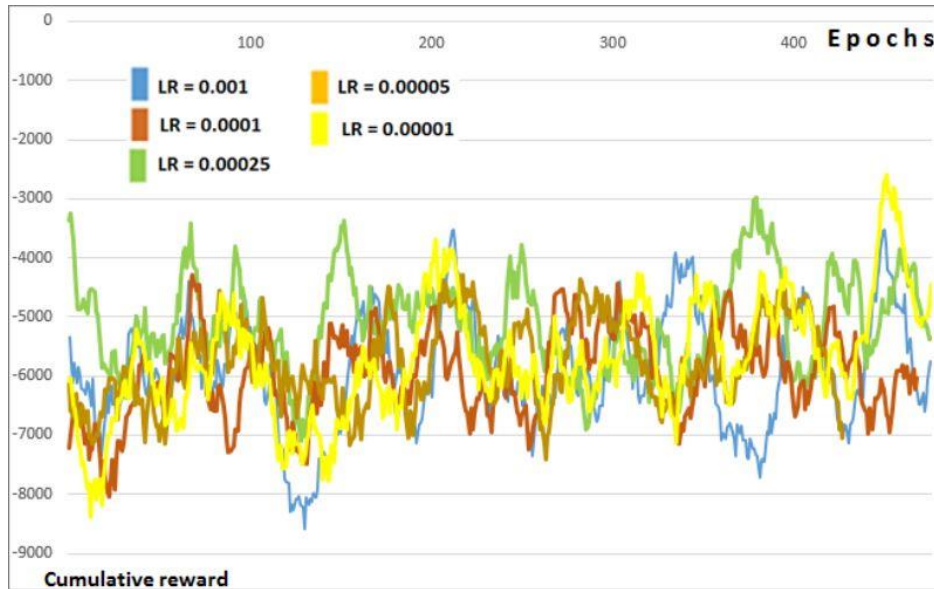
$$Q(s, a) = E[r + \gamma \max_{a'} Q(s', a')] \quad (3)$$

where  $s$  is the state,  $a$  is the action,  $r$  is the immediate reward,  $s'$  is the next state, and  $\gamma$  is the discount factor.

The DQN algorithm aims to minimize the temporal difference error between the predicted and observed rewards:

$$L(\theta) = E[(r + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta))^2] \quad (4)$$

where  $\theta$  represents the parameters of the neural network, and  $\theta$ -denotes the target network parameters.



**Figure3.** Evolution of the average of DQN cumulative reward with different learning rates [20]

### 2. Deep Deterministic Policy Gradients (DDPG)

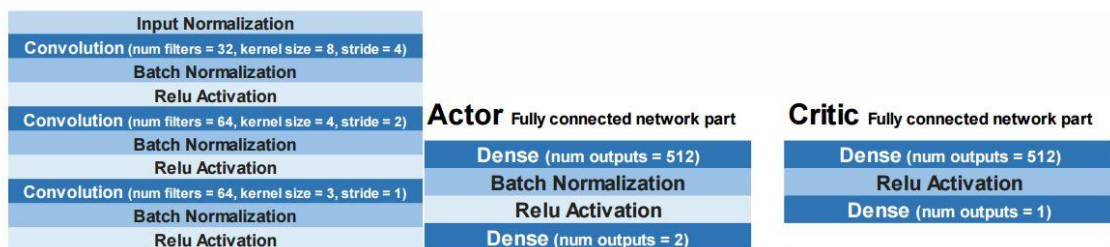
DDPG is a technique created by Lillicrap et al. in 2016 for continuous action spaces [21]. DDPG makes use of an actor-critic design where the actor network is trained to learn a policy that maps states to specific actions, and the critic network evaluates these actions' quality according to approximated value functions.

Mathematically, the actor network is trained to maximize the expected cumulative reward:

$$\nabla_{\theta} J \approx E[\nabla_{\theta} \mu(s|\theta) \nabla_a Q(s, a|\theta) | s=s_t, a=\mu(s_t)] \tag{5}$$

where  $\mu(s|\theta)$  represents the policy function,  $Q(s, a|\theta)$  is the action-value function, and  $J$  denotes the expected return [22].

For DDPG, the architecture of the actor and critic neural networks are the same for the visual features extraction, which uses three blocks each of which contain a convolutional layer, a batch normalization layer, and a Relu activation layer [20], see the Figure4 below.



**Figure 4.** The DDPG Network architecture

### 3. Actor-Critic Methods

Actor-Critic methods merge features from value-based and policy-based techniques by using different networks for the actor (policy) and critic (value

function). These algorithms are aimed at maximizing expected cumulative rewards while also estimating the value function to steer policy improvement.

Mathematically, the policy gradient for actor-critic methods can be expressed as:

$$\nabla_{\theta} J \approx E[\nabla_{\theta} \log \mu(a|s; \theta) Q(s, a; \theta)] \quad (6)$$

where  $\mu(a|s; \theta)$  represents the policy function,  $Q(s, a; \theta)$  is the action-value function, and  $J$  denotes the expected return [23].

#### 4. Other DRL Algorithms

In addition to DQN, DDPG, and Actor-Critic methods, various other DRL algorithms have been proposed and applied to autonomous driving tasks. These include Deep Q-Learning from Demonstrations (DQfD) [24], Twin Delayed Deep Deterministic Policy Gradients (TD3) [25], and Soft Actor-Critic (SAC) [26], among others.

#### D. Applications of Reinforcement Learning in Autonomous Driving

RL techniques have diverse applications in self-driving systems, including lane keeping [27], traffic light [28] and intersection management [29], collision avoidance [30], and path planning [31]. RL algorithms help vehicles navigate through complex environments, predict potential hazards, ensure safe passage, and optimize trajectories to minimize congestion. This versatility of RL algorithms makes them a crucial tool for enhancing the autonomy and safety of driving systems.

#### E. Recent Advancements in DRL

Benchmark datasets [16] and standardized evaluation metrics [17] are essential for enhancing research and development in DRL for self-driving vehicles. Benchmark datasets, like Waymo Open Dataset, Udacity Self-Driving Car, and CARLA, provide standard test and training environments and scenarios for benchmarking RL algorithms. Furthermore, established evaluation metrics, including safety measures, success rates, and computational efficiency, ensure that comparisons between different techniques and approaches are justifiable.

In summary, the recent developments in DRL techniques for self-driving cars have resulted in significant advancements, such as enhancing the capabilities of autonomous driving systems and addressing critical challenges. Researchers have leveraged various techniques, including imitation learning, domain adaptation, multi-agent learning, meta-learning, and benchmark datasets, to unlock new opportunities and overcome remaining hurdles towards mass deployment of safe and efficient autonomous vehicles.

#### F. Related Works

Sallab et al. (2017) [32], presented a framework for autonomous driving utilizing deep reinforcement learning. They address the challenges inherent in developing such agents and delineate three primary task categories: recognition, prediction, and decision-making. Their framework integrates Recurrent Neural Networks and attention models to address partially observable scenarios and emphasize pertinent information.

Liang et al. (2018) [33], addressed the challenge of devising optimal driving policies for autonomous urban driving. They propose a CIRL approach that

leverages encoded experiences mimicking human demonstrations to enhance exploration efficiency and tailor adaptive policies and steering-angle reward designs. The method has shown significant performance improvements compared to previous approaches on the CARLA driving benchmark.

Wang et. al. (2018) [34], explored the application of deep reinforcement learning in developing real-world autonomous driving systems using the deep deterministic policy gradient algorithm. They employed TORCS as the environment, incorporating a range of sensors and reward mechanisms, and devised a network architecture for both the actor and critic components. The model underwent evaluation across various modes in TORCS, demonstrating positive results both quantitatively and qualitatively.

Nagesh Rao et. al. (2019) [35], proposed a method for autonomous highway driving employing deep reinforcement learning and a modified version of the DDQN algorithm to train the decision-making neural network. The approach aims to mitigate unforeseen behaviors and adapt to diverse driving scenarios.

Pusse et. al. (2019) [36], introduced HyLEAP, a hybrid approach that combines deep reinforcement learning and approximate POMDP planning to address pedestrian collision-free navigation in self-driving cars. The performance evaluation is based on GIDAS, and the paper discusses the pros and cons of each method.

Folkers et. al. (2019) [37], presented a control approach based on deep reinforcement learning for autonomous vehicles. They discuss the training of a neural network agent with proximal policy optimization in a simulated environment and its application to a full-sized research vehicle for autonomous exploration of a parking lot, turning maneuvers, and obstacle avoidance. The paper also compares various model-based control approaches and references previous research that utilized deep learning methods for steering a vehicle solely based on camera images.

Moghadam et. al. (2019) [38], proposed an approach to autonomous driving that utilizes a hierarchical architecture with deep reinforcement learning. Their method aims to generate high-level sequential commands for lower-level controllers, ensuring consistent performance in uncertain environments.

Spielberg et. al. (2019) [39], introduced a novel adaptive, model-free controller for general discrete-time processes, utilizing deep reinforcement learning. Their proposed controller learns the control policy in real time through interactions with the process. The effectiveness and advantages of the controller are demonstrated through simulations across various scenarios.

You et. al. (2019) [40], discussed the planning problem of autonomous vehicles in traffic. The authors propose a stochastic Markov decision process model that incorporates road geometry to accommodate diverse driving styles. They also design the reward function.

Semnani et. al. (2020) [41], proposed a hybrid algorithm combining deep reinforcement learning (RL) and force-based motion planning (FMP) for distributed motion planning in dense and dynamic environments, where each agent has a fixed final position that cannot be exchanged with another agent. The proposed algorithm outperforms both deep RL and FMP algorithms, resulting in up



to 50% more successful scenarios than deep RL and up to 75% less extra time to reach the goal than FMP.

Liu et. al. (2020) [42], proposed a method for learning personalized discretionary lane-change initiation for fully autonomous driving based on reinforcement learning. The proposed offline algorithm employs a reinforcement learning technique to learn how to initiate lane changes from traffic context, the action of a self-driving vehicle, and in-vehicle user feedback. A multi-dimensional driving scenario is considered to represent a more realistic lane-change trade-off. The results show that the lane-change initiation model obtained by this method can reproduce the personal lane-change tactic, and the performance of the customized models is much better than that of the non-customized models.

Muzahid et. al. (2020) [43], discussed a conceptual framework based on reinforcement learning for threat assessment of multiple vehicle collisions in autonomous driving. They emphasize the importance of real-time crash risk prediction in ensuring a secure and effective autonomous driving system. The paper advocates for cross-disciplinary efforts that integrate various technological fields, such as robotics, artificial intelligence, machine learning, IoT, and reinforcement learning.

Rong et. al. (2020) [44], conducted a study on safe reinforcement learning with policy-guided planning for autonomous driving. The research addresses uncertainty and complexity within autonomous driving and demonstrates the effectiveness of this approach through numerical experiments. The authors introduce a hierarchical structure to work with formal specifications to ensure safety standards are met. The study concludes that the implemented approach is effective in achieving safe reinforcement learning.

Liao et. al. (2020) [45], investigated decision-making strategies on highways for autonomous vehicles using deep reinforcement learning (DRL). The study explores the dueling deep Q-network (DDQN) method to derive the highway decision-making strategy and conducts a series of simulation experiments to evaluate its effectiveness.

Kim et. al. (2020) [46], examined the application of deep reinforcement learning in developing intelligent self-driving policies aimed at minimizing injury and collision incidents in unexpected traffic scenarios. The study demonstrates that the trained agents surpassed human drivers and an autonomous emergency braking system in terms of collision avoidance and reducing injury severity.

Hishmeh et. al. (2020) [47], analyzed the performance of various deep reinforcement learning algorithms in the context of autonomous driving within a simulated environment. Their objective is to develop a short-term planner to complement other components for long-term planning and ensuring safety.

Lu et. al. (2020) [48], discussed a proposed approach to address the challenges of autonomous decision-making and motion planning for intelligent vehicles in complex traffic scenarios. The approach comprises two layers: a decision-making layer employing a kernel-based least-squares policy iteration algorithm with uneven sampling and pooling strategy, and a lower layer concentrating on lateral motion planning using a dual heuristic programming algorithm. The effectiveness and efficiency of this approach are demonstrated through extensive simulations.

Josef et. al. (2020) [49], introduced a deep reinforcement learning approach for local planning in unknown rough terrains for unmanned ground vehicles. The approach demonstrated superior performance compared to potential fields or local motion planning search space methods.

Palanisamy et. al. (2020) [50], proposed the utilization of Partially Observable Markov Games (POSG) to devise learning-based solutions for connected autonomous driving under realistic assumptions. They present a taxonomy of multi-agent learning environments and offer an extensible set of CAD simulation environments to develop algorithms for CAD systems in multi-agent settings.

Duan et. al. (2020) [51], explored the decomposition of driving tasks into three maneuvers and the learning of sub-policies for each maneuver. They design a master policy to select the maneuver to execute in the current state. All policies, including the master policy and maneuver policies, are represented by fully-connected neural networks and trained using asynchronous parallel reinforcement learners. The paper concludes by demonstrating how this method can safely and smoothly drive a car on a highway.

Emuna et. al. (2020) [52], examined a deep reinforcement learning approach aimed at simulating human driver behavior to develop "human-like driving policies." The study concentrates on mixed traffic scenarios, encompassing both autonomous vehicles and human-controlled vehicles. The authors advocate for autonomous vehicles to demonstrate human-like driving behavior to maintain efficient and safe traffic flow.

Fernando et. al. (2021) [53], scrutinized the significance of precise behavior modeling in autonomous driving and scrutinizes the primary approaches and notable advancements made by researchers, with a focus on the potential of deep inverse reinforcement learning. The authors offer quantitative and qualitative assessments to substantiate their insights and explore promising avenues for future research breakthroughs.

Orgovan et. al. (2021) [54], focused on developing an algorithm for autonomous drifting using Machine Learning, specifically Reinforcement Learning. The algorithm utilized a model-free learning method known as Twin Delayed Deep Deterministic Policy Gradients (TD3), which was trained on six different tracks within the CARLA simulator, specifically designed for autonomous driving applications. The study also highlighted the effectiveness of Deep Reinforcement Learning (DRL) in addressing motion planning challenges, with inputs to the network including factors like destination, vehicle parameters, and outputs controlling vehicle actions such as steering, torque, and braking.

Chukamphaeng et. al. (2021) [16], discussed the utilization of end-to-end reinforcement learning for autonomous vehicles, which employs a single reinforcement learning model to develop the autonomous car. Additionally, it delineates the design of a novel efficient reward function aimed at accelerating the agent's learning process and constructing the car with only essential perceptions and sensors. The paper concludes that end-to-end reinforcement learning holds promise as an approach for autonomous driving vehicles.

Zou et. al. (2021) [55], introduced a deep imitation reinforcement learning (DIRL) framework comprising two main components: the perception module and the control module. The framework employs a deep deterministic policy gradient

algorithm (DDPG) for controlling self-driving vehicles via vision-based policies. The authors validated the effectiveness of the DRL framework using TORCS, an open racing car simulator.

Cao et. al. (2022) [56], introduced a method that combines reinforcement learning (RL) with a baseline rule-based driving policy to develop more intelligent driving policies for autonomous vehicles. Termed "confidence-aware reinforcement learning" (CARL), the proposed method is assessed through a case study involving driving in a two-lane roundabout scenario. The study demonstrates that the proposed approach surpasses both the pure RL policy and the baseline rule-based policy in terms of performance.

Chen et. al. (2022) [57], suggested an approach to address complex urban scenarios using an interpretable deep reinforcement learning method, aiming to reduce the sample complexity of reinforcement learning. Their method incorporates a latent environment model that generates a semantic bird's eye mask to connect with specific intermediate properties and elucidate the behaviors of the learned policy. Comparative tests conducted in a realistic driving simulator demonstrate that the method performs significantly better in urban scenarios with surrounding vehicles compared to numerous baseline approaches.

**Table 1. Summary of Related Work.**

Author& Year	Algorithm	Pros.	Cons.	Results
Sallab 2017 [32]	DRL	Novel framework addresses DRL limitations with RNNs and attention.	Challenges include observability, computational cost, and interpretability.	Successful lane keeping in both action types.
Liang 2018 [33]	CIRL	CIRL outperforms, extensive reward discussion.	Under-explored urban dynamics, rule reliance, exploration limits.	Framework achieves temporal abstraction, fine-grained control.
Wang 2018 [34]	DDPG	Low-cost autonomous driving with vision and RL integration.	Challenges in perception; limited to synthetic simulators.	Successful DRL implementation in TORCS simulator.
Nagesh Rao 2019 [35]	DRL, DDQN	Autonomous highway system mitigates unforeseen behaviors.	Relies on simulated traffic; limits real-world applicability.	Successful highway driving with improved DRL performance.
Pusse 2019 [36]	HyLEAP: DRL and approximate POMDP hybrid.	Good safety, smoothness; handles unseen situations.	Needing training data, requiring neural network updates, and facing uncertainty in POMDP planning.	Needs training data; NN updates; POMDP uncertainty.
Folkers 2019 [37]	DRL with proximal policy optimization.	Successful autonomous driving; NN training framework.	Long training, simulation reliance; limits real-world adaptability.	Successful driving with robust neural network framework.
Moghadam 2019 [38]	Hierarchical DRL for consistent high-level	Consistent high-level decisions in uncertain environments.	Complexity in uncertain environments; reliability	Hierarchical DRL for decision-making in autonomous driving.

	decisions.		challenges.	
Spielberg 2019 [39]	Deep neural network for actor and critic representation.	Good control sans detailed model; efficient neural networks.	Data volume requirement; risk of overfitting.	DRL control outperforms conventional methods in simulations.
You 2019[40]	RL with tabular and function approximation methods.	IRL for driving behavior; generates human-like policies.	Limited scenarios; data representation challenges.	Develops IRL algorithms for diverse driving behaviors.
Semnani 2020 [41]	Hybrid DRL-FMP for dynamic settings.	DRL-FMP Hybrid improves success rate, time-efficiency.	FMP lacks time optimization; GA3C-CADRL-NSL falters in dense scenarios.	Hybrid DRL-FMP excels in success rate, time, collision avoidance.
Liu 2020[42]	Off-policy CB algorithm with neural network policy modeling.	Custom CB algorithm outperforms non-customized version.	Limited real-world validation	Custom CB algorithm enhances driving smoothness, safety.
Muzahid 2020[43]	RL	Improves AV safety via threat assessment framework.	Limited to CARLA simulator, lacks generalization.	Proposes safety enhancement via threat prediction.
Rong 2020[44]	policy-guided trajectory planner	Combines RL with low-level planning for safe, optimal driving.	Lacks formal rule translation, ignores learning conflicts, limited applicability.	Effective handling of minimum-violation scenarios.
Liao 2020[45]	DRL with DDQN algorithm	Efficient DRL highway strategy with DDQN: Tested efficacy and safety.	Limited exploration, adaptability analysis, and risk discussion.	Successful highway overtaking with DDQN algorithm.
Kim 2020[46]	DRL	DRL-based agents excel in collision avoidance.	Insufficient discussion on real-world challenges and adaptability.	Self-driving agents outperformed humans and AEB in safety.
Hishmeh 2020[47]	CARLA simulation with Coach DRL Framework.	Utilizing Dueling DQN for AV obstacle avoidance.	Lack of specificity regarding addressed challenges or limitations in autonomous driving systems.	Dueling DQN evaluated, achieving Atari benchmark results, using CARLA simulation.
Lu 2020[48]	Hierarchical RL for decision-making and planning.	Hierarchical RL enhances efficiency and decision-making.	Optimizing behavior decisions amid vehicle dynamics complexities.	Comparative analysis of decision-making algorithms' simulation outcomes.
Josef 2020[49]	DRL for close-range terrain navigation	Novel DRL method enhances robot navigation in rough terrain.	Reliance on heavy computation limits practicality.	Novel DRL-based terrain navigation with improved performance.
Palanisamy 2020[50]	POSG	Realistic POSG for connected autonomous driving.	Deep RL struggles in dynamic environments.	Experiments validate proposed method.
Duan	Supervised	High performance,	Neglect of traffic	Effective performance,

2020[51]	imitation for driving decisions.	flexibility, adaptability, and task integration.	rules, safety, and environmental factors.	minimal collisions in diverse scenarios.
Emuna 2020[52]	DRL	Algorithm promising for autonomous driving, validated in simulation and real-world.	Generalization in RL, state-space variations, moving obstacles.	Algorithm achieves expert-like autonomous driving, validated in simulation and real vehicle, compared to others.
Fernando 2021[53]	Two-stage encoding with FCN, LSTM, and MaxEnt refinement.	Enhanced accuracy with deep learning, novel encoding, and improved MaxEnt formulation.	Limited applicability, uncertain effectiveness, require more research.	Advanced behavior prediction techniques demonstrated, with room for improvement.
Orgovan 2021[54]	RL	Increase RL awareness, propose drifting solution, evaluate experimentally.	Limited generalizability, results vary with parameters.	RL agent handles drifting, needs practical validation.
Chukamp haeng2021[16]	TD3, SAC	RL benchmark environment introduced.	Complexity, sensor evaluation, reward design, end-to-end learning, and validation highlighted.	Comparison of RL algorithms, optimal configurations with SAC and TD3, using direction-guided rewards
Zou 2021[55]	Deterministic strategies, actor-critic framework, empirical reuse in DQN.	Vision-based algorithm enhances efficiency, stability, safety, and generalizability in autonomous driving.	Future work needed for algorithm transferability and scalability to diverse driving conditions.	Algorithm surpasses traditional methods in stability, safety, efficiency, and adaptability.
Cao 2022[56]	CARL	CARL method improves safety, tackles RL uncertainty, utilizes self-learning, and enhances algorithm performance with more data.	CARL requires baseline policy, high RL confidence, limited scenario evaluation.	CARL outperformed RL and rule-based policy in roundabout simulation.
Chen 2022[57]	DRL	Generalization, efficiency, and exploration improved.	Sensitivity to sensor data, expertise, and hyperparameters.	Superior performance validated across two tasks with interpretability insights.

## G. Discussion

Table 1 provides a comprehensive review of various algorithms, their associated advantages, disadvantages, and results in the realm of autonomous driving. It highlights the diverse array of techniques utilized, ranging from deep reinforcement learning (DRL) to imitation learning and hierarchical reinforcement learning (RL), each aimed at tackling specific challenges such as lane keeping, urban dynamics, and autonomous highway driving.

While each algorithm offers unique benefits like improved safety, efficiency, and adaptability, they also come with their own set of limitations, including computational costs, reliance on simulated environments, and challenges in real-world applicability. Notably, the effectiveness of these approaches is assessed through simulations and real-world experiments, showcasing their potential for practical implementation in autonomous driving systems.

However, the review reveals common gaps across these methodologies, including limited generalizability, uncertainty in effectiveness, and the need for further research to address scalability and transferability issues. These findings emphasize the ongoing challenges in autonomous driving research and underscore the importance of continued exploration and development in this rapidly evolving field.

## H. Conclusion

This paper presents a comprehensive review on deep RL frameworks for autonomous driving, that have become a promising and thriving research field. Deep RL provides the necessary robustness to design adaptive and intelligent driving policies for the complex and unstructured dynamic environment. The results of the review indicate that RL algorithms are superior to traditional methods and can become a game-changer for autonomous driving. Study outcomes reveal that deep RL has not yet provided satisfactory solutions for dynamic and real-world environments. While deep RL incurs fewer assumptions about the behavior of physical and social phenomena, it is still subject to sample randomness and distributional shift. Also, these techniques are computationally expensive and require abundant data and computational resources. Further research is needed to address pressing practical issues to boost the autonomous driving industry's development and make the future of autonomous vehicles more promising. Future research can be extended by exploring potential techniques that can bridge the gap between simulation and the real world, handling non-stationary environments, requiring less computation, and improving the scalability of these systems.

## I. References

- [1] C. Cui *et al.*, "A survey on multimodal large language models for autonomous driving," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 958–979.
- [2] L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, "End-to-end autonomous driving: Challenges and frontiers," *arXiv preprint arXiv:2306.16927*, 2023.
- [3] Z. Bao, S. Hossain, H. Lang, and X. Lin, "A review of high-definition map creation methods for autonomous driving," *Eng Appl Artif Intell*, vol. 122, p. 106125, 2023, doi: <https://doi.org/10.1016/j.engappai.2023.106125>.
- [4] D. Atrushi and A. M. Abdulazeez, "Human Gait Recognition Based on Deep Learning: A Review," *Indonesian Journal of Computer Science*, vol. 13, no. 1, 2024.
- [5] C. Al-Atroshi and A. M. Abdulazeez, "Predictions of Early Hospitalization of Diabetes Patients Based on Deep Learning: A Review: Machine Learning," *Indonesian Journal of Computer Science*, vol. 13, no. 1, 2024.
- [6] Z. Gu *et al.*, "Safe-state enhancement method for autonomous driving via direct hierarchical reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

- 
- [7] X. He *et al.*, "Fear-neuro-inspired reinforcement learning for safe autonomous driving," *IEEE Trans Pattern Anal Mach Intell*, 2023.
- [8] S. Salih and A. M. Abdulazeez, "Classification of Diabetic Retinopathy Images through Deep Learning Models-Color Channel Approach: A Review," *Indonesian Journal of Computer Science*, vol. 13, no. 1, 2024.
- [9] J. Lu, L. Han, Q. Wei, X. Wang, X. Dai, and F.-Y. Wang, "Event-triggered deep reinforcement learning using parallel control: A case study in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, 2023.
- [10] J. Hu, H. Kong, T. Liu, and Y. Meng, "Autonomous Motion Decision-making based on Deep Reinforcement Learning for Autonomous Driving," in *2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, 2022, pp. 1–6. doi: 10.1109/CVCI56766.2022.9964721.
- [11] B. Ferzo and A. M. Abdulazeez, "Image Denoising Techniques Using Unsupervised Machine Learning and Deep Learning Algorithms: A Review," *Indonesian Journal of Computer Science*, vol. 13, no. 1, 2024.
- [12] Z. Huang, H. Liu, J. Wu, and C. Lv, "Conditional predictive behavior planning with inverse reinforcement learning for human-like autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, 2023.
- [13] H. M. Zangana and A. M. Abdulazeez, "Developed Clustering Algorithms for Engineering Applications: A Review," *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, vol. 4, no. 2, pp. 147–169, 2023.
- [14] A. K. Shakya, G. Pillai, and S. Chakrabarty, "Reinforcement learning algorithms: A brief survey," *Expert Syst Appl*, vol. 231, p. 120495, 2023, doi: <https://doi.org/10.1016/j.eswa.2023.120495>.
- [15] S. Aradi, "Survey of Deep Reinforcement Learning for Motion Planning of Autonomous Vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 740–759, Feb. 2022, doi: 10.1109/TITS.2020.3024655.
- [16] N. Chukamphaeng, K. Pasupa, M. Antenreiter, and P. Auer, "Learning to Drive with Deep Reinforcement Learning," in *KST 2021 - 2021 13th International Conference Knowledge and Smart Technology*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 147–152. doi: 10.1109/KST51265.2021.9415770.
- [17] S. Mohammed, A. Argun, and G. Ascheid, "A Unified Approach to Autonomous Driving in a High-Fidelity Simulator Using Vision-Based Reinforcement Learning," in *2024 IEEE/SICE International Symposium on System Integration (SII)*, 2024, pp. 1093–1098. doi: 10.1109/SII58957.2024.10417385.
- [18] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [19] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.
- [20] F. Youssef and B. Houda, "Comparative Study of End-to-end Deep Learning Methods for Self-driving Car," *International Journal of Intelligent Systems and Applications*, vol. 12, no. 5, pp. 15–27, Oct. 2020, doi: 10.5815/ijisa.2020.05.02.
- [21] D. Tagesson, "A Comparison Between Deep Q-learning and Deep Deterministic Policy Gradient for an Autonomous Drone in a Simulated Environment." 2021.
- [22] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

- [23] V. Veeriah, H. van Seijen, and R. S. Sutton, "Forward actor-critic for nonlinear function approximation in reinforcement learning," in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 2017, pp. 556–564.
- [24] T. Hester *et al.*, "Deep q-learning from demonstrations," in *Proceedings of the AAAI conference on artificial intelligence*, 2018.
- [25] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*, PMLR, 2018, pp. 1587–1596.
- [26] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*, PMLR, 2018, pp. 1861–1870.
- [27] A. Feher, S. Aradi, and T. Becsi, "Q-learning based Reinforcement Learning Approach for Lane Keeping," in *2018 IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2018, pp. 31–36. doi: 10.1109/CINTI.2018.8928230.
- [28] B. Qin, W. He, B. Zhang, and J. Li, "A Multi-Agent Reinforcement Learning Framework with Recurrent Communication Module for Traffic Light Control," in *2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, 2021, pp. 117–121. doi: 10.1109/ICISCAE52414.2021.9590701.
- [29] M. K. Ghaith, M. M. Rehaan, N. Shouman, Y. Abdalla, and O. M. Shehata, "Utilizing Multi-Agent Deep Reinforcement Learning for Autonomous Intersection Management Systems: A Promising Approach," in *2023 4th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*, 2023, pp. 115–119. doi: 10.1109/AIRC57904.2023.10303022.
- [30] J. Xia, X. Zhu, Z. Liu, Y. Luo, Z. Wu, and Q. Wu, "Research on Collision Avoidance Algorithm of Unmanned Surface Vehicle Based on Deep Reinforcement Learning," *IEEE Sens J*, vol. 23, no. 11, pp. 11262–11273, 2023, doi: 10.1109/JSEN.2022.3222575.
- [31] Z. Shen, Y. Duan, Y. Cong, W. Li, H. Du, and W. Zhu, "Deep Reinforcement Learning-based Multi-AMR Path Planning Algorithm," in *2023 42nd Chinese Control Conference (CCC)*, 2023, pp. 1984–1987. doi: 10.23919/CCC58697.2023.10240541.
- [32] A. El Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep Reinforcement Learning framework for Autonomous Driving," Apr. 2017, doi: 10.2352/ISSN.2470-1173.2017.19.AVM-023.
- [33] X. Liang, T. Wang, L. Yang, and E. Xing, "CIRL: Controllable Imitative Reinforcement Learning for Vision-based Self-driving," 2018.
- [34] S. Wang, D. Jia, and X. Weng, "Deep Reinforcement Learning for Autonomous Driving," Nov. 2018, [Online]. Available: <http://arxiv.org/abs/1811.11329>
- [35] S. Nagesh Rao, H. E. Tseng, and D. Filev, *Autonomous Highway Driving using Deep Reinforcement Learning*. 2019. doi: 10.0/Linux-x86\_64.
- [36] Institute of Electrical and Electronics Engineers, *Hybrid Online POMDP Planning and Deep Reinforcement Learning for Safer Self-Driving Cars*. 2019.
- [37] Institute of Electrical and Electronics Engineers, *Controlling an Autonomous Vehicle with Deep Reinforcement Learning*. 2019.



- [38] M. Moghadam and G. H. Elkaim, "A Hierarchical Architecture for Sequential Decision-Making in Autonomous Driving using Deep Reinforcement Learning," Jun. 2019, [Online]. Available: <http://arxiv.org/abs/1906.08464>
- [39] S. Spielberg, A. Tulsyan, N. P. Lawrence, P. D. Loewen, and R. Bhushan Gopaluni, "Toward self-driving processes: A deep reinforcement learning approach to control," *AICHE Journal*, vol. 65, no. 10, Oct. 2019, doi: 10.1002/aic.16689.
- [40] C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning," *Rob Auton Syst*, vol. 114, pp. 1–18, Apr. 2019, doi: 10.1016/j.robot.2019.01.003.
- [41] S. H. Semnani, H. Liu, M. Everett, A. De Ruiter, and J. P. How, "Multi-Agent Motion Planning for Dense and Dynamic Environments via Deep Reinforcement Learning," *IEEE Robot Autom Lett*, vol. 5, no. 2, pp. 3221–3226, Apr. 2020, doi: 10.1109/LRA.2020.2974695.
- [42] Z. Liu, Z. Wang, B. Yang, and K. Nakano, "Learning Personalized Discretionary Lane-Change Initiation for Fully Autonomous Driving Based on Reinforcement Learning," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 457–463. doi: 10.1109/SMC42975.2020.9283222.
- [43] A. Jafar Md Muzahid, S. Fauzi Kamarulzaman, and M. Abdur Rahim, "Learning-Based Conceptual framework for Threat Assessment of Multiple Vehicle Collision in Autonomous Driving," 2020, doi: 10.1109/ETCCE51779.2020.9350869/20/\$31.00.
- [44] J. Rong and N. Luan, "Safe Reinforcement Learning with Policy-Guided Planning for Autonomous Driving," in *2020 IEEE International Conference on Mechatronics and Automation, ICMA 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 320–326. doi: 10.1109/ICMA49215.2020.9233522.
- [45] J. Liao, T. Liu, X. Tang, X. Mu, B. Huang, and D. Cao, "Decision-making strategy on highway for autonomous vehicles using deep reinforcement learning," *IEEE Access*, vol. 8, pp. 177804–177814, 2020, doi: 10.1109/ACCESS.2020.3022755.
- [46] M. Kim, S. Lee, J. Lim, J. Choi, and S. G. Kang, "Unexpected collision avoidance driving strategy using deep reinforcement learning," *IEEE Access*, vol. 8, pp. 17243–17252, 2020, doi: 10.1109/ACCESS.2020.2967509.
- [47] L. Hishmeh and F. Awad, "Deer in the Headlights: Short Term Planning via Reinforcement Learning Algorithms for Autonomous Vehicles," in *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, Institute of Electrical and Electronics Engineers Inc., Apr. 2020, pp. 255–260. doi: 10.1109/ICICS49469.2020.239561.
- [48] Y. Lu, X. Xu, X. Zhang, L. Qian, and X. Zhou, "Hierarchical Reinforcement Learning for Autonomous Decision Making and Motion Planning of Intelligent Vehicles," *IEEE Access*, vol. 8, pp. 209776–209789, 2020, doi: 10.1109/ACCESS.2020.3034225.
- [49] S. Josef and A. Degani, "Deep Reinforcement Learning for Safe Local Planning of a Ground Vehicle in Unknown Rough Terrain," *IEEE Robot Autom Lett*, vol. 5, no. 4, pp. 6748–6755, Oct. 2020, doi: 10.1109/LRA.2020.3011912.
- [50] IEEE Computational Intelligence Society, International Neural Network Society, Institute of Electrical and Electronics Engineers, and IEEE World Congress on

---

Computational Intelligence (2020: Online), *Multi-Agent Connected Autonomous Driving using Deep Reinforcement Learning*. 2020.

- [51] J. Duan, S. E. Li, Y. Guan, Q. Sun, and B. Cheng, "Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data," *IET Intelligent Transport Systems*, vol. 14, no. 5, pp. 297–305, May 2020, doi: 10.1049/iet-its.2019.0317.
- [52] R. Emuna, A. Borowsky, and A. Biess, "Deep Reinforcement Learning for Human-Like Driving Policies in Collision Avoidance Tasks of Self-Driving Cars," Jun. 2020, [Online]. Available: <http://arxiv.org/abs/2006.04218>
- [53] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Deep Inverse Reinforcement Learning for Behavior Prediction in Autonomous Driving: Accurate Forecasts of Vehicle Motion," *IEEE Signal Process Mag*, vol. 38, no. 1, pp. 87–96, Jan. 2021, doi: 10.1109/MSP.2020.2988287.
- [54] L. Orgován, T. Bécsi, and S. Aradi, "Autonomous drifting using reinforcement learning," *Periodica Polytechnica Transportation Engineering*, vol. 49, no. 3, pp. 292–300, Sep. 2021, doi: 10.3311/PPTR.18581.
- [55] Q. Zou, K. Xiong, Q. Fang, and B. Jiang, "Deep imitation reinforcement learning for self-driving by vision," *CAAI Trans Intell Technol*, vol. 6, no. 4, pp. 493–503, Dec. 2021, doi: 10.1049/cit2.12025.
- [56] Z. Cao, S. Xu, H. Peng, D. Yang, and R. Zidek, "Confidence-Aware Reinforcement Learning for Self-Driving Cars," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 7419–7430, Jul. 2022, doi: 10.1109/TITS.2021.3069497.
- [57] J. Chen, S. E. Li, and M. Tomizuka, "Interpretable End-to-End Urban Autonomous Driving With Latent Deep Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5068–5078, Jun. 2022, doi: 10.1109/TITS.2020.3046646.