Evaluating Scenario-Based Decision-Making for Interactive Autonomous Driving Using Rational Criteria: A Survey

Zhen Tian, Zhihao Lin, Dezong Zhao, *Senior Member, IEEE*, Wenjing Zhao, David Flynn, *Senior Member, IEEE*, Shuja Ansari, and Chongfeng Wei, *Senior Member, IEEE*

***Abstract*—Autonomous vehicles (AVs) promise substantial gains in safety, reliability, and decarbonization, yet safe and efficient interaction in dynamic, heterogeneous traffic remains a key barrier to large-scale deployment. Deep reinforcement learning (DRL) has emerged as a data-driven approach for learning adaptive decision policies that handle complex, unpredictable environments better than rule-based methods. However, different scenarios impose distinct requirements, necessitating scenario-specific algorithms. This survey systematically reviews DRL for four typical scenarios (highways, on-ramp merging, roundabouts, and unsignalized intersections), summarizes road features and recent advances, and evaluates methods using five criteria: driving safety, driving efficiency, training efficiency, unselfishness, and interpretability (DDTUI). Each DDTUI criterion is analyzed with respect to the reviewed algorithms. In addition, a dedicated scenario-centric learning transferability analysis is introduced that systematically evaluates whether each reviewed method demonstrates scene-specific learning improvements and assesses how effectively their designs transfer across the four scenarios. Finally, the challenges for future DRL-based decision-making algorithms are summarized.**

***Index Terms*—Interactive autonomous driving, decision making, deep reinforcement learning, typical scenarios, rationale evaluation.**

I. Introduction

AUTONOMOUS vehicles (AVs) face major challenges in making reliable decisions when interacting with human driven vehicles (HDVs), largely because HDV intentions are difficult to predict. Road crashes cause millions of deaths and severe injuries worldwide, and since 2021 more than 900 Tesla crashes involving driver assistance systems have been reported [1]. Despite unresolved safety issues, the AV fleet is projected to exceed 50 million by 2024 [2], which underscores the urgency of safer decision making. Robust decision systems can reduce crashes caused by human error, such as fatigue, distraction, and delayed reactions [3].

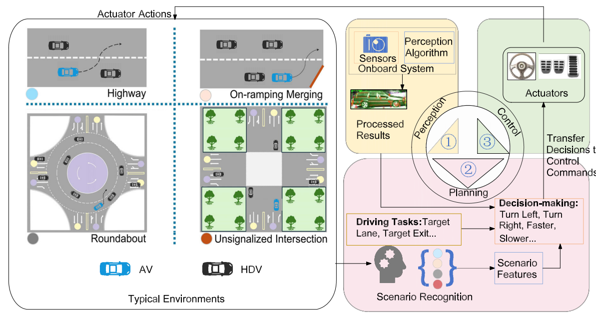


Fig. 1. Autonomous driving in different scenarios.

Typical driving scenarios include highways, on ramp merging, roundabouts, and unsignalized intersections, each with distinct road features and requirements, as in Fig. 1. For example, on ramp merging requires completing lane changes before any obstructed roadway, while roundabout navigation requires exiting at the intended point. Meeting such requirements depends on precise and timely decision making in real time. Operational decision support follows the perception, planning, and control pipeline. Perception uses onboard sensors and processes the data with detection algorithms such as YOLO methods [4], [5]. The planning module selects maneuvers, and produces feasible trajectories. The control module converts these trajectories into control commands.

In the control stage, common controllers include PID, Linear Quadratic Regulator (LQR), Model Predictive Control (MPC), Pure Pursuit Controller (PPC), Fuzzy Logic Control (FLC), and the Stanley controller. PID reduces steady state error and damps oscillations but adapts poorly to rapidly changing conditions [6]. LQR minimizes a quadratic cost to balance tracking and control effort via optimal state feedback; it is often used for yaw control but is limited by linear model assumptions [7]. MPC predicts over a finite horizon and optimizes commands, which is effective under uncertainty and constraints [8]. PPC uses a simple kinematic model to compute

Received 8 December 2024; revised 18 April 2025 and 12 August 2025; accepted 1 November 2025. Date of publication 15 December 2025; date of current version 18 February 2026. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) Innovation Fellowship under Grant EP/S001956/2, in part by the Royal Academy of Engineering/Leverhulme Trust Research Fellowship under Grant LTRF-2425-21-151, in part by the EPSRC Digital Twinning Research Hub for Decarbonising Transport (TransiT) under Grant EP/Z533221/1, and in part by the EPSRC Research Hub for Decarbonised Adaptable and Resilient Transport Infrastructures (DARe) under Grant EP/Y024257/1. The Associate Editor for this article was F. R. Yu. *(Zhen Tian and Zhihao Lin contributed equally to this work.) (Corresponding author: Dezong Zhao.)*

The authors are with the School of Engineering, University of Glasgow, G12 8QQ Glasgow, U.K. (e-mail: 2620920z@student.gla.ac.uk; 2800400l@student.gla.ac.uk; dezong.zhao@glasgow.ac.uk; wenjing.zhao@glasgow.ac.uk; david.flynn@glasgow.ac.uk; shuja.ansari@glasgow.ac.uk; chongfeng.wei@glasgow.ac.uk).

Digital Object Identifier 10.1109/TITS.2025.3636070

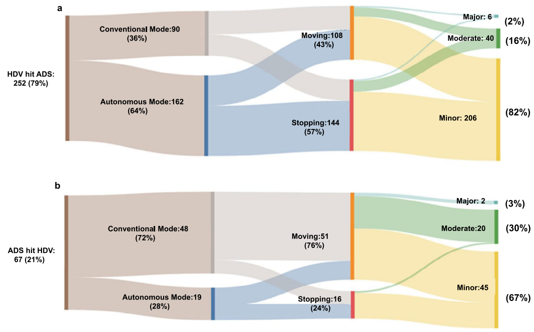


Fig. 2. Rear-end accident conditions between ADS and HDV: (a) Rear-end accidents that HDV hit an ADS from behind; (b) Rear-end accidents that ADS hit an HDV from behind [15].

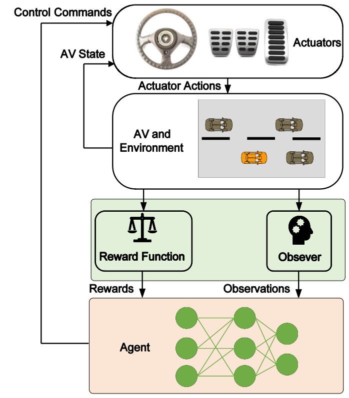


Fig. 3. DRL-based autonomous driving system.

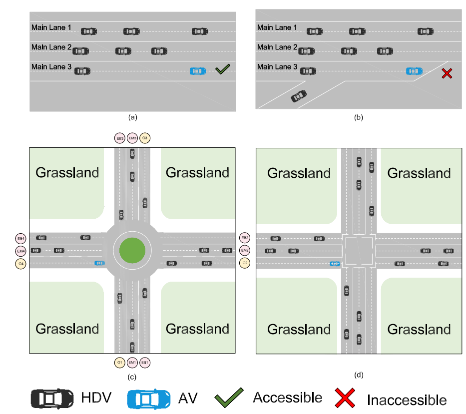


Fig. 4. Example scenarios of autonomous driving: (a) highway; (b) on-ramp merging; (c) roundabout; (d) unsignalized intersection.

early steering targets [9]. FLC encodes expert rules when accurate models are unavailable [10]. The Stanley controller combines lateral cross track and heading errors to refine steering [11].

Interactions between AVs and HDVs are complex, so vehicles must make continuous decisions such as lane changes and braking. Common approaches include model based control, simple guidance strategies, and learning based methods.

Model based approaches fall into four groups. (1) Intention or trajectory prediction relies heavily on rule based classification. For example, [12] predicts within a fixed window, which may be shorter than an actual lane change. (2) Robust control, such as time varying min max MPC, addresses nonlinearities [13] but acts on worst case assumptions that are rare in practice and can degrade driving quality through overly slow behavior. (3) Game theory, including cooperative and non cooperative games, depends on equilibrium models that do not capture uncertainty and irregular interactions in real traffic. (4) Collision avoidance and Voronoi diagram methods [14] struggle to respond safely to movable objects. Real world crash data involving vehicles with ADS indicate that 79 % of accidents involve HDVs striking AVs and 21 % involve AVs striking HDVs [15], as in Fig. 2. Therefore, model based methods alone are insufficient for interactive driving with HDVs. achieving collision-free interactions with HDVs are still to be addressed.

Simple guidance methods are attractive because they avoid predicting HDV intentions and avoid overly cautious assumptions [16]. A common example is the artificial potential field based on attractive and repulsive forces. However, it often treats surrounding risk as uniform and ignores target reachability, whereas in practice the front of a vehicle is more dangerous and not all targets are reachable [17].

Learning-based methods enhance collision-free interaction by exploring actions under uncertainty and diverse HDV behaviors. Within machine learning [18], supervised learning enables tasks such as classification [19] but requires costly labeling, unsupervised learning avoids labels but often generalizes poorly in dynamic traffic, and reinforcement learning is well suited to sequential decision making, allowing agents to acquire safe control policies in changing environments [20].

Deep reinforcement learning (DRL) combines deep learning [21], [22] with reinforcement learning. By using deep neural networks to approximate value functions and policies, DRL learns directly from sensory inputs and handles complex, real time decision making better than classical RL. For example, [23] applies DRL to collision free path planning in the presence of surrounding obstacles.

Fig. 3 shows the DRL loop for autonomous driving. The agent observes the environment, selects actions, and the environment returns new states and rewards. The reward function reflects safety, efficiency, and driving rules, and this feedback guides the policy toward better driving behavior.

DRL has shown strong capability in emergency handling, a critical requirement for real-world driving. For example, [24] demonstrates that DRL can learn collision avoidance under sudden changes. Recent work covers many scenarios [25], [26], [27], [28], [29], [30], but each setting poses different demands. Highways emphasize collision avoidance while keeping high speed, whereas ramps contain blocked regions not present on highways (Fig. 4). Scenario-specific algorithms are therefore needed, and evaluations should reflect the priorities of users, manufacturers, and broader traffic systems to ensure meaningful progress.

Multi-agent reinforcement learning (MARL) extends single-agent RL to Markov games [31], enabling more collaborative behaviors. This leads to more efficient merging and lane changing with lower collision risk [32], as well as signal-free intersection management where vehicles negotiate right of way [33]. Viewing traffic as a collaborative multi-agent game enables distributed decisions and implicit coordination, reducing deadlocks and improving safety and flow [34].

DRL offers strong potential for complex decision making in autonomous driving, but it often converges more slowly than classical control. The delay stems from iterative interaction and gradient optimization over deep networks, with higher computational cost [35]. Biologically inspired architectures can model nonlinear dynamics and learn from raw data, with low interpretable decision-making [36]. Efficiency can be improved by simplifying network designs, such as replacing recurrent models with compact convolutional structures [37] and removing unnecessary parameters, since excess attributes add computation [38]. Larger hidden layers and oversized networks further raise complexity [39].

This survey reviews DRL based algorithms for autonomous interactive driving, organized by four scenarios: highways, on ramp merging, roundabouts, and unsignalized intersections. The adaptation to real world conditions using five factors—driving safety [40], [41], driving efficiency [42], training efficiency [29], unselfishness [43], [44], and interpretability [45], [46] (DDTUI) has been accessed. A consistent protocol has been applied: a factor is credited only when the algorithm both incorporates it and verifies it experimentally; for example, safety requires evidence such as fewer collisions or consistently maintained safe distance d_s. In addition, this paper introduces a standalone scenario-centric chapter on learning and transferability (SCLT). For each of the four scenarios and for every reviewed method, the chapter records whether the framework contains scene-specific learning improvements. Scenario-specific improvements are examined through tailored state representations, specialized objectives beyond weight tuning, and scenario-focused learning mechanisms that are validated experimentally. Cross-scenario transferability is evaluated by analyzing structural dependencies in topology, rules, and dynamics. Learning advances are distinguished from domain-specific heuristics through four summary tables and transfer analysis, with design generalization being clarified across highways, merging zones, roundabouts, and intersections.

To guide future work, methods have been grouped into four families that reflect properties of DRL and autonomous driving: *network structure improved DRL* and *policy improved DRL* (rooted in DRL design), and *robust control assisted DRL* and *traffic model assisted DRL* (rooted in control theory and traffic modeling). The remainder of the survey is organized as follows: Section II describes scenario features and driving tasks; Section III explains the rationale for DDTUI; Sections IV to VII evaluate methods for the four scenarios using DDTUI; Section VIII provides conclusions and discussion.

II. Road Features and Driving Tasks

This section provides the road features and driving tasks for AVs in the scenarios of highways, on-ramping merging, roundabouts, and unsignalized intersections.

A. Highways

1) Road Features and Driving Tasks: Highways support long-distance travel with minimal interruption and are built for safety, efficiency, and reduced environmental impact. Safety is aided by wide lanes and clear signage, efficiency by lane layouts that smooth traffic flow, and environmental impact by careful routing. The U.S. Interstate Highway System provides nationwide mobility and economic connectivity [47], while Germany's Autobahn, including sections without speed limits, shows that high-speed travel can still be managed safely [48].

2) Example Highway Scenario: Fig. 4(a) illustrates a three-lane highway where the AV travels in lane 3 and interacts with HDVs across all lanes. With no external disturbances considered, safety risks mainly result from potential collisions. In car following, the AV adjusts acceleration to maintain headway, but overly conservative behavior reduces efficiency. When headway becomes limited, the AV may change lanes to maintain throughput, which introduces interaction risks. The central challenge is therefore to avoid collisions while sustaining a consistently high speed.

B. On Ramp Merging

1) Road Features and Driving Tasks: Ramps connect surface streets to highways and manage the transition between different speed regimes. Key design goals are safety, efficiency, and compact space use. Vehicles must accelerate or decelerate smoothly, merge with limited disruption, and fit within constrained urban layouts, supported by designs such as cloverleaf interchanges and HOV ramps [49], [50].

2) Comparison With Highways: Highways enable sustained high-speed travel on long, mostly straight multi-lane segments, while ramps handle speed changes and merging on shorter, often curved sections. As a result, interactions on ramps require earlier maneuver commitment than on highways.

3) Example Ramp Scenario: Fig. 4(b) shows a three-lane setting with two mainline lanes and one ramp lane. The AV must merge into the mainline before the ramp ends, while interacting with HDVs and avoiding a static obstruction that blocks the ramp. Waiting for large gaps increases safety but reduces efficiency and ramp capacity. The task is to avoid the conflicts while maintaining reasonable speed.

C. Roundabouts

1) Road Features and Driving Tasks: Roundabouts are designed to improve flow and safety by reducing approach speeds and eliminating severe conflict points. Examples range from the roundabout with Folon's obelisk, featuring a central island and circulating lanes [51], to the large multilane junction at Place Charles de Gaulle in Paris, where twelve avenues converge around the Arc de Triomphe [52].

2) Example Roundabout Scenario: Fig. 4(c) shows an AV entering from EB4 with three possible exits: O1, O2, and O3. For O1, the AV can stay in the outer lane. For O2, it may either remain in the outer lane for a safer but longer route or move to the inner lane for a shorter path while managing interaction risks with circulating vehicles. For O3, the AV must merge

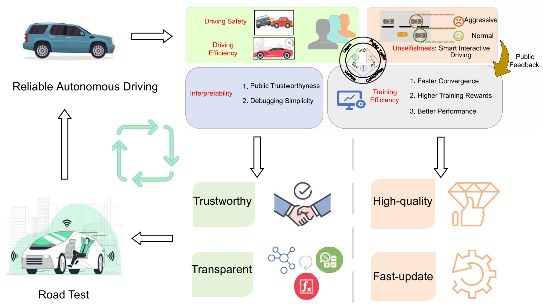


Fig. 5. The importance and necessaries of achieving DDTUI in real-world autonomous driving.

to the inner lane early and then return to the outer lane near the exit. The main challenge is balancing safety and efficiency while navigating circular trajectories among HDVs.

D. Unsignalized Intersections

1) Road Features and Driving Tasks: Unsignalized intersections are key points where two or more roads meet without traffic signals. They manage flows from different directions by assigning priority and organizing movements to reduce congestion and enhance safety. A representative design is the *Diverging Diamond Interchange (DDI)* [53].

2) An Example of an Unsignalized Intersection: Fig. 4(d) shows a three-lane unsignalized intersection serving moderate to heavy traffic. Vehicles approach from all four directions, each with dedicated lanes for different movements. The surrounding area is open grassland, and the central space where the roads meet is wide enough to support turning maneuvers from any direction. This central zone helps prevent bottlenecks and maintains smooth traffic flow.

III. Rationale of the Evaluation Factors

To evaluate adaptation to real-world driving, we consider five rational factors: driving safety, driving efficiency, training efficiency, unselfishness, and interpretability. As shown in Fig. 5, safety and efficiency form the foundation of any autonomous driving system. Training efficiency supports rapid convergence and deployment. Unselfishness reflects socially compliant interaction with HDVs, and interpretability enables trust, diagnosis, and accountability.

A. Driving Safety

Driving safety is essential for autonomous vehicles [54], [55]. We assess safety mainly through collision frequency with surrounding vehicles [56], [57]. Collision avoidance depends on relative speed, distance, and trajectory to estimate risk [58]. Some approaches search for decisions that avoid collisions while maintaining efficiency [59], [60], while others use rule-based commands under unsafe conditions, such as stopping when interaction zones or stop lines are detected [61].

Recent progress includes Bayesian calibration for improving IDM fidelity [62], a reference driver model for highway merging safety assessment [63], a "reasonable driver" standard for evaluating automated vehicle behavior [64], and a survey on explainable AI for transparency and trust [65]. These advances enhance model calibration, scenario-specific safety assessment, standardized evaluation, and interpretability.

B. Driving Efficiency

Driving efficiency refers to an AV's ability to maintain high average speed while adapting to traffic. Its influence extends to road capacity, user experience, and energy use. Efficient operation increases throughput and reduces congestion [66], [67], [68], shortens travel time and improves comfort [69], [70], [71], and lowers energy consumption [72], [73].

Recent advances improve efficiency through better routing and flow control. A DRL framework reduces distance and turns to produce near-optimal routes and shorter travel time [74]. A stability-informed artificial potential field uses a Lyapunov threshold to adjust repulsion, avoiding local minima and generating smoother, shorter paths in congestion [75]. Sparse AV control shows that even a small proportion of AVs can damp stop-and-go waves and raise average speed in mixed traffic [76]. For roundabouts, an evolutionary coordination method balances mainline and merging flows to reduce queues and delay [77]. These studies demonstrate how advanced decision strategies can enhance driving efficiency.

C. Training Efficiency

Training efficiency directly affects the time and resources needed to develop a fully functional autonomous driving system. Faster training shortens development cycles and allows more effort to be spent on fine-tuning and extensive testing. Many studies improve training speed by adding auxiliary training mechanisms or simplifying network structures [29], [78], [79], [80]. Higher efficiency also reduces computational load, which lowers hardware wear and maintenance needs.

D. Unselfishness

Unselfishness reflects an AV's willingness to consider the intentions of surrounding HDVs and supporting cooperative flow. It measures how well the AV anticipates others and balances its own progress with the needs of nearby drivers. Intention recognition is central to cooperative behavior. Reviews of interaction-aware driving highlight predictors for tasks such as car following and lane changing [81], [82], [83]. Although much prior work optimizes only the ego vehicle [84], [85], evidence increasingly supports policies that yield, create gaps, or modulate speed to reduce conflicts. Such cooperation avoids both aggressive and overly cautious behavior, smooths traffic, and improves safety and efficiency for all road users.

E. Algorithm Interpretability

Interpretability is essential because DRL policies must make decisions that are logical and explainable [45], [46]. Clear structure improves transparency and supports safety assurance.

Common tools include behavior and policy visualization [86], [87] and surrogate models that provide human-readable explanations [88], [89], [90]. Rule-based evaluations such as *FAST* examine fairness, accountability, sustainability, and transparency [91]. Structure-adapted methods use standardized benchmarks and sanity checks to validate explanations [92], [93], [94], [95], and some approaches modify network architectures to improve clarity [96], [97]. Human-grounded studies explore how easily people understand the reasoning behind decisions [98]. Incorporating traffic models and explicit mathematical structure can also enhance interpretability.

Recent work integrates large language models to attach rationales to actions and align policies with traffic rules. Examples include an LLM-guided driving policy, DriveMLM, which generates decisions with textual explanations [99], and DriveGPT4, which predicts controls and answers queries in real time [100]. Other recent developed models inject domain knowledge by a chain of instructions, or couple Teacher LLM with an attention-based Student DRL policy using risk metrics and retrieved scenarios. Evaluations on CARLA [101] and nuPlan [102] show that these frameworks make DRL policies more explainable for autonomous driving.

IV. Deep Reinforcement Learning-Based Decision-Making on Highways

A. Single-Factor Methods for Highway Driving

Many studies address only one of the five key factors. For example, [103] integrates a Double Deep Q-Network (DDQN) with two safety modules: a handcrafted module based on safety rules that maintain a distance (d_s), and a learned module that extracts safety patterns from driving data. Combining these modules strengthens the policy's safety.

Moreover, deep deterministic policy gradients (DDPG) have been used to improve driving efficiency by overtaking surrounding vehicles in [104]. The overtaking-oriented training is achieved by adding a high reward for overtaking maneuvers. The reward function for overtaking is formulated as [104]:

$$R_{\text{overtaking}} = R_{\text{lane\_keeping}} + 100 \times (n - \text{RacePos}) \tag{1}$$

where $R_{\text{lane\_keeping}}$ is the reward for lane-keeping, n is the total number of vehicles in a given episode, and *RacePos* reflects the number of vehicles in front of the AV. Therefore, the larger the *RacePos*, the smaller the $R_{\text{overtaking}}$. Although safety rewards are applied, the collision rate increases with the frequency of overtaking.

Furthermore, non-linear model predictive control (NMPC) has been integrated with DDQN to maintain safe highway driving in [105]. NMPC inherently incorporates vehicle dynamics as constraints into its optimization, ensuring that the control inputs from the DRL agent remain within safe and feasible bounds [105]:

$$\begin{aligned}
\min_{x(t),u(t)} \quad & \int_0^T \left(e(t)^\top Q\, e(t) + r\,\delta^2(t) + r\,u^2(t)\right) \mathrm{d}t \\
\text{s.t.} \quad & \dot{x}(t) = f\,(x(t), u(t)), \quad e_{y_{\min}} \le e_y(t) \le e_{y_{\max}}, \\
& e_{\psi_{\min}} \le e_\psi(t) \le e_{\psi_{\max}}, \\
& \delta_{\min} \le \delta(t) \le \delta_{\max}, \quad u_{\min} \le u_1(t) \le u_{\max}
\end{aligned} \tag{2}$$

where T is the prediction horizon, $e(t)$ is the error vector to be regulated to zero, and $Q = \text{diag}(q_1, q_2)$ is a diagonal matrix of tracking weights. The control effort weight is denoted by r. The steering angle is represented by $\delta(t)$, and the control input is $u(t)$. The state vector is $x(t)$, and f represents the system dynamics. $e_y(t)$ and $e_\psi(t)$ are the lateral position error and heading angle error, respectively. The variables $e_{y_{\min}}$, $e_{y_{\max}}$, $e_{\psi_{\min}}$, $e_{\psi_{\max}}$, $\delta_{\min}$, $\delta_{\max}$, $u_{\min}$, and $u_{\max}$ are the minimum and maximum admissible values for the lateral position error, heading angle error, steering angle, and control input, respectively. NMPC improves the interpretability of safe control by providing a clear mathematical formulation that integrates the system's constraints with the agent's decision-making [106], [107], [108].

In [109], a policy gradient method with hard constraints ensures highway safety by keeping the vehicle away from risky boundaries. Cooperative lane changing improves unselfishness in [110]. Interpretability is enhanced by combining DRL with imitation learning that uses expert demonstrations [111]. Training efficiency increases when a spatial attention module is integrated into a deep Q network [112].

B. Dual-Factor Methods for Highway Driving

Additionally, two of the five considered factors are integrated in some recent studies. The Intelligent Driver Model (IDM) [113] has been incorporated into the DDQN for highway driving in [114]. The IDM prevents collisions during car-following and therefore, the integration of DDQN with IDM enhances both the driving safety and interpretability in highway driving. The IDM is formulated as [114]:

$$U_{\text{IDM}} = U_{\max}\left[1 - \left(\frac{v_{\text{FV}}}{v_e}\right)^4 - \left(\frac{g^*}{g}\right)^2\right] \tag{3}$$

where $U_{\max}$ is the maximum acceleration of the AV, v_e is the expected velocity, and g is the gap between the AV and the HDV. The desired gap g^* between the AV and the front HDV is formulated as [114]:

$$g^* = d_s + v_{\text{AV}} T_e - \frac{v_{\text{AV}} \Delta v}{2\sqrt{U_{\max} b}} \tag{4}$$

where T_e is the expected time gap, Δv is the velocity difference between the AV and the front vehicle (FV), and b is the comfortable deceleration.

The reward function of DDQN has been adapted to improve driving safety and efficiency in [115]. Specifically, a penalty is applied when the vehicle goes off-road or the time-to-collision (TTC) falls below a threshold [116]. The reward for driving efficiency is formulated as [115]:

$$R = \frac{v_0}{v_{\max}} \tag{5}$$

where $v_{\max}$ is the maximum velocity, and v_0 is the current velocity. This reward function helps maintain a relatively high driving velocity, thus increasing driving efficiency. Moreover, driving safety and altruism have been achieved using a level-k game-based DQN in [117]. The level-k game models the reasoning interaction between AVs and HDVs, promoting unselfish decision-making. A crash penalty is implemented in the DQN to prevent frequent collisions between AVs and HDVs. Additionally, unselfishness and training efficiency have been considered in [118]. Unselfishness is achieved through a cooperative multi-goal credit function-based policy gradient (PG). This adapted PG accounts for the goals of all vehicles, optimizing overall performance during training. Training efficiency is improved by a multi-agent reinforcement learning (MARL) curriculum, which reduces the number of trainable parameters and lowers computational costs.

Unselfishness and driving efficiency on highways are achieved in [119]. Unselfishness is promoted through MARL by considering each vehicle's state. Driving efficiency is enhanced by a reward function that selects actions to increase the average velocity of all vehicles. Driving safety and driving efficiency have been achieved using multi-objective approximate policy iteration (MO-API) in [120]. Driving safety is ensured by monitoring collisions, while driving efficiency has been assessed by comparing the v_0 with the v_e. In [121], driving efficiency and unselfishness are considered in highway driving. Driving efficiency is achieved by a reward based on v_0, $v_{\max}$, and $v_{\min}$. Unselfishness is achieved by penalizing unnecessary lane changes to reduce disturbances to HDVs. Driving safety and training efficiency have been addressed in [122]. Safety is maintained by ensuring a d_s between vehicles using rule-based constraints, while training efficiency is improved by incorporating a multi-head attention mechanism.

C. Three-Factor Methods for Highway Driving

Furthermore, three of the five considered factors are combined in a few recent papers. Driving safety, interpretability, and driving efficiency have been improved in [123]. Driving safety and interpretability are enhanced by using a collision penalty and the IDM. Driving efficiency is ensured by a reward based on the velocity difference between $v_{\max}$ and v_0. The reward at time step t is formulated as [123]:

$$R_t = -\text{Collision} - 0.1 \times (v_{\max}^t - v_0^t) - 0.4 \times (L-1)^2 \tag{6}$$

where Collision, $v_{\max}^t$, and v_0^t are the occurrence of a collision, maximum velocity, and current velocity at time t, respectively. L represents the relative position the target lane, where $L = 1$ indicates that the vehicle has successfully reached the target lane. A collision results in a negative reward, and a larger difference between v_0 and $v_{\max}$ also leads to a negative reward. Additionally, if the vehicle does not drive in the target lane, a penalty is applied.

A multi-reward-based DQN has been proposed to achieve safe, efficient, and unselfish driving in [124]. Three rewards are combined: speed reward, limited lane-changing reward, and overtaking reward. The speed reward is a normalized reward based on the current speed relative to the minimum and maximum speed limits [124]:

$$R_v = \frac{(v_0 - v_{\min}) * r_v}{v_{\max} - v_{\min}} \tag{7}$$

where R_v represents the reward for speed, encouraging higher speeds within safe limits. $v_{\min}$ is the minimum speed of the agent vehicle, and r_v is the base reward for speed. The limited lane-changing reward function is designed to minimize the number of lane changes, promoting safer driving and reducing the disturbance to surrounding vehicles:

$$R_l = -r_l \text{ if lane-change occurs, otherwise } 0 \tag{8}$$

where $-r_l$ is the penalty value for a lane change. The overtaking reward function encourages the agent vehicle to overtake more vehicles, improving driving efficiency:

$$R_o = r_o \text{ if an overtaking occurs, otherwise } 0 \tag{9}$$

where r_o is the reward value for overtaking.

Interpretability, driving safety, and driving efficiency have been achieved in [125]. Safety and efficiency are enhanced by penalizing frequent lane changes and tracking the desired velocity v_d, respectively. Interpretability is achieved through a car-following process using a proportional-derivative (PD) controller with transparent mathematical formulation [125]:

$$d_{\text{des},i} = \alpha_i^j v_l^{j+1} \tag{10}$$

$$a_{\text{cf},i} = K_p(x_l^{j+1} - x_i^j) + K_d(v_l^{j+1} - v_i^j) \tag{11}$$

where $d_{\text{des},i}$ is the desired following distance for the i-th vehicle, α_i^j is a sensitivity parameter with random values from $\mathcal{N}(1.3, 0.02)$, v_l^{j+1} is the speed of the leading vehicle in the $(j+1)$-th lane, $a_{\text{cf},i}$ is the acceleration command, K_p and K_d are the proportional and derivative gains, and x_l^{j+1} and x_i^j are the positions of the leading and i-th vehicles, respectively.

In [126], safety and efficiency are promoted by penalizing collisions and rewarding high average velocity, while interpretability is provided by a risk potential field that models and visualizes surrounding risk. In [127], adaptive cruise control achieves safety and interpretability by maintaining the headway d_s with explicit formulations, and efficiency by rewarding high speed states. In [128], safety uses a collision penalty, efficiency uses the gap between v_0 and $v_{\min}$, and training efficiency is improved with a long short term memory assisted DDQN. Compared with [123], which rewards efficiency via $v_{\max} - v_0$ and integrates safety and interpretability through collision penalties and lane deviation, and [125], where a PD controller offers clear mathematical transparency, the approach in [128] is distinctive for faster convergence but may warrant further study of robustness and stability under diverse traffic given the added complexity of the long short term memory component.

D. Four-Factor Methods for Highway Driving

Moreover, four of the five considered factors have been included in some studies. Driving safety, driving efficiency, training efficiency, and interpretability have been considered in [129]. Driving safety and driving efficiency are achieved by a reward function that maintains a d_s from the leading vehicle while tracking the v_d. Interpretability is ensured through safety-based driving rules [129]:

$$\begin{aligned} t_{f_{\min}} &= \inf\{t : t > 2(v - v_{f_{\text{target}}})/a_{d_{\max}}\}, \\ t_{b_{\min}} &= \inf\{t : t > 2(v_{b_{\text{target}}} - v)/a_{d_{\max}}\} \\ d_{\text{target}_{\min}} &= \min\{(v - v_f)t_f/2, (v_b - v)t_b/2\}, \\ \Delta d_{\text{target}} &= \min\{|x_{\text{AV}} - x_{f_{\text{target}}}|, |x_{\text{AV}} - x_{b_{\text{target}}}|\} \end{aligned} \tag{12}$$

where $t_{f_{\min}}$ and $t_{b_{\min}}$ are the minimum safe time intervals between the AV and the vehicles in front and behind in the target lane, respectively. v is the speed of the AV; $v_{f_{\text{target}}}$ and $v_{b_{\text{target}}}$ are the speeds of the front and behind vehicles in the target lane, respectively. $a_{d_{\max}}$ is the maximum deceleration. $d_{\text{target}_{\min}}$ is the minimum distance between the AV and the FV in the target lane, and Δd_{target} is the actual distance between the AV and the nearest vehicle in the target lane. x_{AV}, $x_{f_{\text{target}}}$, and $x_{b_{\text{target}}}$ represent the horizontal coordinates of the AV, the front target vehicle, and the vehicle behind in the target lane, respectively. By implementing these safety rules, the decision-making of the AV becomes more transparent and interpretable. Training efficiency is achieved through the potential-based reward shaping function. The total reward function and reward shaping function are given by [129]:

$$R' = R(s, a, s') + \beta F(s, a, s') \tag{13}$$

$$F(s, a, s') = \gamma\phi(s') - \phi(s) \tag{14}$$

$$F(s, a, s', t, t') = \gamma\phi(s', t') - \phi(s, t) \tag{15}$$

where R' is the new reward criterion, $R(s, a, s')$ is the original reward function, β is a weighting factor, $F(s, a, s')$ is the potential-based reward shaping function, s and s' are the current and next state, respectively, a is the action taken, and γ is the discount factor. $\phi(s)$ is the potential function mapping the state to a real number, and t and t' are the time corresponding to s and s', respectively. (16) combines the original reward function with an additional shaping term. (17) defines the shaping function as the difference between the discounted potential of the next state and the current state. (18) extends (17) by including time as a parameter and therefore allows for dynamic potential functions.

Highway studies addressing multiple DDTUI factors show several common patterns. [130] and [131] jointly consider safety, efficiency, unselfishness, and training efficiency: both embed safety and efficiency into the reward function, use MARL to promote cooperative behavior, and improve training efficiency through experience sharing (single-agent reuse in [130] and dynamic coordination graphs in [131]). Reference [132] similarly covers all four factors by penalizing collisions and road departures, rewarding overtaking, using MARL for cooperation, and employing parameter sharing to accelerate learning.

Some works incorporate interpretability alongside other factors. Reference [133] improves safety, efficiency, and unselfishness through reward shaping, while enhancing interpretability with an autonomous emergency braking module. Reference [134] combines safety, efficiency, interpretability, and training efficiency via a safety layer, speed-ratio features, an SVM for explainable boundaries, and spatial attention for faster learning.

Other works cover four factors without interpretability. Reference [135] uses collision penalties, speed ratios, and MARL to improve safety, efficiency, and cooperation, and increases training efficiency with a distributional DQN. Reference [136] enhances safety, efficiency, and unselfishness via reward design and improves interpretability through rule-based constraints on unsafe lane changes.

E. Five-Factor Methods for Highway Driving

Additionally, all the five factors are addressed in some studies, such as [137]. Driving safety and efficiency, and unselfishness are achieved by reducing collisions, increasing speed, and minimizing lane-change frequency through rewards. Training efficiency is improved through a convolutional neural network-based LSTM. Interpretability is enhanced by using spatio-temporal image representations for HDVs, which increase the interpretability of the inputs. The DRL-based decision making methods in highway driving based on DDTUI are summarized in Table I and categorized in Table II.

V. Deep Reinforcement Learning-Based Decision-Making in On-Ramping Merging

A. Single-Factor Methods for On-Ramping Merging

Driving efficiency has been considered using Q-learning in [138]. The remaining time of AV on the ramp lane is reduced by optimizing the reward function, thus promoting fast lane-changing to the main lane. The reward function is formulated as [138]:

$$r_t = \mu\bar{v}_t + \omega\bar{q}_t, \quad \mu > 0, \quad \omega < 0 \tag{16}$$

where r_t represents the reward after taking action a_t; $\bar{v}_t$ denotes the average speed in the merging area during time step t; $\bar{q}_t$ indicates the average queue length at the on-ramp during time steps t and $t+1$; μ is a positive weight assigned to the speed reward, and ω is a negative weight for the queue length reward. These rewards help balance the trade-off between enhancing vehicle mobility on the mainline and reducing delays at the on-ramp. Driving efficiency has also been improved in [139] by reducing the total travel time reward (R_{TTT}), represented by the summation of the total number of vehicles at each time step. Driving safety has been achieved through a safety factor in [140]. The safety factor is a negative reward when the relative distances between AV and HDV are small. Driving safety has been achieved in [141], by giving rewards for each state having a d_s and penalties for collisions.

B. Dual-Factor Methods for On-Ramping Merging

Driving efficiency and unselfishness have been considered in [142]. Driving efficiency is achieved by using the average velocity of AVs as part of the reward, and unselfishness is achieved using MARL to maximize general profits. Interpretability and driving efficiency have been addressed in [143]. Interpretability is achieved by using DDPG to tune a traditional controller's parameters, keeping the traditional controller as the main system to ensure transparency. Driving efficiency is enhanced by reducing the error state, which reflects the gap between actual and critical traffic density. A smaller error state leads to higher traffic flow and average speed.

C. Three-Factor Methods for On-Ramping Merging

Several merging studies address three DDTUI factors. Reference [144] improves efficiency with a travel-time reward and enhances training efficiency through a teacher–student framework, where traditional control supervises the DQN policy. Reference [145] improves efficiency via interval-based speed rewards and gains both unselfishness and interpretability by integrating ramp metering with Q-learning. Reference

TABLE I
EVALUATION OF THE DRL-BASED DECISION MAKING IN HIGHWAY DRIVING

Reference	Safety	Efficiency	Training Efficiency	Unselfishness	Interpretability
[103]	Safety modules	-	-	-	-
[104]	-	Overtaking reward	-	-	-
[105]	NMPC constraints	-	-	-	-
[109]	Hard constraints	-	-	-	-
[110]	-	-	-	Local interactions	-
[111]	-	-	-	-	Imitation learning
[112]	-	-	Attention module	-	-
[114]	IDM integration	-	-	-	IDM integration
[115]	TTC threshold	Velocity reward	-	-	-
[117]	Crash penalty	-	-	Level-k game	-
[118]	-	-	MARL curriculum	Cooperative function	-
[119]	-	Average velocity	-	MARL	-
[120]	Collision monitoring	Velocity comparison	-	-	-
[121]	-	Velocity reward	-	Lane change penalty	-
[122]	Rule-based	-	Attention mechanism	-	-
[123]	IDM & collision	Velocity difference	-	-	IDM integration
[124]	Speed-limit reward	Overtaking reward	-	Lane-change limit	-
[125]	Lane change penalty	Velocity tracking	-	-	PD controller
[126]	Reward function	Reward function	-	-	Risk potential field
[127]	Adaptive cruise	High-speed reward	-	-	ACC formulations
[128]	Collision penalty	Velocity difference	LSTM-DDQN	-	-
[129]	Safety rules	Reward function	Reward shaping	-	Safety rules
[130]	Reward function	Reward function	MARL reuse	Joint policy	-
[131]	Reaction time	Lane-changing point	DCG efficiency	MARL	-
[132]	Collision penalties	Overtaking reward	Parameter sharing	MARL	-
[133]	Collision rewards	Velocity ratio	-	Lane change limit	Emergency braking
[134]	Safety layer	Velocity ratio	Attention mechanism	-	SVM boundaries
[135]	Collision rewards	Velocity ratio	Distributional DQN	MARL	-
[136]	Collision rewards	Velocity difference	-	Lane change penalty	Rule-based
[137]	Collision reduction	Speed increase	CNN-LSTM	Lane change limit	Representations

'-' indicates that the corresponding factor was not explicitly addressed in the study.

TABLE II
CATEGORIZATION OF DRL BASED APPROACHES FOR HIGHWAY DRIVING

Category	**References & Analysis**
Network structure improved DRL	[112], [128], [135], [137]. Architectural changes aim to enhance spatio–temporal representation and sampling efficiency. Attention [112] and LSTM-based models [128] improve temporal reasoning, while distributional DQN [135] and CNN–LSTM hybrids [137] strengthen sample efficiency and interpretability.
Policy improved DRL	[104], [109]–[111], [115], [117]–[121], [123], [124], [130]–[132]. These works optimize policies through reward shaping, objective design, imitation learning, and MARL. Examples include overtaking and safety rewards [104], [115], imitation-guided policies [111], and cooperative strategies using level-k or multi-agent reasoning [117], [130].
Robust control assisted DRL	[103], [105]–[109], [125], [127], [129], [133], [134], [136]. Classical control principles (e.g., NMPC [105], PD control [125]) and rule-based or safety-layer mechanisms help enforce dynamic feasibility and safety. Some combine handcrafted safety rules with learned modules for robustness [103].
Traffic model assisted DRL	[113], [114], [123], [126]. Traffic models such as IDM support realistic car-following [113], [114], and risk potential fields assist in modeling surrounding dynamics for safer planning [126].

[146] targets safety, efficiency, and unselfishness: safety is encouraged through penalties for small headways, efficiency by maintaining appropriate gaps, and unselfishness through cooperative MARL.

D. Four-Factor Methods for On-Ramping Merging

Four-factor methods provide more comprehensive designs. In [147], DDPG-assisted ramp metering and speed limits improve efficiency and training efficiency, while the transparency of these controls supports unselfishness and interpretability. Reference [148] offers a balanced four-factor solution: an APF module improves safety and interpretability by quantifying and visualizing risk, and MPC–DDQN integration enhances both efficiency and training efficiency. Compared with this, [144] achieves efficiency and training efficiency but lacks risk visualization, while [145] improves efficiency and interpretability through RM but provides weaker safety mechanisms. Reference [146] strengthens safety and efficiency via headway management and promotes cooperation through MARL, though it provides neither MPC-level control nor interpretable risk modeling. Reference [149] addresses efficiency, training efficiency, and unselfishness by combining driver-intention modeling with DDPG. Finally, [150] achieves safety, efficiency, training efficiency, and interpretability via reward design and an IDM module, with independent PPO improving training speed.

TABLE III
EVALUATION OF THE DRL-BASED DECISION MAKING IN ON-RAMPING MERGING

Ref.	Safety	Efficiency	Training Efficiency	Unselfishness	Interpretability
[138]	-	Reward function	-	-	-
[139]	-	Travel time reward	-	-	-
[140]	Safety factor	-	-	-	-
[141]	Collision-free driving	-	-	-	-
[142]	-	Average velocity reward	-	MARL	-
[143]	-	Error state reduction	-	-	Traditional controller
[146]	Distance penalty	Distance minimization	-	MARL	-
[144]	-	Trip time difference	Teacher-student model	-	Traditional control
[145]	-	Speed comparison	-	Ramp metering	Ramp metering
[147]	-	DDPG-assisted RM	DDPG	RM and VSL	RM and VSL
[148]	APF	MPC with DDQN	MPC with DDQN	-	APF
[149]	Collision penalty	Stop maneuver penalty	DIM with DDPG	HDV intentions	-
[150]	Safety reward	Efficiency reward	IPPO	-	IDM
[151]	Crash evaluation	Stable speed assessment	Safety supervisor	MARL	Rule-based constraints
[152]	Collision rewards	Velocity ratio	Adversarial constraints	Nash-based game	transparent game process
[153]	DRAC	Velocity ratio reward	Multi-state rep.	Vehicle coop.	DRAC

TABLE IV
CATEGORIZATION OF DRL BASED APPROACHES FOR ON RAMP MERGING

Category	**References & Analysis**
Network structure improved DRL	No reviewed work explicitly modifies the network architecture for ramp merging.
Policy improved DRL	[138]–[142], [146], [152]. Policies are refined mainly through reward shaping and decision logic. Q-learning optimizes remaining ramp time [138] and travel time [139]. Safety is strengthened by penalizing small gaps [140], [141], and MARL encourages cooperative merging [142], [146], [152].
Robust control assisted DRL	[143]–[145], [147]–[149], [151], [153]. Classical control supports learning by improving stability and constraint handling. Examples include tuning controller parameters [143], teacher–student supervision [144], and integrating ramp metering, VSL, APF, or MPC to guide safe and efficient merging [145], [147]–[149], [151], [153].
Traffic model assisted DRL	[150]. Traffic flow models such as IDM are embedded to represent realistic ramp dynamics and improve safety and efficiency.

E. Five-Factor Methods for On-Ramping Merging

In [151], driving safety and efficiency have been achieved through collision and stable speed assessment rewards, respectively. Training efficiency is improved by using a safety supervisor, filtering detectable collision cases. Interpretability is enhanced through rule-based safety constraints, and unselfishness is achieved using MARL to maximize general profits. Similarly, all factors have been addressed in [152], where driving safety and efficiency are achieved via collision rewards and a velocity ratio, respectively. Training efficiency is enhanced by adversarial constraints, while unselfishness and interpretability is enhanced through a transparent Nash-based game that considers HDV's profits. Finally, in [153], driving safety and interpretability have been achieved using the deceleration rate to avoid a crash (DRAC), which has a detailed mathematical formulation and is transparent. Driving efficiency is improved by using (5) as an efficiency reward. Unselfishness is addressed by considering the cooperation intentions of other vehicles, and training efficiency is improved using multi-state representations to enhance the agent's learning capabilities. The DRL-based decision making methods in on-ramp merging based on DDTUI are summarized in Table III and categorized in Table IV.

VI. DEEP REINFORCEMENT LEARNING-BASED DECISION-MAKING AT ROUNDABOUTS

A. Single-Factor Methods for Roundabout Driving

Driving efficiency in roundabout driving has been improved using soft actor-critic (SAC) with higher peak rewards in [154]. Training efficiency has been achieved through action repeat and asynchronous advantage in [155]. Action repeat improves efficiency by allowing the agent to repeat the same action for several time steps, decreasing the frequency of making new decisions. Asynchronous advantage enables each agent to share its interaction experience with others. Training efficiency has been further improved by embedding the operational design domain (ODD) into DQN in [156]. ODD guides the training to more targeted scenarios, reducing unnecessary exploration and accelerating convergence.

B. Dual-Factor Methods for Roundabout Driving

Dual-factor methods typically target efficiency, safety, training efficiency, or interpretability. Reference [157] improves efficiency and training efficiency via the Conditional Representation Model, which classifies states as safe or unsafe to guide learning. Reference [158] enhances training efficiency and interpretability by using expert-labeled data for supervised guidance. Reference [159] improves safety and efficiency using rewards based on desired speed and allowable relative distance.

Training efficiency and interpretability are jointly improved in [160] through optimization-embedded DRL, which provides transparent model-based updates. Reference [161] improves safety and unselfishness using collision and off-road penalties with MARL to promote cooperative behavior. Reference [162] achieves safety and interpretability via collision penalties and a gradual training curriculum that progresses from sparse to dense traffic.

C. Three-Factor Methods for Roundabout Driving

Three-factor approaches mainly combine safety, efficiency, and training efficiency. Reference [163] achieves this using dedicated safety and efficiency rewards and accelerates convergence through trust region policy optimization (TPRO). Reference [164] enhances safety and efficiency through non-collision lane-change and velocity-based rewards, while training efficiency improves through LSTM-enhanced actor–critic networks. Reference [165] strengthens training efficiency via reward normalization and boosts safety and efficiency using multi-environment training, which increases success rates and reduces crashes.

D. Four-Factor Methods for Roundabout Driving

Four-factor methods provide the most comprehensive coverage.Reference [166] addresses safety using d_s, efficiency using a velocity-ratio reward, training efficiency through synthetic representations, and unselfishness through MARL.Reference [56] covers safety, efficiency, interpretability, and training efficiency using crash penalties, high-speed rewards, IDM for transparent car-following, and an interval-prediction module to reduce training cost.Reference [167] integrates safety, efficiency, training efficiency, and interpretability through collision and stop penalties, hybrid DDPG–DQN–NMPC training, and NMPC-based transparent decision mechanisms.

Relative to these, [167] offers the strongest overall balance: it provides interpretable control via NMPC, fast convergence through hybrid learning, and robust safety–efficiency trade-offs. Three-factor methods such as [163], [164], and [165] achieve competitive performance but lack the explicit interpretability provided by NMPC in [167]. Similarly, while [56] and [166] cover four factors, their interpretability and training-efficiency mechanisms are less cohesive than those in [167]. Overall, [167] stands as a strong baseline for roundabout driving due to its balanced treatment of all four dimensions.

E. Five-Factor Methods for Roundabout Driving

All five factors have been considered in [107], where driving safety and interpretability are ensured by a rule-based action inspector. Driving efficiency is enhanced via high-speed rewards. Training efficiency is achieved through a Kolmogorov-Arnold network-enhanced DQN. Unselfishness is promoted through rule-based route planning that considers the varying distributions of HDVs on the roundabout. The DRL-based decision making methods in roundabouts based on DDTUI are summarized in Table V and categorized in Table VI.

VII. Deep Reinforcement Learning-Based Decision-Making at Unsignalized Intersections

A. Single-Factor Methods for Intersection Driving

Traffic efficiency has been improved by using the difference between v_0 and v_d as a reward in [168]. Additionally, a penalty is applied when the velocity drops below a threshold, further boosting traffic efficiency. In [169], driving efficiency has been achieved by applying a constant penalty as long as the AV has not reached the target exits.

B. Dual-Factor Methods for Unsignalized Intersection Driving

Dual-factor methods for intersections focus mainly on combinations of efficiency, interpretability, training efficiency, and unselfishness.

Driving and training efficiency are improved in [170] by using total waiting time (TWT) in the reward and accelerating learning through a background-removal ResNet. Several works combine efficiency with interpretability: [171] uses a velocity-difference reward for efficiency and IDM for transparent car-following; [172] enhances efficiency with a velocity-based reward and interpretability through rule-based safety policies; [173] improves efficiency using safe-distance rewards and gains interpretability via a model-based TD3 structure; [174] promotes efficiency with a speed-ratio reward and interpretability through gridded coordination-zone representations.

Driving and training efficiency are also addressed in [175] using DQN with shared and task-specific submodules to enable knowledge transfer across tasks. Training efficiency and unselfishness are improved in [176] through incentive communication-assisted MARL, where agents exchange custom messages to coordinate and maximize collective benefit. Reference [177] enhances efficiency and training efficiency using an adaptive D2-TSP module with DDQN to optimize bus speeds and reduce passenger waiting time. Similarly, [178] improves both factors through cooperative intersection management combined with DQN and vehicle connectivity.

TABLE V
EVALUATION OF THE DRL-BASED DECISION MAKING AT ROUNDABOUTS

Ref.	Safety	Efficiency	Training Efficiency	Unselfishness	Interpretability
[154]	-	SAC with higher peak rewards	-	-	-
[155]	-	-	Action repeat, asynchronous advantage	-	-
[156]	-	-	ODD-embedded DQN	-	-
[157]	-	CRM	CRM	-	-
[158]	-	-	Expert guidance	-	Expert guidance
[159]	Allowable relative distance	v_d	-	-	-
[160]	-	-	Optimization-embedded DRL	-	Model-based optimization
[161]	Collision penalties	-	-	MARL	-
[162]	Collision penalties	-	-	-	Gradual training
[163]	Safety rewards	Efficiency rewards	TPRO	-	-
[164]	Non-collision rewards	Velocity difference rewards	LSTM-embedded actor-critic	-	-
[165]	Fewer crashes	Higher success rates	Reward normalization	-	-
[166]	Safety distance	Velocity ratio	Synthetic representation	MARL	-
[56]	Crash penalties	High-speed rewards	Interval prediction	-	IDM
[167]	Collision penalties	Vehicle-stop penalties	DDPG, DQN, NMPC integration	-	NMPC
[107]	Rule-based inspector	High-speed rewards	KAN-DQN	Rule-based planning	Rule-based inspector

TABLE VI
CATEGORIZATION OF DRL BASED APPROACHES FOR ROUNDABOUT DRIVING. SOME REFERENCES CONTRIBUTE TO MULTIPLE CATEGORIES

Category	**References & Analysis**
Network structure improved DRL	[107], [156], [157], [164]. Architectural changes enrich features or guide training: embedding the operational design domain into DQN to focus on targeted scenarios [156]; a conditional representation model for state characterization [157]; adding LSTM to the actor critic to capture temporal dynamics [164]; a Kolmogorov Arnold network enhanced DQN for improved feature mapping [107].
Policy improved DRL	[154], [155], [158]–[163], [165]. Studies refine the learned policy via reward design and training strategies: higher peak rewards with SAC to promote efficiency [154]; action repeat and asynchronous advantage style updates to boost training efficiency [155]; expert guided labeling and imitation [158]; safe driving criteria embedded in rewards [159]; optimization embedded DRL to tailor behavior for roundabouts [160]–[163], [165].
Robust control assisted DRL	[166], [167]. Conventional control is combined with learning to enforce constraints and stabilize training: DRL with NMPC and related controllers for safe decisions under physical limits [167]; rule based action inspectors and synthetic representations to improve robustness [166].
Traffic model assisted DRL	[56]. Established traffic models improve safety and interpretability; IDM guides car following aligned with realistic dynamics.

C. Three-Factor Methods for Unsignalized Intersection Driving

In [179], driving efficiency, unselfishness, and training efficiency have been addressed. Driving efficiency is enhanced by penalizing each low-speed state, while unselfishness is achieved through MARL for maximizing overall profits. Training efficiency is improved with multi-agent DQN, which offers faster convergence than baseline algorithms. In [180], driving efficiency, training efficiency, and interpretability have been integrated. Driving efficiency is increased by rewarding each high-velocity state, and training efficiency is achieved by combining deep Q-learning with transfer learning. Interpretability is improved by using the IDM for safe vehicle following.

In [181], driving safety, driving efficiency, and training efficiency have been incorporated. Driving safety is promoted through collision penalties, and driving efficiency is enhanced by rewarding velocities higher than a baseline. Training efficiency is improved by using a spatial and temporal attention module with SAC. In [182], all three aspects have also been addressed. Driving safety and efficiency are improved by rewarding goal attainment and penalizing collisions. Training efficiency is enhanced using a randomized prior function (RPF) for each ensemble member, leading to a better Bayesian posterior [185]. Compared with this approach, [179] emphasizes driving efficiency by penalizing low speed states and promotes unselfishness with multi agent reinforcement learning, while training efficiency is improved via multi agent DQN. However, its safety mechanism is less explicit and does not adopt the goal oriented reward strategy in [182]. Likewise, [180] raises efficiency by rewarding high velocity states and speeds training by combining Q learning with transfer learning; interpretability is provided by IDM based following, but it lacks the robust uncertainty handling enabled by the randomized prior function in [182]. Furthermore, [181] improves safety and efficiency through collision penalties and velocity based rewards and uses a spatio temporal attention module with SAC to enhance training efficiency, yet it focuses on standard scenarios and does not explicitly address edge cases where RL may exploit reward loopholes. Overall, while each method has strengths, [182] offers a more balanced integration of safety, efficiency, and training efficiency, largely due to its use of a randomized prior function for uncertainty.

D. Four-Factor Methods for Unsignalized Intersection Driving

In [33], driving safety, driving efficiency, training efficiency, and unselfishness have been incorporated. Driving safety is enhanced through autonomous intersection management (AIM), and driving efficiency is improved by applying a constant penalty until the AV reaches the exits. Training efficiency is improved by embedding AIM and LSTM into the learning, and unselfishness is achieved through MARL. In [183], driving safety, driving efficiency, training efficiency, and interpretability have been integrated. Driving safety is promoted through collision penalties, and driving efficiency is enhanced by rewarding goal attainment. Training efficiency is improved through the Mix-Attention Network, synthetic representation mechanism, and replay memory mechanism. The interpretability is ensured by using the IDM.

E. Five-Factor Methods for Intersection Driving

In [184], driving safety has been promoted through collision penalties, while driving efficiency is enhanced by penalizing each state with velocity lower than the $v_{\min}$. Training efficiency is improved using value decomposition-based multi-agent deep Q-learning. Unselfishness is achieved by employing MARL to minimize joint profits, and interpretability is ensured through the IDM. The DRL-based decision making on unsignalized intersections based on DDTUI is summarized in Table VII, while these methods are categorized in Table VIII.

VIII. Scenario-Centric Learning and Transferability Analysis

This section provides a systematic evaluation of scenario-centric learning improvements and cross-scenario transferability for DRL-based autonomous driving methods across the four investigated scenarios. The SCLT framework distinguishes genuine scene-specific learning advances from domain-specific heuristics by analyzing framework-level evidence rather than relying on heterogeneous evaluation metrics.

The SCLT evaluation methodology employs a comprehensive multi-dimensional assessment framework. For scenario-specific learning improvements, each method is evaluated across three critical dimensions: (i) Tailored State Representations: whether the approach incorporates scenario-specific environmental features beyond generic vehicle dynamics, such as merging gap analysis, circular trajectory encoding, or intersection conflict mapping; (ii) Specialized Learning Objectives: whether the algorithm extends beyond standard reward tuning to include scenario-focused mechanisms, such as yield-priority protocols, gap acceptance strategies, or multi-exit navigation optimization. For cross-scenario transferability assessment, the analysis examines structural dependencies across multiple levels: topology dependencies, such as adaptation to different road geometries from linear highways to circular roundabouts; rule dependencies, such as algorithmic reliance on scenario-specific traffic rules versus universal driving principles; and dynamic dependencies, such as adaptability to varying interaction patterns from high-speed merging to low-speed intersection negotiations. Each reviewed method receives a systematic scoring across these dimensions, enabling quantitative comparison of learning advancement levels and transferability potential. The framework further groups methods into learning tiers and transferability classes based on their cumulative scores. Tables IX–XII summarise the SCLT assessments for highways, on-ramps, roundabouts, and unsignalised intersections.

A. Highway Scenario Learning Analysis

Highway scenarios offer structured multi-lane settings with longitudinal and lane-change decisions. The SCLT analysis shows differing levels of scenario-specific learning gains across methods.

1) Scenario-Specific Learning Assessment: Several highway methods demonstrate genuine scenario-specific learning adaptations. The spatio–temporal image representations in [137] provide tailored state representations for multi-lane highways by encoding lane structure and vehicle formation patterns. Its CNN–LSTM architecture introduces specialized learning objectives that capture the temporal dynamics of highway interactions.

The IDM integration approaches [114], [123] embed calibrated car-following models for high-speed longitudinal control, tuning parameters such as time gaps and deceleration to highway conditions. However, they offer little bespoke state representation beyond standard vehicle-dynamics features. The safety-based driving rules in [129] provide both tailored states and specialised objectives: minimum time gaps and lane-

TABLE VII
EVALUATION OF THE DRL-BASED DECISION MAKING AT UNSIGNALIZED INTERSECTIONS

Ref.	Safety	Efficiency	Training Efficiency	Unselfishness	Interpretability
[168]	-	Velocity difference reward	-	-	-
[169]	-	Time penalty	-	-	-
[170]	-	Total waiting time	Background removal ResNet	-	-
[171]	-	Velocity difference reward	-	-	IDM
[172]	-	Velocity-based reward	-	-	Safety-based rule policy
[173]	-	Safe distance reward	-	-	MPC with TD3
[174]	-	Velocity ratio reward	-	-	Gridded coordination zone
[175]	-	Goal attainment reward	DQN with subtasks	-	-
[176]	-	-	Incentive communication	MARL	-
[177]	-	D2-TSP	DDQN	-	-
[178]	-	CIM-enhanced DQN	CIM-enhanced DQN	-	-
[179]	-	Low-speed penalty	Multi-agent DQN	MARL	-
[180]	-	High-velocity reward	DQL with transfer learning	-	IDM
[181]	Collision penalties	High-velocity reward	SAC with attention	-	-
[182]	Collision penalties	Goal attainment reward	RPF	-	-
[33]	AIM	Constant time penalty	AIM and LSTM	MARL	-
[183]	Collision penalties	Goal attainment reward	Mix-Attention Network	-	IDM
[184]	Collision penalties	Low-velocity penalty	VD-MADQL	MARL	IDM

TABLE VIII
CATEGORIZATION OF DRL BASED APPROACHES FOR UNSIGNALIZED INTERSECTION DRIVING

Category	**References & Analysis**
Network structure improved DRL	[170], [175], [181]–[183]. Architectural enhancements aim to improve convergence and feature extraction, including background-removal ResNet [170], shared and task-specific DQN branches [175], spatial–temporal attention with SAC [181], ensemble priors [182], and mix-attention networks [183].
Policy improved DRL	[168], [169], [171]–[174], [177]–[180], [184]. These works refine reward shaping and decision policies: differential velocity rewards for efficiency [168], [169], waiting-time or distance-based incentives for coordination [171]–[174], [177], [178], and transfer learning or MARL to accelerate convergence and encourage cooperation [179], [180], [184].
Robust control assisted DRL	[33], [183]. Classical control is integrated to enhance safety and stability, for example through AIM-style scheduling or controller-based decision refinement.
Traffic model assisted DRL	[171], [172], [174], [180], [183], [184]. Traffic models such as IDM and rule-based safety policies guide DRL toward realistic following and coordinated movement at intersections.

distance metrics encode highway-specific spatial relations, while potential-based rewards target typical highway lane changes. Some methods show partial scenario-specific adaptation. The overtaking-oriented training in [104] introduces specialized learning objectives through rewards that prioritize position advancement, a key metric for highway efficiency, but lacks tailored state representations.

Most methods, including [103], [115], [117], [118], show limited scenario-specific learning gains, relying on generic rewards and standard multi-agent mechanisms without modelling highway-specific structures or interactions.

2) Cross-Scenario Transferability Assessment: The transferability analysis reveals substantial variation across method categories. Network structure improvements, including atten-

TABLE IX
SCLT ASSESSMENT FOR HIGHWAY SCENARIO DRL METHODS

Reference	Tailored State Rep.	Specialized Objectives	Learning Level	Transferability	Primary Adaptation
[137]	High	High	Comprehensive	High	Spatio-temporal encoding
[129]	High	High	Comprehensive	Medium-High	Safety rule formulation
[114]	Medium	High	Partial	Medium	IDM integration
[123]	Medium	High	Partial	Medium	IDM + reward design
[104]	Low	High	Partial	Medium	Overtaking-oriented rewards
[105]	Low	Medium	Partial	Medium-High	NMPC constraints
[125]	Low	Medium	Partial	Medium-High	PD controller integration
[126]	Low	Medium	Partial	Medium	Risk potential field
[112]	Low	Low	Minimal	High	Attention mechanism
[128]	Low	Low	Minimal	High	LSTM architecture
[134]	Low	Low	Minimal	High	SVM boundaries
[103]	Low	Low	Minimal	Medium	Safety modules
[115]	Low	Low	Minimal	Medium	Generic rewards
[117]	Low	Low	Minimal	Medium	Level-k game
[118]	Low	Low	Minimal	Medium	MARL framework

Learning Level: Comprehensive (both dimensions high), Partial (one dimension high), Minimal (both dimensions low).
Transferability: High (direct application), Medium-High (minor adaptation), Medium (moderate adaptation), Low (major adaptation).

TABLE X
SCLT ASSESSMENT FOR ON-RAMPING MERGING SCENARIO DRL METHODS

Reference	Tailored State Rep.	Specialized Objectives	Learning Level	Transferability	Primary Adaptation
[148]	High	High	Comprehensive	Medium-High	APF risk visualization
[144]	Medium	High	Partial	High	Teacher-student framework
[149]	Low	High	Partial	Medium	Driver intention modeling
[145]	Medium	High	Partial	Medium	Ramp metering integration
[147]	Medium	High	Partial	Medium	RM and VSL control
[153]	Medium	High	Partial	Medium	DRAC safety assessment
[150]	Low	Medium	Partial	Medium	IDM integration
[143]	Low	Medium	Partial	Medium-High	Traditional controller tuning
[138]	Low	High	Partial	Medium	Queue length optimization
[146]	Low	Medium	Partial	High	MARL cooperation
[142]	Low	Low	Minimal	High	Generic MARL
[151]	Low	Low	Minimal	High	Rule-based constraints
[152]	Low	Low	Minimal	Medium	Nash-based game
[139]	Low	Low	Minimal	Medium	Travel time optimization
[140]	Low	Low	Minimal	Medium	Generic safety factor
[141]	Low	Low	Minimal	Medium	Generic collision avoidance

TABLE XI
SCLT ASSESSMENT FOR ROUNDABOUT SCENARIO DRL METHODS

Reference	Tailored State Rep.	Specialized Objectives	Learning Level	Transferability	Primary Adaptation
[107]	High	High	Comprehensive	Medium	Rule-based action inspector
[156]	Medium	High	Partial	High	ODD embedding
[167]	Low	High	Partial	High	NMPC integration
[56]	Medium	High	Partial	Medium	IDM and path prediction
[157]	Low	Medium	Partial	Medium	Conditional representation
[160]	Low	Medium	Partial	High	Optimization-embedded DRL
[166]	Low	Medium	Partial	High	Synthetic representation
[163]	Low	Low	Minimal	High	Trust region optimization
[164]	Low	Low	Minimal	High	LSTM actor-critic
[165]	Low	Low	Minimal	High	Multi-environment training
[155]	Low	Low	Minimal	High	Action repeat mechanism
[154]	Low	Low	Minimal	High	Peak reward optimization
[159]	Low	Low	Minimal	Medium	Generic reward design
[161]	Low	Low	Minimal	High	Generic MARL
[158]	Low	Low	Minimal	High	Expert guidance
[162]	Low	Low	Minimal	Medium	Gradual training mode

tion mechanisms [112], [134] and LSTM-based architectures [128], [137], show high transferability because spatial–temporal attention and sequence modeling are broadly applicable across autonomous driving scenarios.

TABLE XII
SCLT ASSESSMENT FOR UNSIGNALIZED INTERSECTION SCENARIO DRL METHODS

Reference	Tailored State Rep.	Specialized Objectives	Learning Level	Transferability	Primary Adaptation
[174]	High	High	Comprehensive	Medium	Gridding coordination zone
[33]	Low	High	Partial	Medium	AIM integration
[177]	Low	High	Partial	Medium	D2-TSP algorithm
[183]	Medium	High	Partial	High	Mix-Attention Network
[176]	Low	High	Partial	High	Communication-assisted MARL
[181]	Low	Medium	Partial	High	Spatial-temporal attention
[182]	Low	Medium	Partial	High	Randomized prior function
[178]	Low	Medium	Partial	High	Cooperative intersection mgmt
[175]	Low	Medium	Partial	High	Sub-task knowledge sharing
[179]	Low	Low	Minimal	High	Generic MARL
[180]	Low	Low	Minimal	High	Transfer learning
[184]	Low	Low	Minimal	High	Value decomposition
[173]	Low	Low	Minimal	Medium	Model-based TD3
[171]	Low	Low	Minimal	Medium	IDM integration
[172]	Low	Low	Minimal	Medium	Rule-based policy
[170]	Low	Low	Minimal	High	Background removal ResNet
[168]	Low	Low	Minimal	Medium	Generic reward design
[169]	Low	Low	Minimal	Medium	Constant penalty

Policy-level improvements exhibit medium transferability. Reward shaping and multi-objective formulations can be adapted with parameter adjustments, but highway-specific objectives—such as overtaking rewards [104] or high-speed efficiency metrics—require extensive modification for low-speed or multi-directional settings like intersections.

Control-integration methods demonstrate medium–high transferability. NMPC integration [105] and safety-constraint frameworks [129], [136] provide generalizable mathematical structures, although the associated safety bounds and constraint parameters must be recalibrated for each scenario's geometry and operating speed.

Traffic-model integration approaches show medium transferability. IDM-based methods [114], [123] generalize moderately well because car-following principles apply across scenarios, but substantial parameter adaptation is required for environments with different interaction patterns, such as roundabouts or intersections.

B. On-Ramping Merging Scenario Learning Analysis

On-ramp merging poses challenges involving acceleration, gap acceptance, and spatial constraints. The SCLT analysis shows distinct patterns of learning specialisation across the reviewed methods.

1) Scenario-Specific Learning Assessment: Several methods show clear merging-specific adaptations. The APF–DDQN framework in [148] offers tailored states and specialised objectives: the APF models risk regions for gap assessment and collision avoidance, while the MPC–DDQN setup adds temporal decision-making suited to merging.

The teacher–student approach in [144] also reflects meaningful scenario-specific adaptation. It incorporates traditional merging control strategies as supervisory signals, and its tailored state representations embed domain knowledge related to acceleration, gap acceptance, and merging feasibility. The DIM-based method in [149] offers specialized learning objectives by modeling HDV intentions to support cooperative gap acceptance, a key requirement in merging contexts. However, its state representation remains largely generic beyond standard multi-agent encodings. The RM-based approaches in [145] and [147] show partial scenario-specific adaptation. Their learning objectives incorporate traffic-flow optimization mechanisms designed for ramp environments, addressing spatial constraints and ramp-specific flow patterns, although their state representations remain limited.

Some methods provide only minimal scenario-specific improvements. The queue-length optimization in [138] adds merging-relevant reward terms such as queue length and merging-zone speed but lacks tailored state representations. Most remaining approaches, including [139], [140], [141], [142], rely primarily on generic reward shaping and standard safety rules without explicitly modeling merging-specific environmental features or learning mechanisms.

2) Cross-Scenario Transferability Assessment: Transferability varies across merging-specific methods. The APF-based risk visualization approach [148] has medium–high transferability: its risk visualization can extend to intersections and roundabouts with geometric adjustments, though spatial constraints need reconfiguration. The teacher–student learning framework [144] shows high transferability, as supervisory controllers can be swapped for diverse scenarios. Traffic flow optimization through RM and VSL [145], [147] has medium transferability; the concepts generalize, but models and parameters require substantial revision outside infrastructure-regulated settings. Cooperative learning mechanisms in MARL [142], [146], [151] demonstrate high transferability, with multi-agent coordination naturally extending to intersections and roundabouts. The DRAC-based safety assessment [153] shows medium transferability: its collision-risk formulation adapts to other scenarios, but thresholds must be recalibrated for different speeds and interactions.

C. Roundabout Scenario Learning Analysis

Roundabout scenarios demand specialized learning due to circular topology, multi-exit navigation, and complex yielding. SCLT analysis shows limited but notable scenario-specific learning across methods.

1) Scenario-Specific Learning Assessment: The roundabout literature contains fewer fully developed scenario-specific adaptations compared with highway and merging scenarios. The rule-based action inspector in [107] provides both tailored state representations and specialized learning objectives. Its rule-based route planning explicitly handles multi-exit navigation and varying HDV distributions on the roundabout, representing a strong scenario-specific learning adaptation.

The ODD embedding method in [156] demonstrates specialized learning objectives by incorporating roundabout-relevant operational constraints into a DQN framework. It also offers tailored state representations by integrating safety and operational boundaries specific to circular driving environments.

The NMPC-based approaches in [167] introduce specialized objectives through predictive control tailored for roundabout navigation. The NMPC module addresses circular-trajectory planning and multi-exit decisions, though its state representation remains largely limited to standard vehicle-dynamics features adapted to circular paths.

The IDM integration in [56] provides specialized objectives by adapting car-following dynamics to roundabout-specific behavior. Its interval prediction model for feasible path precomputation serves as a tailored state representation that encodes geometric constraints and multi-exit structure.

Several methods show partial scenario-specific adaptation. The CRM approach in [157] includes specialized learning objectives through roundabout-specific safe–unsafe state classification. However, its state representation remains largely generic, relying on general safety-encoding mechanisms rather than explicitly modeling roundabout geometry or conflict patterns.

The majority of reviewed methods, including [154], [155], [159], [160], show minimal scenario-specific learning, relying on generic rewards, standard training techniques, and conventional multi-agent frameworks without roundabout-specific adaptations.

2) Cross-Scenario Transferability Assessment: Transferability varies among roundabout-specific methods. The rule-based action inspector approach [107] has medium transferability: its route-planning concepts extend to intersections, but circular-geometry assumptions need major adaptation for highways or merging scenarios. The NMPC integration [167] shows high transferability, as its predictive control framework can be adapted to other scenarios by adjusting models and constraints, making it broadly applicable across autonomous driving contexts. The ODD embedding approach [156] is highly transferable, as incorporating operational design domains can be applied to any scenario by specifying scenario-specific boundaries and constraints. The IDM-based approaches [56] demonstrate medium transferability. While car-following principles apply broadly, the interval prediction mechanisms and path precomputation strategies require scenario-specific adaptations for non-circular environments. The MARL coordination mechanisms [161], [166] has high transferability, as multi-agent cooperation principless for roundabout scenarios can be effectively transferred to intersection and merging scenarios. The training efficiency improvements through action repeat [155] and trust region optimization [163] are also highly transferable, as these training acceleration techniques are scenario-agnostic and broadly applicable.

D. Unsignalized Intersection Scenario Learning Analysis

Unsignalized intersections are highly complex, demanding advanced coordination and conflict resolution strategies. The SCLT analysis reveals limited scenario-specific learning developments despite this complexity.

1) Scenario-Specific Learning Assessment: The intersection literature provides relatively few fully developed scenario-specific learning adaptations. The AIM-based method [33] targets multi-directional coordination and conflict resolution but uses standard multi-agent state encoding with limited scenario-specific representations. The gridding-based coordination zone method [174] shows both tailored state representations and specialized objectives. Its spatial discretization and matrix-based risk encoding capture intersection-specific spatial relations and conflict patterns, representing a strong scenario-focused adaptation. The adaptive D2-TSP integration [177] provides specialized objectives by modeling intersection-specific temporal optimization but uses only generic traffic-flow state representations. The Mix-Attention Network [183] introduces specialized objectives through attention mechanisms suited to multi-directional interactions. While its synthetic representation enhances feature extraction, it still lacks highly tailored state definitions. Several methods exhibit partial scenario-specific adaptation. The incentive communication-assisted MARL [176] incorporates specialized objectives through custom message creation to support intersection negotiation, yet its state representation remains based on general multi-agent communication formats.

Most methods, including [168], [169], [170], [171], [172], show minimal scenario-specific learning, relying on generic rewards, standard velocity metrics, and conventional IDM integration without intersection-specific features or specialized learning.

2) Cross-Scenario Transferability Assessment: Transferability varies across intersection-oriented methods. The AIM-based approach [33] exhibits medium transferability: although its coordination principles can extend to roundabouts, the multi-directional conflict resolution mechanisms require substantial redesign for linear highway and ramp environments. The gridding-based coordination zone method [174] also shows medium transferability. Its spatial discretization can be adapted to other scenarios with geometric adjustments, but the conflict-zone formulations remain highly scenario dependent outside intersection settings. Communication-assisted MARL [176] demonstrates high transferability. Its custom messaging and negotiation strategies apply naturally to merging, roundabout circulation, and other multi-agent interactions, making the framework broadly applicable across driving scenarios. Attention-based learning approaches [181], [183] similarly show high transferability. Spatial and temporal attention modules can be refocused to different risk regions, enabling adaptation to highways, ramps, and roundabouts. IDM-integrated methods [171], [180], [183], [184] exhibit medium transferability. While car-following dynamics are

TABLE XIII
OCCURRENCE AND RATIO OF EVALUATION FACTORS ACROSS DIFFERENT SCENARIOS

Scenario	Safety	Eff.	Train Eff.	Unselfish.	Interpret.
Highway	23 (77%)	20 (67%)	11 (37%)	13 (43%)	11 (37%)
Ramp	9 (56%)	14 (88%)	8 (50%)	8 (50%)	9 (56%)
Roundabout	10 (63%)	11 (69%)	12 (75%)	3 (19%)	6 (38%)
Intersection	5 (28%)	17 (94%)	12 (67%)	4 (22%)	6 (33%)
Total	47 (58%)	63 (78%)	44 (54%)	29 (36%)	32 (40%)

Percentages are relative to total studies per factor.

widely applicable, scenario-specific calibration of safety distances and speed profiles is necessary to accommodate varying interaction patterns and operational speeds. The training efficiency improvements through value decomposition [184] and transfer learning [180] demonstrate high transferability, as these training acceleration techniques are scenario-agnostic and can be applied universally across autonomous driving contexts.

IX. Discussion

Table XIII shows the distribution of the five evaluation factors across four scenarios. At unsignalized intersections, 18 studies (94.4%) prioritize efficiency to reduce delay in interaction-dense settings. Ramps show a similar trend, with 14 studies (87.5%) emphasizing efficiency to ease congestion. On highways, safety dominates in 23 studies (76.7%) due to high-speed operation, whereas intersections incorporate safety less frequently. Training efficiency is most prominent at roundabouts (12 studies, 75%) and unsignalized intersections (12 studies, 66.7%), reflecting the need for sample-efficient learning in complex environments. Interpretability appears primarily at ramps (9 studies, 56.25%) and highways (11 studies, 36.7%). Unselfishness is overall less common but receives relatively more attention in highway and ramp settings.

DRL is promising across all scenarios, but deployment depends on hardware, regulations, and public acceptance, each shaped by scenario demands. Highways require strong onboard computing and policies enforcing speed, lane keeping, and headway, backed by safety evidence. Ramp merging demands computational resources for precise gap acceptance and strict rule compliance, validated transparently. Roundabouts need heavier processing for complex trajectories and multi-agent interactions, with adaptable controllers and clear decision explanations. Unsignalized intersections are most challenging, requiring policies that encode local norms, operate on reliable hardware for timely responses, and are supported by extensive testing and open reporting to ensure safety.

SCLT analysis highlights a persistent gap between scenario complexity and scene-specific learning in current DRL research. Only a few methods achieve comprehensive tailoring: 2/25 for highways, 1/16 for on-ramp merging, 1/16 for roundabouts, and 1/18 for unsignalized intersections. Transferability varies: network-structure improvements are broadly portable with minimal changes; policy tweaks transfer moderately but require retuning; control integrations transfer medium–high with scenario-specific calibration; and traffic-model integrations are highly variable, depending on assumptions. Overall, research adapts best to structured highways, while intersections and roundabouts—requiring richer multi-agent coordination and spatial reasoning—remain underexplored. Most methods fall into minimal learning, relying on generic rewards and standard MARL without explicit scene structure, leaving room for deeper, scenario-specific learning.

X. Conclusion and Future Works

This survey provides a comprehensive overview of DRL-based decision-making for autonomous vehicles across highways, on-ramp merging, roundabouts, and unsignalized intersections. Beyond summarizing existing algorithms, we highlight common research patterns, gaps, and emerging trends through the lens of the five DDTUI factors. While safety and efficiency remain widely addressed, recent work increasingly integrates multiple factors within unified frameworks. Approaches such as MARL and DRL–control integration show strong potential to enhance unselfishness, interpretability, and robustness in complex driving environments. Future challenges and opportunities are summarized below.

1) **Balancing all five DDTUI factors in a single framework remains rare.**Only 3 of 16 roundabout studies and 1 of 19 intersection studies achieve full coverage. This underscores the difficulty of designing unified frameworks that maintain safety, efficiency, unselfishness, interpretability, and training efficiency together. Future work should develop architectures that jointly optimize all five factors.
2) **Improving interpretability without sacrificing performance remains limited.**Fewer than 40% of reviewed studies address interpretability, typically with a single mechanism. While models like IDM provide interpretable baselines for car-following, many high-performing DRL policies remain non-transparent models. Future work should combine multiple interpretability tools, such as APF and IDM, to enhance transparency without reducing performance.
3) **Developing richer MARL grounded in human behavior remains a key challenge.**About 50% of studies use MARL to promote unselfish behavior, but real-world traffic involves high uncertainty. Future work should incorporate naturalistic driving data, combining game-theoretic reasoning with learned driving-style classification. Large language models (LLMs) can provide contextual understanding and explicit reasoning though output may be inconsistent. A promising approach is a hybrid pipeline where LLMs identify constraints and DRL performs precise policy optimization within them.
4) **Advancing scenario-centric learning with stronger transferability is needed.**Only 5 of 75 methods achieve comprehensive learning across scenarios, showing a gap between scenario complexity and specialization. Future work should combine tailored states and specialized objectives with modular, transferable components. Intersections need multi-directional conflict resolution, roundabouts require circular-trajectory planning with multi-exit strategies, and both remain underexplored.

References

[1] D. Omeiza, H. Webb, M. Jirotka, and L. Kunze, "Explanations in autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10142–10162, Aug. 2022.

[2] H. A. Ignatious, H.-E. Sayed, and M. A. Khan, "An overview of sensors in autonomous vehicles," *Proc. Comput. Sci.*, vol. 198, pp. 736–741, Dec. 2022.

[3] C. Badué et al., "Self-driving cars: A survey," *Expert Syst. Appl.*, vol. 165, Mar. 2020, Art. no. 113816.

[4] Z. Lin, Q. Zhang, Z. Tian, P. Yu, and J. Lan, "DPL-SLAM: Enhancing dynamic point-line SLAM through dense semantic methods," *IEEE Sensors J.*, vol. 24, no. 9, pp. 14596–14607, May 2024.

[5] Z. Lin et al., "SLAM2: Simultaneous localization and multimode mapping for indoor dynamic environments," *Pattern Recognit.*, vol. 158, Feb. 2025, Art. no. 111054.

[6] H. M. Zhang, "Enhancing vehicle safety—The role of PID control in adaptive cruise control systems," *AIP Conf. Proc.*, vol. 3194, 2024, Art. no. 050023.

[7] J. Chen, W. Zhan, and M. Tomizuka, "Autonomous driving motion planning with constrained iterative LQR," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 2, pp. 244–254, Jun. 2019.

[8] N. Ye, D. Wang, and Y. Dai, "Enhancing autonomous vehicle lateral control: A linear complementarity model-predictive control approach," *Appl. Sci.*, vol. 13, no. 19, p. 10809, Sep. 2023.

[9] B. Li et al., "Adaptive pure pursuit: A real-time path planner using tracking controllers to plan safe and kinematically feasible paths," *IEEE Trans. Intell. Veh.*, vol. 8, no. 9, pp. 4155–4168, Sep. 2023.

[10] I. Emmanuel, "Fuzzy logic-based control for autonomous vehicle: A survey," *Int. J. Educ. Manage. Eng.*, vol. 7, no. 2, pp. 41–49, Mar. 2017.

[11] J. Liu, Z. Yang, Z. Huang, W. Li, S. Dang, and H. Li, "Simulation performance evaluation of pure pursuit, stanley, LQR, MPC controller for autonomous vehicles," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Jul. 2021, pp. 1444–1449.

[12] H. Vijayakumar et al., "A holistic safe planner for automated driving considering interaction with human drivers," *IEEE Trans. Intell. Vehicles*, vol. 9, no. 1, pp. 2061–2076, Jan. 2024.

[13] J. Zhou et al., "Robust min-max model predictive vehicle platooning with causal disturbance feedback," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15878–15897, Sep. 2022.

[14] Z. Lin, Z. Tian, Q. Zhang, H. Zhuang, and J. Lan, "Enhanced visual SLAM for collision-free driving with lightweight autonomous cars," *Sensors*, vol. 24, no. 19, p. 6258, Sep. 2024.

[15] M. Abdel-Aty and S. Ding, "A matched case-control analysis of autonomous vs human-driven vehicle accidents," *Nature Commun.*, vol. 15, no. 1, p. 4931, 2024.

[16] Z. Wang et al., "APF-CPP: An artificial potential field based multi-robot online coverage path planning approach," *IEEE Robot. Autom. Lett.*, vol. 9, no. 11, pp. 9199–9206, Nov. 2024.

[17] S. Fu, "Robot path planning optimization based on RRT and APF fusion algorithm," in *Proc. 8th Int. Conf. Robot. Autom. Sci. (ICRAS)*, Jun. 2024, pp. 32–36.

[18] H. Yu, S. Huo, M. Zhu, Y. Gong, and Y. Xiang, "Machine learning-based vehicle intention trajectory recognition and prediction for autonomous driving," in *Proc. 7th Int. Conf. Adv. Algorithms Control Eng. (ICAACE)*, Mar. 2024, pp. 771–775.

[19] T. Sharma et al., "Deep learning-based object detection and classification for autonomous vehicles in different weather scenarios of Quebec, Canada," *IEEE Access*, vol. 12, pp. 13648–13662, 2024.

[20] 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 Trans. Intell. Vehicles*, vol. 8, no. 4, pp. 2821–2831, Apr. 2023.

[21] M. Ganesan, S. Kandhasamy, B. Chokkalingam, and L. Mihet-Popa, "A comprehensive review on deep learning-based motion planning and end-to-end learning for self-driving vehicle," *IEEE Access*, vol. 12, pp. 66031–66067, 2024.

[22] M. Almehdhar et al., "Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 869–906, 2024.

[23] K. Yeom, "Deep reinforcement learning based autonomous driving with collision free for mobile robots," *Int. J. Mech. Eng. Robot. Res.*, vol. 11, no. 5, pp. 338–344, 2022.

[24] A. J. M. Muzahid, S. F. Kamarulzaman, Md. A. Rahman, and A. H. Alenezi, "Deep reinforcement learning-based driving strategy for avoidance of chain collisions and its safety efficiency analysis in autonomous vehicles," *IEEE Access*, vol. 10, pp. 43303–43319, 2022.

[25] H. Yang, Y. Zhou, J. Wu, H. Liu, L. Yang, and C. Lv, "Human-guided continual learning for personalized decision-making of autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 4, pp. 5435–5447, Apr. 2025.

[26] H. Huang, J. Liu, B. Zhang, S. Zhao, B. Li, and J. Wang, "LEAD: Learning-enhanced adaptive decision-making for autonomous driving in dynamic environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 5, pp. 6142–6156, May 2025.

[27] R. Shi, T. Li, Y. Yamaguchi, and L. Zhang, "Traffic scene-informed attribution of autonomous driving decisions," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 7, pp. 1–12, Jul. 2025.

[28] K. Yang, S. Li, M. Wang, and X. Tang, "Interactive decision-making integrating graph neural networks and model predictive control for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 5, pp. 1–15, May 2025.

[29] Z. Tian et al., "Efficient and balanced exploration-driven decision making for autonomous racing using local information," *IEEE Trans. Intell. Vehicles*, vol. 10, no. 3, pp. 1732–1748, Mar. 2025.

[30] Z. Tian, D. Zhao, Z. Lin, D. Flynn, W. Zhao, and D. Tian, "Balanced reward-inspired reinforcement learning for autonomous vehicle racing," in *Proc. 6th Annu. Learn. Dyn. Control Conf.*, 2024, pp. 628–640.

[31] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 1994, pp. 157–163.

[32] W. Zhou, D. Chen, J. Yan, Z. Li, H. Yin, and W. Ge, "Multi-agent reinforcement learning for cooperative lane changing of connected and autonomous vehicles in mixed traffic," *Auto. Intell. Syst.*, vol. 2, no. 1, p. 5, Dec. 2022.

[33] G.-P. Antonio and C. Maria-Dolores, "Multi-agent deep reinforcement learning to manage connected autonomous vehicles at tomorrow's intersections," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7033–7043, Jul. 2022.

[34] M. Bouton, A. Nakhaei, K. Fujimura, and M. J. Kochenderfer, "Cooperation-aware reinforcement learning for merging in dense traffic," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3441–3447.

[35] B. Kiumarsi et al., "Optimal convergence rate of reinforcement learning for linear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 2, pp. 750–763, Feb. 2018.

[36] E. O. Neftci and B. B. Averbeck, "Reinforcement learning in artificial and biological systems," *Nature Mach. Intell.*, vol. 1, no. 3, pp. 133–143, Mar. 2019.

[37] M. Lin, T. Chen, H. Chen, B. Ren, and M. Zhang, "When architecture meets AI: A deep reinforcement learning approach for system of systems design," *Adv. Eng. Informat.*, vol. 56, Apr. 2023, Art. no. 101965.

[38] P. Zhang, C. Wang, N. Kumar, and L. Liu, "Space-air-ground integrated multi-domain network resource orchestration based on virtual network architecture: A DRL method," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2798–2808, Mar. 2022.

[39] Z. Wang, Z. Luo, and Y. Lei, "DRL-based RIS-assisted integrated sensing and communication joint design of beamforming and reflection phase shift," *Phys. Commun.*, vol. 71, Aug. 2025, Art. no. 102669.

[40] L. Abualigah, S. Ekinci, D. Izci, and R. A. Zitar, "Modified elite opposition-based artificial hummingbird algorithm for designing FOPID controlled cruise control system," *Intell. Autom. Soft Comput.*, vol. 38, no. 2, pp. 169–183, 2023.

[41] S. Tang et al., "A survey on automated driving system testing: Landscapes and trends," *ACM Trans. Softw. Eng. Methodol.*, vol. 32, no. 5, pp. 1–62, Sep. 2023.

[42] J. Wu, Z. Song, and C. Lv, "Deep reinforcement learning-based energy-efficient decision-making for autonomous electric vehicle in dynamic traffic environments," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 1, pp. 875–887, Mar. 2024.

[43] W. Yue, X. Wu, C. Li, N. Cheng, P. Duan, and Z. Han, "Navigating the impact of connected and automated vehicles on mixed traffic efficiency: A driving behavior perspective," *IEEE Internet Things J.*, vol. 11, no. 23, pp. 37770–37784, Dec. 2024.

[44] Y. Liu, X. Zhao, Y. Tian, and J. Sun, "Sociality probe: Game-theoretic inverse reinforcement learning for modeling and quantifying social patterns in driving interaction," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 12, pp. 20841–20853, Dec. 2024.

[45] X. Huang et al., “A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability,” *Comput. Sci. Rev.*, vol. 37, Aug. 2020, Art. no. 100270.

[46] F. Fan, J. Xiong, M. Li, and G. Wang, “On interpretability of artificial neural networks: A survey,” *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 6, pp. 741–760, Jun. 2021.

[47] F. Calamunci and J. Lonsky, “The road to crime: An unintended consequence of the interstate highway system,” *Econ. J.*, vol. 135, no. 667, pp. 748–772, Mar. 2025.

[48] M. Gross, “Speed tourism: The German autobahn as a tourist destination and location of ‘unruly rules,”’ *Tourist Stud.*, vol. 20, no. 3, pp. 298–313, Sep. 2020.

[49] N. Khadka and H. Hassan, “Exploring the effects of connected and autonomous vehicles on traffic safety and operation at full-cloverleaf interchanges,” *J. Transp. Saf. Secur.*, vol. 17, no. 1, pp. 1–29, Jan. 2025.

[50] T. Schönhofer, B. Kaltenhäuser, and K. Bogenberger, “Simulation approach to estimate the potential of high-occupancy toll lanes on German freeways,” *Transp. Res. Rec.*, vol. 2678, no. 12, pp. 764–775, 2024.

[51] A. Pratelli and R. R. Souleyrette, “Visibility, perception and roundabout safety,” *WIT Trans. Built Environ.*, vol. 107, pp. 577–588, Jun. 2009.

[52] M. Naderi, M. Papageorgiou, I. Karafyllis, and I. Papamichail, “Automated vehicle driving on large lane-free roundabouts,” in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2022, pp. 1528–1535.

[53] L. Elefteriadou, “Interchanges and alternative intersection designs,” in *An Introduction To Traffic Flow Theory*. Cham, Switzerland: Springer, 2024, pp. 295–308.

[54] S. S. Channamallu, D. Almaskati, S. Kermanshachi, and A. Pamidimukkala, “Autonomous vehicle safety: A comprehensive analysis of crash injury determinants,” in *Proc. Int. Conf. Transp. Develop.*, Jun. 2024, pp. 767–779.

[55] I. C. Burnett, “Traffic collisions in north carolina: Weather, human factors, and economic analysis, 2013 to 2019,” Ph.D. thesis, Dept. Econ., North Carolina State Univ., Raleigh, NC, USA, 2013.

[56] J. Wang, J. Wu, X. Zheng, D. Ni, and K. Li, “Driving safety field theory modeling and its application in pre-collision warning system,” *Transp. Res. C, Emerg. Technol.*, vol. 72, pp. 306–324, Nov. 2016.

[57] R. Nahata, D. Omeiza, R. Howard, and L. Kunze, “Assessing and explaining collision risk in dynamic environments for autonomous driving safety,” in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 223–230.

[58] M. Wang, L. Zhang, Z. Zhang, and Z. Wang, “A hybrid trajectory planning strategy for intelligent vehicles in on-road dynamic scenarios,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 2832–2847, Mar. 2023.

[59] A. Botros and S. L. Smith, “Spatio-temporal lattice planning using optimal motion primitives,” *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11950–11962, Nov. 2023.

[60] W. Chen, Y. Chen, S. Wang, F. Kong, X. Zhang, and H. Sun, “Motion planning using feasible and smooth tree for autonomous driving,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 6270–6282, May 2023.

[61] F. Bouchard et al., “A rule-based behaviour planner for autonomous driving,” in *Proc. Int. Joint Conf. Rules Reasoning*, 2022, pp. 263–279.

[62] C. Zhang and L. Sun, “Bayesian calibration of the intelligent driver model,” *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 8, pp. 9308–9320, Aug. 2024.

[63] C. Wang, F. Guo, S. Zhao, Z. Zhu, and Y. Zhang, “Safety assessment for autonomous vehicles: A reference driver model for highway merging scenarios,” *Accid. Anal. Prev.*, vol. 206, Oct. 2024, Art. no. 107710.

[64] P. Koopman and W. H. Widen, “A reasonable driver standard for automated vehicle safety,” in *Proc. Int. Conf. Comput. Saf.*, 2023, pp. 355–361.

[65] A. Kuznietsov, B. Gyevnar, C. Wang, S. Peters, and S. V. Albrecht, “Explainable AI for safe and trustworthy autonomous driving: A systematic review,” *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 12, pp. 19342–19364, Dec. 2024.

[66] R. Gajjar and D. Mohandas, “Critical assessment of road capacities on urban roads—A Mumbai case-study,” *Transp. Res. Proc.*, vol. 17, pp. 685–692, Nov. 2016.

[67] M. A. S. Kamal, T. Hayakawa, and J. Imura, “Road-speed profile for enhanced perception of traffic conditions in a partially connected vehicle environment,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 6824–6837, Aug. 2018.

[68] M. A. S. Kamal, S. Taguchi, and T. Yoshimura, “Efficient driving on multilane roads under a connected vehicle environment,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2541–2551, Sep. 2016.

[69] D. A. Hensher, “Valuation of travel time savings,” in *A Handbook of Transport Economics*. Ann Arbor, MI, USA: Edwards, 2011.

[70] F. Steck, V. Kolarova, F. Bahamonde-Birke, S. Trommer, and B. Lenz, “How autonomous driving may affect the value of travel time savings for commuting,” *Transp. Res. Rec.*, vol. 2672, no. 46, pp. 11–20, Dec. 2018.

[71] C. Zhai, F. Luo, Y. Liu, and Z. Chen, “Ecological cooperative look-ahead control for automated vehicles travelling on freeways with varying slopes,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1208–1221, Feb. 2019.

[72] S. A. Birrell, M. Fowkes, and P. A. Jennings, “Effect of using an in-vehicle smart driving aid on real-world driver performance,” *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1801–1810, Aug. 2014.

[73] C. Sun, J. Guanetti, F. Borrelli, and S. J. Moura, “Optimal eco-driving control of connected and autonomous vehicles through signalized intersections,” *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3759–3773, May 2020.

[74] F. Paparella, G. Olivieri, G. Volpe, A. M. Mangini, and M. P. Fanti, “A deep reinforcement learning approach for route planning of autonomous vehicles,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2024, pp. 2047–2052.

[75] M. Zhao, X. Li, Y. Lu, H. Wang, and S. Ning, “A path-planning algorithm for autonomous vehicles based on traffic stability criteria: The AS-IAPF algorithm,” *Mech. Sci.*, vol. 15, no. 2, pp. 613–631, Nov. 2024.

[76] L. Luo, Y. Liu, Y. Feng, H. X. Liu, and Y.-E. Ge, “Stabilizing traffic flow by autonomous vehicles: Stability analysis and implementation considerations,” *Transp. Res. C, Emerg. Technol.*, vol. 158, Jan. 2024, Art. no. 104449.

[77] C. Antonya, C. Iclodean, and I.-A. Roşu, “Enhancing roundabout efficiency through autonomous vehicle coordination,” *Vehicles*, vol. 7, no. 1, p. 9, Jan. 2025.

[78] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, “Bottleneck transformers for visual recognition,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16514–16524.

[79] R. Xu, J. Joshi, and C. Li, “NN-EMD: Efficiently training neural networks using encrypted multi-sourced datasets,” *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 4, pp. 2807–2820, Jul. 2022.

[80] H. Touvron et al., “ResMLP: Feedforward networks for image classification with data-efficient training,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 5314–5321, Apr. 2023.

[81] J. Fang, F. Wang, J. Xue, and T. Chua, “Behavioral intention prediction in driving scenes: A survey,” *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 8, pp. 8334–8355, Aug. 2024.

[82] X. Li, W. Wang, and M. Roetting, “Estimating driver’s lane-change intent considering driving style and contextual traffic,” *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3258–3271, Sep. 2019.

[83] D. Xu, H. Zhao, F. Guillemard, S. Geronimi, and F. Aioun, “Aware of scene vehicles—Probabilistic modeling of car-following behaviors in real-world traffic,” *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2136–2148, Jun. 2019.

[84] P. Jardin, I. Moisidis, S. S. Zetina, and S. Rinderknecht, “Rule-based driving style classification using acceleration data profiles,” in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Dec. 2020, pp. 1–6.

[85] R. Vogel et al., “You can’t drive my car—A method to fingerprint individual driving styles in a sim-racing setting,” in *Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET)*, Jul. 2022, pp. 1–9.

[86] F. Lateef, M. Kas, and Y. Ruichek, “Saliency heat-map as visual attention for autonomous driving using generative adversarial network (GAN),” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5360–5373, Jun. 2022.

[87] N. Ding, C. Zhang, and A. Eskandarian, “SalienDet: A saliency-based feature enhancement algorithm for object detection for autonomous driving,” *IEEE Trans. Intell. Vehicles*, vol. 9, no. 1, pp. 2624–2635, Jan. 2024.

[88] Z. Cui, M. Li, Y. Huang, Y. Wang, and H. Chen, “An interpretation framework for autonomous vehicles decision-making via SHAP and RF,” in *Proc. CAA Int. Conf. Veh. Control Intell.*, 2022, pp. 1–7.

[89] P. M. Dassanayake, A. Anjum, A. K. Bashir, J. Bacon, R. Saleem, and W. Manning, “A deep learning based explainable control system for reconfigurable networks of edge devices,” *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 7–19, Jan. 2022.

[90] M. Zemni, M. Chen, É. Zablocki, H. Ben-Younes, P. Pérez, and M. Cord, "OCTET: Object-aware counterfactual explanations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 15062–15071.

[91] Z. Chen, F. Xiao, F. Guo, and J. Yan, "Interpretable machine learning for building energy management: A state-of-the-art review," *Adv. Appl. Energy*, vol. 9, Feb. 2023, Art. no. 100123.

[92] S. Hooker, D. Erhan, P.-J. Kindermans, and B. Kim, "A benchmark for interpretability methods in deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9737–9748.

[93] R. Tomsett, D. Harborne, S. Chakraborty, P. Gurram, and A. Preece, "Sanity checks for saliency metrics," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, 2020, pp. 6021–6029.

[94] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 9505–9515.

[95] A. Ghorbani, A. Abid, and J. Zou, "Interpretation of neural networks is fragile," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 3681–3688.

[96] A. A. Ismail, M. K. Gunady, L. Pessoa, H. C. Bravo, and S. Feizi, "Input-cell attention reduces vanishing saliency of recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 10814–10824.

[97] M. Wu, M. Hughes, S. Parbhoo, M. Zazzi, V. Röth, and F. Doshi-Velez, "Beyond sparsity: Tree regularization of deep models for interpretability," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1670–1678.

[98] T. Speith and M. Langer, "A new perspective on evaluation methods for explainable artificial intelligence (XAI)," in *Proc. IEEE 31st Int. Requirements Eng. Conf. Workshops (REW)*, Sep. 2023, pp. 325–331.

[99] W. Wang et al., "DriveMLM: Aligning multi-modal large language models with behavioral planning states for autonomous driving," 2023, *arXiv:2312.09245*.

[100] Z. Xu et al., "DriveGPT4: Interpretable end-to-end autonomous driving via large language model," *IEEE Robot. Autom. Lett.*, vol. 9, no. 10, pp. 8186–8193, Oct. 2024.

[101] X. Mu, T. Qin, S. Zhang, C. Xu, and M. Yang, "Pix2Planning: End-to-end planning by vision-language model for autonomous driving on Carla simulator," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2024, pp. 2383–2390.

[102] J. Cheng, Y. Chen, X. Mei, B. Yang, B. Li, and M. Liu, "Rethinking imitation-based planners for autonomous driving," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 14123–14130.

[103] A. Baheri, S. Nageshrao, H. E. Tseng, I. Kolmanovsky, A. Girard, and D. Filev, "Deep reinforcement learning with enhanced safety for autonomous highway driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 1550–1555.

[104] M. Kaushik, V. Prasad, K. M. Krishna, and B. Ravindran, "Overtaking maneuvers in simulated highway driving using deep reinforcement learning," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1885–1890.

[105] N. Albarella, D. Lui, A. Petrillo, and S. Santini, "A hybrid deep reinforcement learning and optimal control architecture for autonomous highway driving," *Energies*, vol. 16, no. 8, p. 3490, Apr. 2023.

[106] M. Hua et al., "Optimizing distributed energy system with an enhanced reinforcement learning–model predictive control algorithm," *Available SSRN J.*, 2024. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4876862

[107] Z. Lin et al., "A conflicts-free, speed-lossless KAN-based reinforcement learning decision system for interactive driving in roundabouts," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 10, pp. 16377–16390, Oct. 2025.

[108] J. Gómez-Romero, "Explaining deep reinforcement learning-based methods for control of building HVAC systems," *Methods*, vol. 15, p. 52, Oct. 2022.

[109] Y. Zhu, Z. Li, J. Wang, Y. Zhao, and M. Li, "Behavior safety decision-making based on deep deterministic policy gradient and its verification method," *Symmetry*, vol. 17, no. 1, p. 132, Jan. 2025.

[110] C. Yu, X. Wang, J. Hao, and Z. Feng, "Reinforcement learning for cooperative overtaking," in *Proc. Int. Conf. Auto. Agents Multiagent Syst.*, 2019, pp. 341–349.

[111] M. B. Ozcelik, B. Agin, O. Caldiran, and O. Sirin, "Decision making for autonomous driving in a virtual highway environment based on generative adversarial imitation learning," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2023, pp. 1–6.

[112] S. Zhang, Y. Wu, and H. Ogai, "Spatial attention for autonomous decision-making in highway scene," in *Proc. 59th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, Sep. 2020, pp. 1435–1440.

[113] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, pp. 1805–1824, Aug. 2000.

[114] S. Nageshrao, H. E. Tseng, and D. Filev, "Autonomous highway driving using deep reinforcement learning," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 2326–2331.

[115] J. Zhao, T. Qu, and F. Xu, "A deep reinforcement learning approach for autonomous highway driving," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 542–546, 2020.

[116] W. Zhao et al., "A spatial-state-based omni-directional collision warning system for intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 10, pp. 14344–14358, Oct. 2024.

[117] B. M. Albaba and Y. Yildiz, "Driver modeling through deep reinforcement learning and behavioral game theory," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 2, pp. 885–892, Mar. 2022.

[118] J. Yang, "Variational constrained reinforcement learning with application to planning at roundabout," in *Proc. ICLR Conf. Blind Submission*, 2019. [Online]. Available: https://openreview.net/forum?id=H1lJJnR5Ym

[119] M. Schutera, N. Goby, D. Neumann, and M. Reischl, "Transfer learning versus multi-agent learning regarding distributed decision-making in highway traffic," 2018, *arXiv:1810.08515*.

[120] X. Xu, L. Zuo, X. Li, L. Qian, J. Ren, and Z. Sun, "A reinforcement learning approach to autonomous decision making of intelligent vehicles on highways," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 10, pp. 3884–3897, Oct. 2020.

[121] Z. Bai, W. Shangguan, B. Cai, and L. Chai, "Deep reinforcement learning based high-level driving behavior decision-making model in heterogeneous traffic," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 8600–8605.

[122] T. Liu, Q. Liu, H. Liu, and X. Ren, "Combining deep reinforcement learning with rule-based constraints for safe highway driving," in *Proc. China Autom. Congr. (CAC)*, Nov. 2022, pp. 2785–2790.

[123] 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.

[124] W. Yuan, M. Yang, Y. He, C. Wang, and B. Wang, "Multi-reward architecture based reinforcement learning for highway driving policies," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3810–3815.

[125] S. Aradi, T. Becsi, and P. Gaspar, "Policy gradient based reinforcement learning approach for autonomous highway driving," in *Proc. IEEE Conf. Control Technol. Appl. (CCTA)*, Aug. 2018, pp. 670–675.

[126] H. Wang, S. Yuan, M. Guo, C.-Y. Chan, X. Li, and W. Lan, "Tactical driving decisions of unmanned ground vehicles in complex highway environments: A deep reinforcement learning approach," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 235, no. 4, pp. 1113–1127, Mar. 2021.

[127] A. M. Naveen, R. Ravish, and S. Ranga Swamy, "Distributional reinforcement learning for automated driving vehicle," in *Proc. IEEE 2nd Mysore Sub Sect. Int. Conf. (MysuruCon)*, Oct. 2022, pp. 1–6.

[128] J. Wang, T. Yang, L. Tang, Y. Wang, and R. Xiong, "Learning an efficient and safe policy for highway driving using supervised learning and reinforcement learning," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Aug. 2019, pp. 112–117.

[129] K. Lv, X. Pei, C. Chen, and J. Xu, "A safe and efficient lane change decision-making strategy of autonomous driving based on deep reinforcement learning," *Mathematics*, vol. 10, no. 9, p. 1551, 2022.

[130] R. Rădulescu, M. Legrand, K. Efthymiadis, D. M. Roijers, and A. Nowé, "Deep multi agent reinforcement learning in a homogeneous open population," in *Proc. 30th Benelux Conf. Artif. Intell. (BNAIC)*, 2019, pp. 90–105.

[131] C. Yu et al., "Distributed multiagent coordinated learning for autonomous driving in highways based on dynamic coordination graphs," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 735–748, Feb. 2020.

[132] M. Kaushik, N. Singhania, S. Phaniteja, and K. M. Krishna, "Parameter sharing reinforcement learning architecture for multi agent driving," in *Proc. 4th Int. Conf. Adv. Robot.*, Jul. 2019, pp. 1–7.

[133] M. Molaie, A. Amirkhani, and H. Kashiani, "Auto-driving policies in highway based on distributional deep reinforcement learning," in *Proc. 5th Int. Conf. Pattern Recognit. Image Anal. (IPRIA)*, Apr. 2021, pp. 1–6.

[134] G. Chen, Y. Zhang, and X. Li, "Attention-based highway safety planner for autonomous driving via deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 162–175, Jan. 2024.

[135] K. Min, H. Kim, and K. Huh, "Deep distributional reinforcement learning based high-level driving policy determination," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 3, pp. 416–424, Sep. 2019.
[136] B. Gangopadhyay, H. Soora, and P. Dasgupta, "Hierarchical program-triggered reinforcement learning agents for automated driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10902–10911, Aug. 2022.
[137] S. Cheng, B. Yang, Z. Wang, and K. Nakano, "Spatio-temporal image representation and deep-learning-based decision framework for automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24866–24875, Dec. 2022.
[138] B. Liu, Y. Tang, Y. Ji, Y. Shen, and Y. Du, "A deep reinforcement learning approach for ramp metering based on traffic video data," *J. Adv. Transp.*, vol. 2021, pp. 1–13, Oct. 2021.
[139] M. Yang, Z. Li, Z. Ke, and M. Li, "A deep reinforcement learning-based ramp metering control framework for improving traffic operation at freeway weaving sections," in *Proc. Transp. Res. Board Annu. Meeting*, Washington, DC, USA, 2019, pp. 13–17.
[140] P. Wang and C.-Y. Chan, "Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
[141] Y. Lin, J. McPhee, and N. L. Azad, "Anti-jerk on-ramp merging using deep reinforcement learning," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 7–14.
[142] F. Deng, J. Jin, Y. Shen, and Y. Du, "Advanced self-improving ramp metering algorithm based on multi-agent deep reinforcement learning," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3804–3809.
[143] M. Cheng, C. Zhang, H. Jin, Z. Wang, and X. Yang, "Adaptive coordinated variable speed limit between highway mainline and on-ramp with deep reinforcement learning," *J. Adv. Transp.*, vol. 2022, pp. 1–16, Sep. 2022.
[144] Z. Hu and W. Ma, "Guided deep reinforcement learning for coordinated ramp metering and perimeter control in large scale networks," *Transp. Res. C, Emerg. Technol.*, vol. 159, Feb. 2024, Art. no. 104461.
[145] D. Deng, B. Yu, D. Xu, Y. Chen, and Y. Kong, "Automated traffic state optimization in the weaving area of urban expressways by a reinforcement learning-based cooperative method of channelization and ramp metering," *J. Adv. Transp.*, vol. 2023, pp. 1–13, Apr. 2023.
[146] S. Zhou, W. Zhuang, G. Yin, H. Liu, and C. Qiu, "Cooperative on-ramp merging control of connected and automated vehicles: Distributed multi-agent deep reinforcement learning approach," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2022, pp. 402–408.
[147] C. Wang, Y. Xu, J. Zhang, and B. Ran, "Integrated traffic control for freeway recurrent bottleneck based on deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15522–15535, Sep. 2022.
[148] X. Qi, L. Zhang, P. Wang, J. Yang, T. Zou, and W. Liu, "Learning-based MPC for autonomous motion planning at freeway off-ramp diverging," *IEEE Trans. Intell. Veh.*, vol. 9, no. 10, pp. 6183–6194, Oct. 2024.
[149] Z. E. A. Kherroubi, S. Aknine, and R. Bacha, "Novel decision-making strategy for connected and autonomous vehicles in highway on-ramp merging," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12490–12502, Aug. 2022.
[150] X. Zhang, L. Wu, H. Liu, Y. Wang, H. Li, and B. Xu, "High-speed ramp merging behavior decision for autonomous vehicles based on multiagent reinforcement learning," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 22664–22672, Dec. 2023.
[151] D. Chen et al., "Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11623–11638, Nov. 2023.
[152] X. He, B. Lou, H. Yang, and C. Lv, "Robust decision making for autonomous vehicles at highway on-ramps: A constrained adversarial reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 4103–4113, Apr. 2023.
[153] M. Li, Z. Li, S. Wang, and S. Zheng, "Enhancing cooperation of vehicle merging control in heavy traffic using communication-based soft actor-critic algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 6491–6506, Jun. 2023.
[154] J. Chen, B. Yuan, and M. Tomizuka, "Model-free deep reinforcement learning for urban autonomous driving," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 2765–2771.
[155] G. Bacchiani, D. Molinari, and M. Patander, "Microscopic traffic simulation by cooperative multi-agent deep reinforcement learning," in *Proc. 18th Int. Conf. Auto. Agents MultiAgent Syst.*, 2019, pp. 1547–1555.
[156] B. Montgomery, C. Muise, and S. Givigi, "Hierarchical deep reinforcement learning with cross-attention and planning for autonomous roundabout navigation," in *Proc. Can. Conf. Electr. Comput. Eng.*, 2024, pp. 417–423.
[157] Y. Tian et al. (2019). *Variational Constrained Reinforcement Learning With Application to Planning at Roundabout*. [Online]. Available: https://openreview.net/forum?id=H1e3HlSFDr
[158] W. Wang, L. Jiang, S. Lin, H. Fang, and Q. Meng, "Imitation learning based decision-making for autonomous vehicle control at traffic roundabouts," *Multimedia Tools Appl.*, vol. 81, no. 28, pp. 39873–39889, Nov. 2022.
[159] L. Ferrarotti et al., "Autonomous and human-driven vehicles interacting in a roundabout: A quantitative and qualitative evaluation," *IEEE Access*, vol. 12, pp. 32693–32705, 2024.
[160] Y. Zhang, B. Gao, L. Guo, H. Guo, and H. Chen, "Adaptive decision-making for automated vehicles under roundabout scenarios using optimization embedded reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5526–5538, Dec. 2021.
[161] F. Konstantinidis, M. Sackmann, O. De Candido, U. Hofmann, J. Thielecke, and W. Utschick, "Parameter sharing reinforcement learning for modeling multi-agent driving behavior in roundabout scenarios," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 1974–1981.
[162] R. Nakaya, T. Harada, Y. Miura, K. Hattori, and J. Matsuoka, "Emergence of cooperative automated driving control at roundabouts using deep reinforcement learning," in *Proc. Annu. Conf. Soc. Instrum. Control Engineers*, 2023, pp. 97–102.
[163] H. Yuan, P. Li, B. Van Arem, L. Kang, H. Farah, and Y. Dong, "Safe, efficient, comfort, and energy-saving automated driving through roundabout based on deep reinforcement learning," in *Proc. IEEE 26th Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2023, pp. 6074–6079.
[164] W. Wang, F. Hui, J. Zhang, and S. Fang, "Deep reinforcement learning method for trajectory planning of connected and autonomous vehicles in the roundabout lane-changing scenario," in *Proc. 4th Int. Symp. Comput. Technol. Inf. Sci. (ISCTIS)*, Jul. 2024, pp. 168–173.
[165] A. P. Capasso, G. Bacchiani, and A. Broggi, "From simulation to real world maneuver execution using deep reinforcement learning," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 1570–1575.
[166] A. Paolo Capasso, G. Bacchiani, and D. Molinari, "Intelligent roundabout insertion using deep reinforcement learning," 2020, *arXiv:2001.00786*.
[167] S. Alighanbari and N. L. Azad, "Deep reinforcement learning with NMPC assistance Nash switching for urban autonomous driving," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 3, pp. 2604–2615, Mar. 2023.
[168] B. Peng, M. F. Keskin, B. Kulcsár, and H. Wymeersch, "Connected autonomous vehicles for improving mixed traffic efficiency in unsignalized intersections with deep reinforcement learning," *Commun. Transp. Res.*, vol. 1, Dec. 2021, Art. no. 100017.
[169] A. Pozzi, S. Bae, Y. Choi, F. Borrelli, D. M. Raimondo, and S. Moura, "Ecological velocity planning through signalized intersections: A deep reinforcement learning approach," in *Proc. 59th IEEE Conf. Decis. Control (CDC)*, Dec. 2020, pp. 245–252.
[170] K.-F. Chu, A. Y. S. Lam, and V. O. K. Li, "Traffic signal control using end-to-end off-policy deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7184–7195, Jul. 2022.
[171] D. Quang Tran and S.-H. Bae, "Proximal policy optimization through a deep reinforcement learning framework for multiple autonomous vehicles at a non-signalized intersection," *Appl. Sci.*, vol. 10, no. 16, p. 5722, Aug. 2020.
[172] Z. Bai, P. Hao, W. ShangGuan, B. Cai, and M. J. Barth, "Hybrid reinforcement learning-based eco-driving strategy for connected and automated vehicles at signalized intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15850–15863, Sep. 2022.
[173] R. Bautista-Montesano, R. Galluzzi, K. Ruan, Y. Fu, and X. Di, "Autonomous navigation at unsignalized intersections: A coupled reinforcement learning and model predictive control approach," *Transp. Res. C, Emerg. Technol.*, vol. 139, Jun. 2022, Art. no. 103662.
[174] D. Li, F. Zhu, T. Chen, Y. D. Wong, C. Zhu, and J. Wu, "COOR-PLT: A hierarchical control model for coordinating adaptive platoons of connected and autonomous vehicles at signal-free intersections based on deep reinforcement learning," *Transp. Res. C, Emerg. Technol.*, vol. 146, Jan. 2023, Art. no. 103933.

[175] S. Kai, B. Wang, D. Chen, J. Hao, H. Zhang, and W. Liu, "A multi-task reinforcement learning approach for navigating unsignalized intersections," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 1583–1588.

[176] B. Zhou, Q. Zhou, S. Hu, D. Ma, S. Jin, and D.-H. Lee, "Cooperative traffic signal control using a distributed agent-based deep reinforcement learning with incentive communication," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 8, pp. 10147–10160, Aug. 2024.

[177] W. X. Hu, H. Ishihara, C. Chen, A. Shalaby, and B. Abdulhai, "Deep reinforcement learning two-way transit signal priority algorithm for optimizing headway adherence and speed," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 7920–7931, Aug. 2023.

[178] A. Lombard, A. Noubli, A. Abbas-Turki, N. Gaud, and S. Galland, "Deep reinforcement learning approach for V2X managed intersections of connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7178–7189, Jul. 2023.

[179] D. Li, J. Wu, M. Xu, Z. Wang, and K. Hu, "Adaptive traffic signal control model on intersections based on deep reinforcement learning," *J. Adv. Transp.*, vol. 2020, pp. 1–14, Aug. 2020.

[180] H. Shu, T. Liu, X. Mu, and D. Cao, "Driving tasks transfer using deep reinforcement learning for decision-making of autonomous vehicles in unsignalized intersection," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 41–52, Jan. 2022.

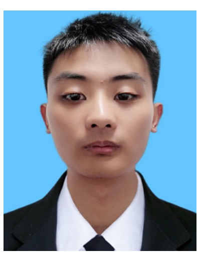
[181] H. Seong, C. Jung, S. Lee, and D. H. Shim, "Learning to drive at unsignalized intersections using attention-based deep reinforcement learning," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 559–566.

[182] C.-J. Hoel, T. Tram, and J. Sjöberg, "Reinforcement learning with uncertainty estimation for tactical decision-making in intersections," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–7.

[183] W. Xiao et al., "Decision-making for autonomous vehicles in random task scenarios at unsignalized intersection using deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 73, no. 6, pp. 7812–7825, Jun. 2024.

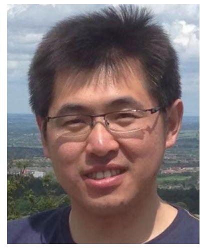
[184] Z. Guo, Y. Wu, L. Wang, and J. Zhang, "Coordination for connected and automated vehicles at non-signalized intersections: A value decomposition-based multiagent deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3025–3034, Mar. 2023.

[185] I. Osband, J. Aslanides, and A. Cassirer, "Randomized prior functions for deep reinforcement learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 8626–8638.



Zhen Tian received the B.Eng. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2020. He is currently pursuing the Ph.D. degree with the James Watt School of Engineering, University of Glasgow, Glasgow. His main research interests include interactive autonomous driving and autonomous racing.

Zhihao Lin received the M.S. degree from the College of Electronic Science and Engineering, Jilin University, Changchun, China. He is currently pursuing the Ph.D. degree with the James Watt School of Engineering, University of Glasgow, U.K. His research interests focus on multi-sensor fusion SLAM systems and robot perception in complex scenarios.

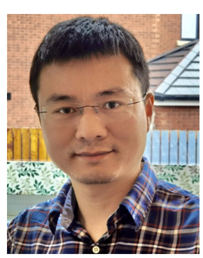


Dezong Zhao (Senior Member, IEEE) received the Ph.D. degree in control engineering from Tsinghua University, Beijing, China, in 2010. He is currently a Professor of autonomous systems with the University of Glasgow and a Turing Fellow with The Alan Turing Institute. He was awarded the Royal Academy of Engineering/Leverhulme Trust Research Fellow in 2025, the Royal Society-Newton Advanced Fellow in 2020, and the EPSRC Innovation Fellow in 2018.

Wenjing Zhao received the Ph.D. degree in traffic engineering from Central South University, Changsha, China, in 2022. She is currently a Post-Doctoral Research Associate with the James Watt School of Engineering, University of Glasgow, Glasgow, U.K. Her research interests include traffic safety, driving behaviour analysis, and connected vehicles.

David Flynn (Senior Member, IEEE) received the Ph.D. degree in microscale magnetic components from Heriot-Watt University, Edinburgh, U.K., in 2007. He is currently a Professor of cyber physical systems with the University of Glasgow. He is also the Co-Founder of U.K.'s EPSRC National Centre for Energy System Integration and U.K. Offshore Robotics and Artificial Intelligence Hub for Offshore Energy Asset Integrity Management.

Shuja Ansari received the M.Sc. (Hons.) and Ph.D. degrees in engineering from Glasgow Caledonian University in 2015 and 2019, respectively. He is currently a Lecturer with the James Watt School of Engineering, University of Glasgow. His research interests include wireless communications, terrestrial/airborne mobile networks, security and privacy in communications, and the Internet of Things.



Chongfeng Wei (Senior Member, IEEE) received the Ph.D. degree in mechanical engineering from the University of Birmingham in 2015. He is currently a Senior Lecturer with the University of Glasgow. His research interests include decision-making and control of intelligent vehicles, human-centric autonomous driving, cooperative automation, and dynamics and control of mechanical systems.