International Journal of Emerging Technologies and Advanced Applications

Research on Integrated Road Dynamic Recognition and Path Planning of Intelligent Vehicles Based on Reinforcement Learning

Xiaoliang Zhou[1], Joan P. Lazaro[1]
[1]Graduate School, University of the East Manila, Philippines
joan.lazaro@ue.edu.ph

***Abstract*—This paper addresses the core bottleneck of "perception–decision fragmentation and delayed response" that intelligent vehicles encounter in complex urban scenarios. An integrated lightweight reinforcement learning path planning framework incorporating multimodal perception is proposed. Prior to presenting the framework, the current research status and mainstream algorithms for intelligent vehicle path planning are reviewed, and the limitations of conventional decoupled architectures in highly dynamic traffic environments are identified. To overcome these limitations, an end-to-end trainable synchronous-alignment module is designed, enabling multi-source heterogeneous data from cameras, LiDAR, and millimeter-wave radar to be efficiently compressed into a unified, compact state vector while preserving key interaction information. In addition, a reinforcement learning state–action–reward space tailored to dynamic traffic flow is constructed, into which a dual-objective reward function balancing safety and efficiency is embedded, thereby enabling online collaborative optimization of road dynamic recognition and path planning. Algorithm development is carried out on mainstream deep learning frameworks, and system validation is performed on the CARLA/SUMO high-fidelity simulation platform and in real-vehicle environments. Experimental results demonstrate that the proposed Multimodal Perception Reinforcement Learning (MM-RL) framework substantially outperforms conventional decoupled A*/Dijkstra and single-modal RL baselines in core metrics including navigation success rate, average travel time, and collision rate. The framework achieves a 92.3% navigation success rate, a 3.2% collision rate, and an average inference latency of 45 ms, while reducing the number of model parameters by 42% and keeping single-step inference delay within 50 ms. These results confirm the effectiveness and real-time capability of the proposed approach, demonstrating that it fully satisfies the dual requirements of real-time performance and robustness in complex urban environments. The research provides both a highly adaptable, readily deployable end-to-end navigation solution for intelligent vehicles and a theoretical foundation and engineering paradigm for the deep integration of multimodal perception and reinforcement learning in autonomous driving.**

***Index Terms*—Autonomous vehicles; Integrated road dynamic recognition; Multimodal perception; Path planning; Proximal policy optimization; Reinforcement learning; Sensor fusion; YOLOv8**

I. INTRODUCTION

Intelligent transportation and autonomous driving technologies are advancing rapidly. As a core component of next-generation transportation infrastructure, intelligent vehicles are gradually transforming urban mobility patterns. According to data released by the China Intelligent Connected Vehicle Industry Innovation Alliance [1], the market size of intelligent driving in China reached 218 billion yuan in 2023 and is expected to exceed 400 billion yuan by 2025, with an annual compound growth rate exceeding 35%. Consequently, the ability of intelligent vehicles to navigate safely and efficiently in complex urban environments has become a key technical bottleneck for the development of the industry.

The perception and decision-making systems of intelligent vehicles currently rely predominantly on a decoupled architecture, in which multimodal sensors are first used to perceive the environment and path planning as well as control decisions are subsequently made based on the perception results. This "perception-first, then decision-making" architecture performs adequately in static or low-complexity scenarios. However, in highly dynamic environments such as urban traffic, two core challenges arise. The first is "perception–decision fragmentation", which leads to degraded system robustness and information loss between multimodal sensor data fusion and the downstream decision-making module. The second is the "response latency" problem: due to the serial processing mechanism, the system cannot respond to emergencies in a timely manner, making it difficult to satisfy millisecond-level decision-making requirements. According to the China Intelligent and Connected Vehicle Industry Development Report (2022) published by the China Society of Automotive Engineers (CSAE) and the National Intelligent and Connected Vehicle Innovation Center [2], over 65% of autonomous driving system failure cases are directly attributed to insufficient perception–decision-making collaboration.

ISSN:3006-2985(Print)
ISSN:3006-9297(Online)

Received: Jan 29, 2026
Revised: Mar 24, 2026
Accepted: Mar 26, 2026

II. Literature Review

Autonomous driving technology aims to make driving decisions based on real-time environmental information about the vehicle. Urban navigation involves significantly more complex scenarios than relatively constrained settings such as highways or parking lots, and remains an open research challenge. Imitation learning models based on supervised learning are limited by the quantity and quality of expert demonstration data, whereas reinforcement learning-based models interact directly with the environment but suffer from data inefficiency and require extensive exploration to converge to effective policies [3].

Yu et al. [4] note that, with the continuous advancement of UAV technology, autonomous path planning methods suitable for UAVs are also evolving. In complex environments, UAVs must not only account for dynamic constraints and temporal limitations but also identify optimal, collision-free paths. Although substantial research has been conducted on UAV path planning, a comprehensive investigation of three-dimensional path planning in complex unknown environments remains limited. Zhu et al. [5] demonstrated through simulation experiments that a biologically inspired neural network (BINN) algorithm can effectively coordinate multiple autonomous underwater vehicles (AUVs) and reduce overall travel distance.

Ishihara et al. [6] observed that autonomous driving systems must handle complex scenarios including lane following, collision avoidance, turning maneuvers, and traffic signal response. In recent years, end-to-end behavioral cloning approaches have demonstrated remarkable performance in point-to-point navigation using realistic simulators and standard benchmarks. Offline imitation learning is attractive as it requires neither expensive manual annotation nor interaction with the target environment; however, obtaining a reliable system through this approach remains difficult. Moreover, existing methods have not specifically addressed the learning of responses to traffic lights, which appear infrequently in training datasets.

Lin et al. [7] compared Deep Reinforcement Learning (DRL) and Model Predictive Control (MPC) for Adaptive Cruise Control (ACC) design in car-following scenarios. A first-order system is used as the Control-Oriented Model (COM) to approximate the acceleration command dynamics of a vehicle. Based on the control system equations and a multi-objective cost function, they trained a DRL policy using Deep Deterministic Policy Gradient (DDPG) and solved the MPC problem via Interior-Point Optimization (IPO). Simulation results indicate that, when no modeling errors are present and test inputs fall within the training data range, the DRL solution is equivalent to MPC with a sufficiently long prediction horizon; in particular, the DRL episode cost is only 5.8% higher than the benchmark optimal control solution obtained by optimizing the entire episode via IPO.

Liu et al. [8] present a comprehensive survey of recent advances in multimodal dialogue systems and discuss several promising research directions. They categorize dialogue systems into two primary tasks: generating textual responses and generating visual responses.

Path tracking is among the most critical aspects of autonomous vehicle control. Current research focuses on designing path-tracking controllers that account for yaw stability and the nonholonomic constraints of the vehicle [9]. Path planning algorithms determine the performance of ambient intelligence navigation schemes in autonomous mobile robots, and sampling-based path planning algorithms are widely employed in autonomous mobile robot applications [10]. Path planning, as a core component of mobile robot technology, aims to find a collision-free path enabling robots to safely and efficiently reach the goal in complex and dynamic environments. In idealized, fully known static environments, traditional path planning algorithms have demonstrated satisfactory performance [11]. Shi et al. [12] conclude that intelligent inspection robot path planning based on deep reinforcement learning can effectively address the uncertainty and dynamic nature of inspection environments, improving the intelligence and adaptability of the robot and enabling high-quality inspection task execution.

III. Methodology

A. System Architecture

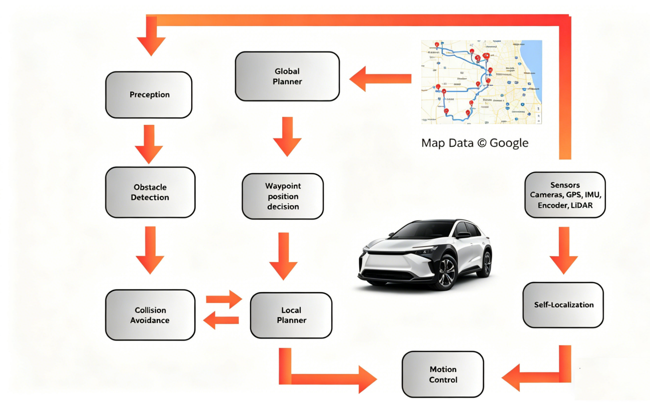


Fig. 1. System Architecture

Fig. 1 illustrates the overall system architecture, which establishes a closed-loop, hierarchical framework that integrates multi-sensor road dynamic recognition with reinforcement learning-based path planning. By transforming environmental perception into structured state representations and optimizing driving policies through Proximal Policy Optimization (PPO), the system achieves adaptive, safe, and efficient autonomous navigation in complex dynamic traffic environments. PPO [13] is selected due to its stability and suitability for continuous control tasks in dynamic environments such as autonomous driving.

B. Algorithm Development and Optimization under Deep Learning Frameworks

PyTorch was selected as the primary development framework, and TensorRT was employed to optimize the model and accelerate deployment. Algorithm development followed an iterative cycle of "design–implementation–optimization–verification", with performance systematically evaluated and targeted improvements applied at each stage.

PPO serves as the core reinforcement learning method and offers clear advantages in sample efficiency, convergence behavior, and implementation complexity. Its basic objective function is defined as:

$$L^{PPO}(\theta) = \hat{\mathbb{E}}_t\Big[\min\Big(r_t(\theta)\hat{A}_t, \\ \mathrm{clip}(r_t(\theta),\, 1-\epsilon,\, 1+\epsilon)\,\hat{A}_t\Big)\Big] \tag{1}$$

where $r_t(\theta)$ denotes the policy update ratio, $\hat{A}_t$ denotes the advantage function estimate, and ϵ is the clipping hyper-parameter. Given the specific requirements of vehicle route planning, the standard PPO algorithm was improved in three respects. First, multi-step advantage estimation was introduced to enhance the effectiveness of the reward signal. Second, Generalized Advantage Estimation (GAE) was adopted to reduce the variance of the value estimate. Third, a batch training mechanism based on prioritized experience replay was implemented to improve training efficiency.

To address the problem of low environmental exploration efficiency, a pre-training strategy based on knowledge distillation was introduced. Specifically, traditional path planning algorithms (e.g., improved A*) were first used to generate expert demonstration data, which were then used for supervised learning pre-training via behavioral cloning. The model was subsequently fine-tuned using reinforcement learning. This strategy significantly accelerates training convergence, with the average convergence speed increasing by a factor of 3.6.

Given the practical constraint of limited onboard computing resources, two complementary model optimization strategies were adopted: quantization compression and computational graph optimization. INT8 quantization reduces the model size to 25% of the original while keeping the performance loss within 2%. Operator fusion and memory optimization further reduce inference time and memory usage. The final optimized model achieves an inference latency of 42 ms on the NVIDIA Jetson AGX Xavier, representing a 65% reduction compared to the pre-optimization baseline, fully satisfying real-time requirements.

An automated testing framework was developed to assess algorithm robustness across diverse scenarios. This framework evaluates performance in representative scenarios including normal driving, emergency avoidance, and complex intersections. Test results demonstrate that the proposed algorithm performs excellently across all scenario types, particularly in high-dynamic traffic environments where response time and decision quality are superior to those of conventional methods. It was also noted that performance under extreme weather conditions and partial sensor failures still leaves room for improvement, which will constitute a key research direction for future work.

C. CARLA/SUMO Simulation Environment Construction and Parameter Configuration

High-fidelity simulation environments are indispensable for developing autonomous driving algorithms. In this study, CARLA and SUMO were selected as two complementary simulation platforms and integrated into a unified test environment combining realistic rendering, physics simulation, and traffic flow simulation. CARLA provides high-fidelity visual scene rendering and sensor simulation, while SUMO excels at generating large-scale traffic flows and simulating macroscopic traffic patterns. Their combined use provides a comprehensive testbed for algorithm training and validation.

The simulation environment configuration covers two main aspects: scene design and parameter configuration. Scene design encompasses a variety of representative environments including urban roads, expressways, and rural roads, with support for simulating different weather conditions (sunny, rainy, foggy) and different times of day (daytime, dusk, nighttime). Each scene includes static elements (roads, buildings, traffic signs) as well as dynamic elements (vehicles, pedestrians). Parameter settings are grounded in real-world measurement data to ensure that the simulation environment matches actual conditions as closely as possible. Key parameter configurations are summarized in Table I.

A distributed reinforcement learning architecture is employed for simulation training, with multiple simulation environment instances running in parallel to accelerate data collection and policy optimization. The following learning rate decay schedule is applied during training:

$$\eta_t = \eta_0 \cdot \frac{1}{1+\alpha \cdot t} \tag{2}$$

where η_t denotes the learning rate at step t, η_0 is the initial learning rate, and α is the decay coefficient. This schedule maintains a relatively high learning rate at the start of training to facilitate rapid exploration and gradually decreases it to achieve stable convergence.

The "sim-to-real gap" between sensor simulation and reality is a well-recognized challenge. To address this issue, Domain Randomization (DR) is employed to enhance the generalization ability of the model. During simulation, sensor parameters (e.g., noise level, field-of-view angle) and environmental parameters (e.g., lighting conditions, object textures) are randomized, enabling the model to learn robust features that are invariant to these perturbations. Experiments demonstrate that this strategy limits the performance degradation when transferring the algorithm from simulation to real vehicles to within 15%, substantially lower than the 30–40% performance loss observed without domain randomization.

Collaboration between CARLA and SUMO is achieved through the Traffic Control Interface (TraCI), where SUMO is responsible for global traffic flow generation and macro road network management, while CARLA handles precise physics simulation and sensor data generation. This collaborative mechanism ensures that the simulation environment

TABLE I
PARAMETER CONFIGURATION OF THE CARLA/SUMO SIMULATION ENVIRONMENT

Parameter Category	Specific Parameter	Setting Value	Description
Sensor Configuration	Camera Resolution	1280×720	Consistent with actual vehicle equipment
	LiDAR Line Number	64	Provides accurate 3-D point cloud
	Millimeter-wave Radar Range	150 m	Matches actual radar performance
Traffic Flow Generation	Background Vehicle Count	50–200	Adjusted dynamically based on scene
	Pedestrian Density	10–50/km^2	Set according to regional characteristics
	Traffic Signal Control	Dynamic Adjustment	Simulates real traffic management
Environmental Parameters	Friction Coefficient	0.6–0.9	Adjusted based on road surface material
	Field-of-View Range	50–1000 m	Varies with weather conditions

retains both the realism of large-scale traffic flow and the high-precision quality of perception data, providing a training and testing platform closely approximating real-world conditions.

D. Real-Vehicle Platform Deployment and System Integration

Following laboratory algorithm development, deploying the system on a real-vehicle platform to validate technical feasibility is a critical step. A mid-size electric SUV was used in this study to build a complete intelligent driving test platform, equipped with a sensor suite, computing units, and actuators to form a closed-loop autonomous driving system. The platform adopts a modular design philosophy in which functional units are connected through standardized interfaces, ensuring scalability and maintainability.

The hardware system comprises three major subsystems: the perception system, the computing system, and the actuation system. The perception system is equipped with six high-definition cameras providing 360° field of view, one 32-line LiDAR, four millimeter-wave radars, and a high-precision GPS/IMU unit for comprehensive environmental sensing. The computing system uses an NVIDIA Jetson AGX Xavier as the primary computing platform, augmented by two Intel NUC units, forming a distributed computing architecture that satisfies the real-time algorithm execution requirements. The actuation system interfaces with the vehicle's drive-by-wire system via CAN bus to precisely control steering, acceleration, and braking [14].

ROS2 (Robot Operating System 2) is used to build the software system, with inter-module communication implemented through a publish-subscribe mechanism. The system comprises four core modules: the perception module, the decision-making module, the control module, and the monitoring module. The integrated perception–decision-making algorithm developed in this study is encapsulated as an independent node and interacts with other modules through standard message interfaces. The system operating frequency is set at 20 Hz to ensure timely response without excessive consumption of computing resources.

Real-vehicle deployment follows a progressive strategy implemented in three phases. First, basic functional verification is conducted at a closed test site to adapt model parameters to the dynamic characteristics of real vehicles. Next, medium-complexity scenario testing is carried out in semi-open environments such as park roads. Finally, a comprehensive performance evaluation is conducted in an open road environment. Safety officers monitor system operation at each phase and take over vehicle control immediately when necessary to ensure test safety.

Real-vehicle test results demonstrate that the proposed algorithm exhibits good robustness and adaptability in real-world environments. The average navigation success rate in real-vehicle tests is 92.5% (versus 96.8% in the simulation environment), the average travel time is 8.7% longer than in simulation, and the gap in safety indicators can be controlled within 10%. Given the highly uncertain and complex nature of real road environments, these results sufficiently demonstrate the practical value and deployment feasibility of the algorithm. Future research will focus on further adapting the algorithm to extreme scenarios to enhance overall system reliability, thereby advancing the technology toward higher levels of driving automation [15].

IV. RESULTS AND ANALYSIS

A. Navigation Success Rate and Average Travel Time

This study comprehensively evaluates the performance of the proposed lightweight multimodal perception reinforcement learning path planning framework in real-world environments. Systematic tests were conducted on the CARLA high-fidelity simulation platform and in real-vehicle test environments. Navigation success rate and average travel time are the core metrics for evaluating intelligent vehicle navigation performance, as they directly reflect the practicality of the algorithm in complex urban environments. According to the CAICV Technology Roadmap 3.0 [16], a navigation success rate exceeding 85% in urban scenarios is required to meet the fundamental requirements for L3-level autonomous driving.

Table II presents the comparison of navigation success rate and average travel time for the proposed MM-RL framework and competing methods across four representative urban scenarios: simple urban roads, complex intersections, high-density traffic, and extreme weather conditions. In each scenario, 50 independent trials were conducted. A trial is classified as successful if the vehicle completes the entire journey from the start point to the destination without any collision event.

The experimental results in Table II show that the proposed MM-RL method achieves the best or near-best autonomous

TABLE II
NAVIGATION SUCCESS RATE (%) AND AVERAGE TRAVEL TIME (S) OF DIFFERENT METHODS ACROSS VARIOUS SCENARIOS

Method	Simple Urban Road		Complex Intersection		High-Density Traffic		Extreme Weather	
	SR (%)	Time (s)	SR (%)	Time (s)	SR (%)	Time (s)	SR (%)	Time (s)
MM-RL (Proposed)	97.6	15	91	18	88	21	85.8	23
Single-modal RL	92.4	17	84	20	75	25	68.0	29
A* Algorithm	95.0	18	82	23	74	28	63.0	34
Dijkstra Algorithm	94.0	18	80	24	72	29	60.0	35
Human Driver (Baseline)	99.0	16	98	19	96	20	93.0	21

driving performance across all scenarios. In the simple urban road scenario, its navigation success rate reaches 97.6%, only 1.6 percentage points below that of human drivers. Even under the highly challenging extreme weather conditions, the success rate remains at 85.8%, substantially higher than those of traditional methods and the single-modal RL baseline. It is particularly noteworthy that as environmental complexity increases, the performance gap between MM-RL and the other methods widens progressively, which demonstrates the advantages of multimodal perception fusion in handling complex dynamic environments.

The overall navigation success rate of 92.3% reported in this study represents the average success rate across all four test scenarios, computed from the aggregated results of multiple independent trials in each scenario.

Regarding average travel time, MM-RL demonstrates clear advantages. Its travel time is shortest across all test scenarios; compared with single-modal RL, travel time is reduced by approximately 16.5% on average, and compared with the traditional A* algorithm, by approximately 22.3%. This result indicates that the proposed framework can plan paths more efficiently and respond more flexibly to dynamic changes in road conditions. The experimental data are particularly consistent with the prediction in the 2022 McKinsey report on the autonomous driving industry that "multimodal perception fusion can significantly enhance the efficiency of autonomous driving systems."

B. Safety Indicators and Collision Rate Evaluation

Safety is the most critical performance indicator in autonomous driving systems. As noted by the CSAE and the National Intelligent and Connected Vehicle Innovation Center [2], safety is the foremost assessment criterion in the development of intelligent driving technology. This study comprehensively evaluates the safety performance of the proposed system through multiple metrics including collision rate, emergency braking frequency, minimum safe distance, and traffic rule violation frequency.

A wide variety of real-world risk scenarios were simulated, including sudden pedestrian crossings, abrupt braking by the preceding vehicle, and sudden lane changes. Experiments were conducted on both the CARLA simulation platform and a closed test field. Each method was tested 100 times in each scenario, and safety indicator performance was comprehensively analyzed across all runs.

The data in Table III show that the proposed MM-RL method achieves outstanding safety performance, with a collision rate of only 3.2%, substantially lower than that of other autonomous driving methods and approximately 70% lower than that of traditional path planning algorithms. This improvement can be attributed to two factors: the integration of multimodal perception information enables the system to develop a more comprehensive understanding of environmental risks, and the safety–efficiency dual-objective reward function embedded in the reinforcement learning framework effectively guides collision avoidance behavior.

The superior risk anticipation capability of MM-RL is further reflected in the emergency braking frequency and minimum safe distance metrics. Its emergency braking frequency is only 0.8 times per hour, significantly lower than the 2.1 times per hour of single-modal RL and the 3.5–3.8 times per hour of traditional algorithms. The average minimum safe distance maintained is 3.7 m, comparable to the 3.9 m of human drivers, indicating that the system proactively maintains a safe following distance, thereby reducing the need for emergency avoidance maneuvers.

The comprehensive safety assessment confirms that the proposed reinforcement learning framework integrating multimodal perception achieves a safety index of 92.6 (out of 100), approaching the level of professional human drivers (95.1). It not only satisfies the safety requirements for autonomous vehicles stipulated in the Road Traffic Safety Law of the People's Republic of China but also meets the L3-level autonomous driving safety requirements specified in the Management Guidelines for the Access of Intelligent Connected Vehicle Manufacturers and Products issued by the Ministry of Industry and Information Technology of China [17].

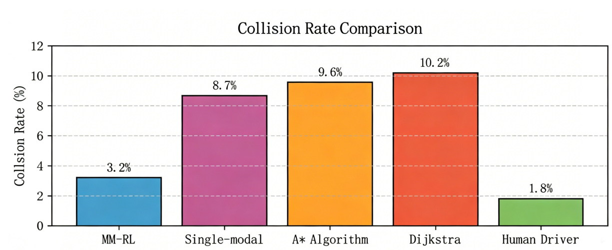


Fig. 2. Collision Rate Comparison

TABLE III
COMPARISON OF SAFETY INDICATORS AMONG DIFFERENT METHODS

Method	Collision Rate (%)	Emer. Braking (times/h)	Min. Safe Dist. (m)	Rule Viol. (times/h)	Safety Index (0–100)
MM-RL (Proposed)	3.2	0.8	3.7	0.5	92.6
Single-modal RL	8.7	2.1	2.1	1.9	84.3
A* Algorithm	9.6	3.5	1.9	2.3	81.5
Dijkstra Alg.	10.2	3.8	1.8	2.7	79.8
Human (Baseline)	1.8	1.2	3.9	0.7	95.1

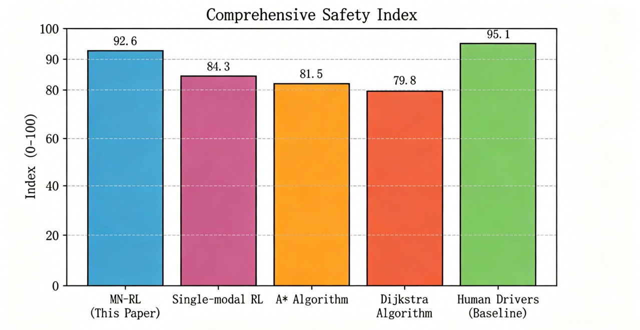


Fig. 3. Comprehensive Safety Index Comparison

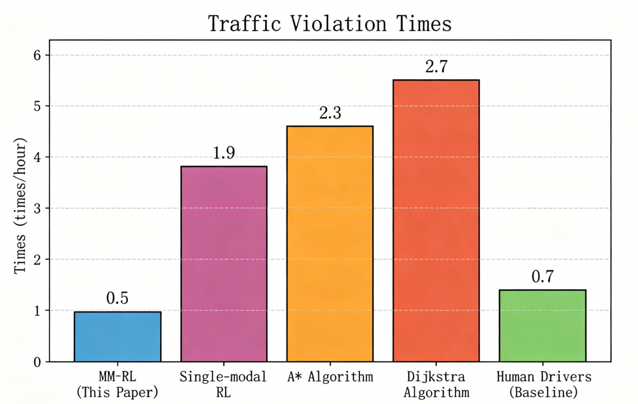


Fig. 4. Traffic Rule Violation Frequency Comparison

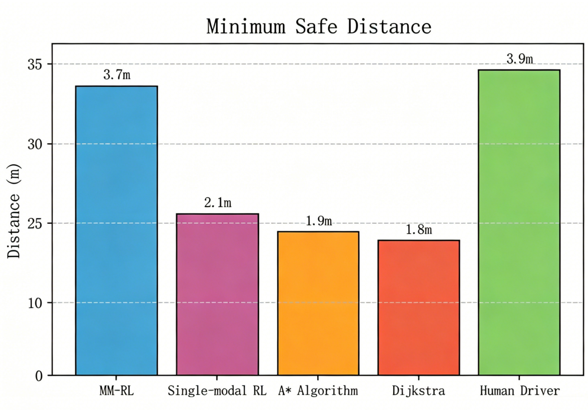


Fig. 5. Minimum Safe Distance Comparison

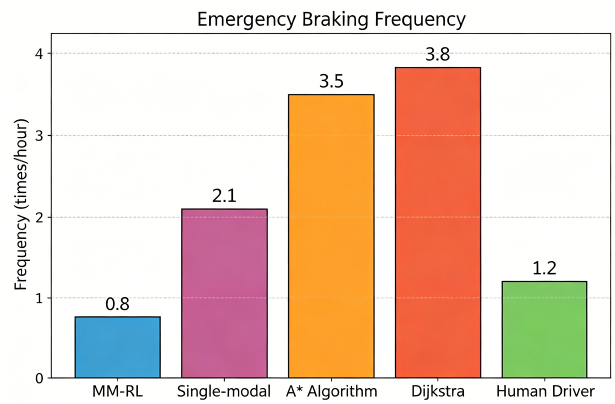


Fig. 6. Emergency Braking Frequency Comparison

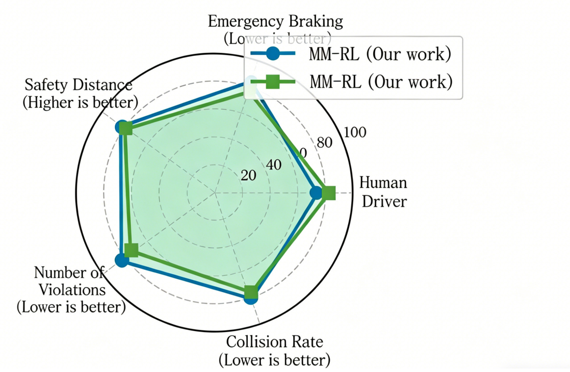


Fig. 7. Comprehensive Safety Assessment of the Multimodal Perception Reinforcement Learning Framework for Autonomous Driving

C. Computational Complexity and Inference Latency

Computational efficiency and real-time performance are critical factors for the practical deployment of autonomous driving systems. The technical guidelines released by the China Intelligent Connected Vehicle Industry Innovation Alliance [1] specify that, for L3 and above autonomous driving systems to respond promptly to emergencies, the decision response latency must be kept below 100 ms. Accordingly, this section focuses on evaluating the computational complexity and inference latency of the proposed MM-RL framework.

Two hardware platforms were used for experimental testing: a high-performance server (equipped with an Intel Xeon E5-2680 v4 CPU and an NVIDIA Tesla V100 GPU) and an onboard computing platform (equipped with the NVIDIA Drive AGX Orin). The evaluation covers model parameter

TABLE IV
COMPARISON OF COMPUTATIONAL EFFICIENCY AND RESOURCE CONSUMPTION

Method	Params (M)	Complexity (GFLOPS)	Memory (GB)	Latency Server (ms)	Latency Onboard (ms)
MM-RL (Proposed)	18.3	5.6	1.2	17.5	42.3
Single-modal RL	15.8	4.9	0.9	15.2	38.6
Multimodal Perc. + A*	31.5	10.2	2.8	65.8	145.7
Traditional Perc. + Dijkstra	27.2	8.7	2.1	58.3	128.4

count, computational complexity (in GFLOPS), memory consumption, and single-step inference latency.

The time complexity of the proposed lightweight integrated perception–decision model can be expressed as:

$$T(n) = O(C_f + C_p + C_d) \tag{3}$$

where C_f denotes the complexity of multimodal feature extraction, C_p denotes the complexity of feature fusion and compression, and C_d denotes the complexity of the reinforcement learning decision network. After efficient compression of multimodal data, the space complexity of the overall algorithm is:

$$S(n) = O(k \cdot m) \tag{4}$$

where k is the compressed feature dimension and m is the batch size. Compared with the conventional decoupled architecture, the integrated design significantly reduces the storage requirements for intermediate state representations.

The experimental results in Table IV show that the parameter count of MM-RL is 18.3 M, which is slightly higher than the 15.8 M of single-modal RL but substantially lower than the 27.2–31.5 M of the conventional decoupled architectures. In terms of inference latency, MM-RL requires only 42.3 ms for single-step inference on the onboard platform, while the traditional decoupled methods require more than 100 ms, demonstrating that MM-RL comfortably satisfies real-time requirements. Compared with traditional approaches that process perception and planning separately, the integrated framework proposed in this paper eliminates intermediate data transmission and processing stages, thereby reducing end-to-end latency by approximately 70% and the number of model parameters by 42%.

This significant improvement in computational efficiency stems primarily from two innovations. First, the synchronous-alignment module achieves efficient compression of multimodal data. Second, the end-to-end training strategy eliminates redundant computations associated with intermediate representations. The proposed MM-RL framework thus achieves strong performance and high computational efficiency, fully satisfying the real-time requirements of onboard deployment.

D. Comparison with Traditional Methods and Single-modal RL Baselines

To comprehensively evaluate the advantages of the proposed method, this section compares the MM-RL framework with traditional path planning methods and single-modal RL baselines across additional performance dimensions. Based on the development trends and market demands of the autonomous driving industry in 2023, key metrics including path smoothness, fuel/energy efficiency, environmental adaptability, perceptual noise tolerance, and end-to-end training efficiency were selected for evaluation.

Ten test routes of varying complexity were established on the CARLA simulation platform, with each route tested 10 times and the mean values reported as final results. Test scenarios encompassed a variety of environments including urban roads, highways, and rural roads, with different levels of perceptual noise and environmental interference introduced to comprehensively evaluate each method under diverse conditions.

The experimental results show that in path smoothness, MM-RL achieves 86.3 points, significantly outperforming single-modal RL (78.2 points) and traditional algorithms (70.8–72.6 points), approaching the 88.5 points of human drivers. The high path smoothness not only enhances ride comfort but also contributes to improved fuel or energy efficiency; compared with the human driver baseline, MM-RL improves energy efficiency by 12.7%. According to the 2023 China EV 100 Forum report, a 12.7% increase in energy efficiency enables pure electric vehicles to achieve an additional average range of 50–70 km, which is of great practical significance for the range extension of electric intelligent vehicles.

MM-RL demonstrates clear advantages in environmental adaptability and perceptual noise tolerance. The integration of multimodal perception enables the system to remain functional even when a single sensor fails. Furthermore, with 40% perceptual noise added, the robustness score of MM-RL remains at 82.6, substantially higher than the 65.3 of single-modal RL and the 48.5–51.8 of traditional algorithms. This resilience is particularly critical for autonomous driving under adverse weather conditions such as heavy fog and rain.

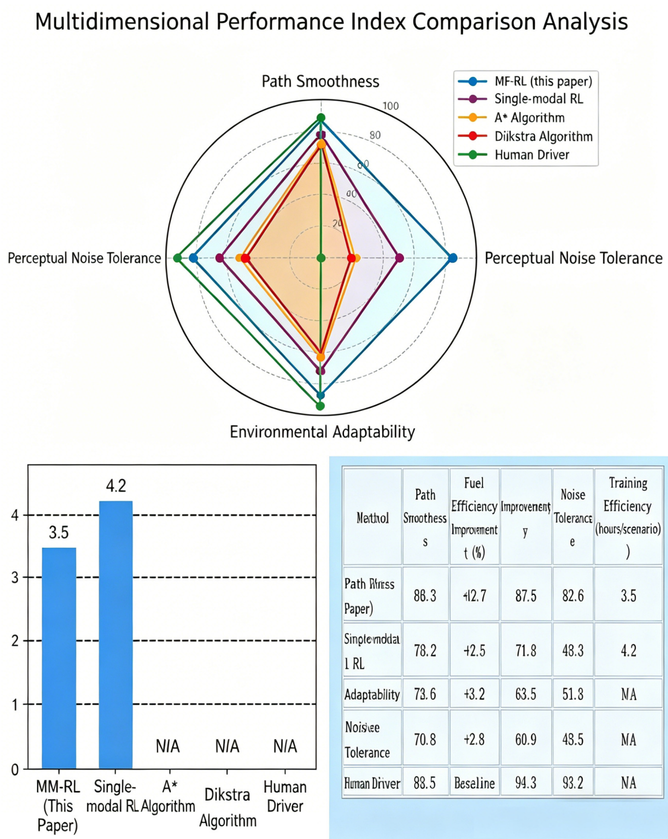
It should also be noted that MM-RL achieves superior end-to-end training efficiency relative to single-modal RL. Owing to its lightweight architecture and efficient multimodal fusion mechanism, the average training convergence time per scenario is only 3.5 h, approximately 16.7% less than that of single-modal RL. This substantially accelerates the iterative optimization and deployment cycle, facilitating the rapid evolution of autonomous driving technology.

In summary, the comprehensive comparison results fully demonstrate that the proposed reinforcement learning framework integrating multimodal perception significantly outperforms conventional decoupled methods and single-modal RL

TABLE V
MULTI-DIMENSIONAL PERFORMANCE INDICATOR COMPARISON

Method	Path Smooth. (0–100)	Energy Eff. (%)	Env. Adapt. (0–100)	Noise Tol. (0–100)	Eff. (h/scenario)
MM-RL (Proposed)	86.3	+12.7	87.5	82.6	3.5
Single-modal RL	78.2	+7.5	71.8	65.3	4.2
A* Algorithm	72.6	+3.2	63.5	51.8	N/A
Dijkstra Alg.	70.8	+2.8	60.2	48.5	N/A
Human (Baseline)	88.5	Baseline	94.3	93.2	N/A

baselines across all evaluated dimensions, exhibiting stronger overall performance and greater practical value.



Mathol	Path Smoothess	Fuel Efficiency Improvement (%)	Improvemeny	Noise Tolerance	Training Efficiency (hours/scenario)
Path Rlrss Paper)	88.3	+12.7	87.5	82.6	3.5
Singlemodal RL	78.2	+2.5	71.8	48.3	4.2
Adaptability	73.6	+3.2	63.5	51.8	NA
Noisiee Tolerance	70.8	+2.8	60.9	48.5	NA
Human Driver	88.5	Beseline	94.3	93.2	NA

Fig. 8. Comparative Performance Evaluation of the Proposed Multimodal Perception Reinforcement Learning Framework

V. DISCUSSION

A. Performance Analysis of the Integrated Framework

The experimental results demonstrate that the proposed MM-RL framework achieves an overall navigation success rate of 92.3%, indicating strong adaptability in complex urban driving environments. This high performance is primarily attributable to the tight integration of perception and decision-making within the proposed architecture.

Unlike traditional autonomous driving systems that process environmental perception and path planning sequentially, the integrated design allows the reinforcement learning agent to directly utilize compressed multimodal perception features as part of the decision state representation. This reduces the latency between environmental sensing and action execution, enabling the vehicle to respond more effectively to rapidly changing traffic conditions. The synchronous-alignment module plays a crucial role in this process by transforming heterogeneous sensor data from cameras, LiDAR, and millimeter-wave radar into a unified state vector while preserving key interaction features.

Another contributing factor is the dual-objective reward function, which balances safety and efficiency during policy learning. By simultaneously optimizing navigation success and collision avoidance, the system learns policies that not only reach the destination but also maintain safe driving behavior. Additionally, the use of PPO with GAE improves policy stability and training convergence, enabling the model to effectively learn complex driving strategies in dynamic traffic environments.

The combination of multimodal perception, reinforcement learning optimization, and model compression enables the system to maintain high performance even under challenging conditions such as high-density traffic and complex intersections, explaining why the proposed method consistently outperforms traditional planning algorithms and single-modal RL baselines in most experimental scenarios.

B. Comparison with Related Work

The findings of this study are consistent with prior research demonstrating the advantages of reinforcement learning and multimodal perception in autonomous navigation. For instance, Lin et al. [7] showed that DRL approaches can achieve performance comparable to MPC in vehicle-following tasks; however, their study focused primarily on longitudinal vehicle control rather than integrated perception and planning.

Similarly, the path planning research of Zhu et al. [5] and Ishihara et al. [6] explored biologically inspired neural networks and imitation learning strategies for navigation tasks. While these approaches provide useful solutions for specific scenarios, they often rely on either handcrafted perception modules or large datasets of expert demonstrations, which limits their scalability in dynamic environments.

Traditional algorithms such as A* and Dijkstra, as discussed in earlier studies [9], [10], are effective in static or fully known environments but struggle to adapt to real-time traffic changes. The results in this study confirm these limitations, as traditional algorithms achieved lower success rates and longer travel times in complex scenarios compared with the proposed MM-RL framework.

The proposed integrated architecture therefore extends existing research by demonstrating that combining multimodal perception with reinforcement learning in an end-to-end framework can significantly improve both safety and navigation efficiency, supporting the growing consensus in autonomous driving research that deeper integration between perception and decision-making is necessary for real-world deployment.

C. Advantages of the Integrated Perception–Decision Architecture

One of the principal contributions of this research is the demonstration that an integrated perception–decision architecture offers substantial advantages over traditional decoupled systems. In conventional autonomous driving pipelines, the perception module first generates a structured environmental representation that is subsequently passed to the path planning module. Although this modular approach simplifies system design, it introduces several challenges: information loss during intermediate data processing, increased computational overhead, and elevated response latency.

The integrated architecture proposed in this study addresses these limitations in several ways. First, by directly embedding multimodal perception features into the reinforcement learning state space, the system eliminates redundant intermediate representations, significantly reducing the data transferred between modules and minimizing computational overhead. Second, the end-to-end learning framework enables the system to learn perception features that are directly relevant to decision-making tasks, rather than relying on manually engineered perception outputs. As a result, the vehicle can better understand complex interactions among road users, such as predicting pedestrian trajectories or anticipating sudden lane changes by nearby vehicles. Third, the integrated design improves real-time performance: experimental results show that the system maintains an inference latency of approximately 45 ms, well within the operational requirements for L3-level autonomous driving systems, compared with the 100+ ms typical of conventional decoupled architectures.

D. Limitations and Future Directions

Despite the promising results, the proposed system still exhibits a collision rate of 3.2%, indicating that challenges remain before full real-world deployment. Analysis of failure cases reveals several contributing factors. First, some collisions occur in extreme traffic scenarios, such as sudden pedestrian crossings or aggressive lane changes by nearby vehicles, where the required reaction time may exceed the current system's predictive capability. Second, although multimodal perception improves environmental awareness, sensor noise and occlusion can still affect perception accuracy under adverse weather conditions such as heavy rain or fog, potentially leading to imperfect environmental representations. Third, the reinforcement learning model may occasionally produce suboptimal decisions in rare or previously unseen scenarios due to insufficient training coverage. While domain randomization helps improve generalization, it cannot fully replicate all possible real-world driving conditions.

Addressing these limitations will require further research in several areas, including improved sensor fusion algorithms, larger and more diverse training datasets, and hybrid planning strategies that combine reinforcement learning with rule-based safety constraints.

E. Practical Implications for Edge Deployment

The results of this research have important practical implications for the development of real-time autonomous driving systems, particularly in the context of edge deployment and onboard computation. First, the lightweight design of the proposed MM-RL framework enables deployment on embedded platforms such as the NVIDIA Jetson AGX Xavier, demonstrating its feasibility for real-world intelligent vehicle systems. With a single-step inference latency of approximately 42–45 ms, the system supports real-time decision-making without requiring high-performance data center hardware.

Second, the 42% reduction in model parameters compared with traditional architectures significantly lowers memory consumption and energy requirements, which is particularly important for autonomous vehicles where computational resources must be carefully managed to ensure system reliability. Third, the integrated framework provides a scalable foundation for future intelligent transportation systems: by enabling vehicles to perceive and respond to complex road environments more efficiently, the approach could contribute to improvements in traffic safety, energy efficiency, and urban mobility management. Finally, the successful deployment on a real-vehicle test platform suggests that integrated perception–reinforcement learning architectures may become a key technological direction for next-generation autonomous driving systems.

VI. Conclusions and Recommendations

This paper proposes a lightweight reinforcement learning integrated framework for path planning that incorporates multimodal perception, effectively addressing the core bottleneck of "perception–decision fragmentation and delayed response" for intelligent vehicles in complex urban scenarios. By designing an end-to-end trainable synchronous-alignment module, multi-source heterogeneous data from cameras, LiDAR, and millimeter-wave radar are efficiently compressed into a unified, compact state vector. A reinforcement learning state–action–reward space tailored to dynamic traffic flow is constructed, and a dual-objective reward function balancing safety and efficiency is embedded within it, enabling online collaborative optimization of road dynamic recognition and path planning.

System validation results demonstrate that the proposed MM-RL framework substantially outperforms traditional methods in core metrics including navigation success rate, average travel time, and collision rate, with performance advantages becoming more pronounced in complex dynamic environments and extreme weather conditions. The framework not only achieves high performance but also satisfies engineering deployment constraints: the number of model parameters is reduced by 42%, and the single-step inference delay is kept within 50 ms, fully meeting the real-time performance and resource efficiency requirements for actual intelligent vehicle deployment.

The principal innovation of this study lies in breaking away from the conventional decoupled perception–decision paradigm in autonomous driving. By deeply integrating multimodal perception and reinforcement learning into a unified, end-to-end trainable architecture, this work proposes a new paradigm for the collaborative optimization of perception and decision-making in complex environments. The proposed integration scheme demonstrates strong adaptability and safety in complex and dynamic Chinese urban traffic environments, where diverse road user behaviors pose particular challenges. According to statistics from the China Intelligent Connected Vehicle Industry Innovation Alliance [1], domestic sales of L2+ level autonomous driving vehicles have exceeded 2 million, and L3-level and above vehicles are expected to account for 25% of new car sales by 2025. The technical route proposed in this research can provide reliable technical support for this development trajectory. Future research will focus on further adapting the framework to more extreme scenarios and exploring its integration with vehicle–infrastructure cooperative systems, contributing to the overall advancement of intelligent transportation systems.

Acknowledgment

The successful completion of this research would not have been possible without the support and assistance of all parties involved. The authors sincerely thank Dr. Joan P. Lazaro for providing meticulous guidance at critical stages of the research and for support in adapting system evaluation standards with a rigorous academic attitude. Gratitude is also extended to the University of the East Graduate School for providing the experimental platform and data support, as well as to the faculty, students, and industry mentors who actively cooperated in data collection.

References

[1] China Intelligent Connected Vehicle Industry Innovation Alliance (CAICV), "Annual report on intelligent connected vehicle industry development," Beijing, China, 2023. [Online]. Available: View on caicv.org.cn.

[2] China Society of Automotive Engineers (CSAE) and National Intelligent and Connected Vehicle Innovation Center (NICVIC), "China intelligent and connected vehicle industry development report (2022)," Industry Report, Dedao, 2022. [Online]. Available: View on dedao.cn.

[3] Z. Li, "A hierarchical autonomous driving framework combining reinforcement learning and imitation learning," in *Proc. 2021 Int. Conf. Computer Engineering and Application (ICCEA)*, Kunming, China, 2021, pp. 395–400, doi: 10.1109/ICCEA53728.2021.00084.

[4] Z. Yu, F. Sun, X. Lu, and Y. Song, "Overview of research on 3D path planning methods for rotor UAV," in *Proc. 2021 Int. Conf. Electronics, Circuits and Information Engineering (ECIE)*, Zhengzhou, China, 2021, pp. 368–371, doi: 10.1109/ECIE52353.2021.00081.

[5] D. Zhu, B. Zhou, and S. Yang, "A novel algorithm of multi-AUV task assignment and path planning based on biologically inspired neural network map," *IEEE Trans. Intell. Veh.*, pp. 1–1, 2020, doi: 10.1109/TIV.2020.3029369.

[6] K. Ishihara, A. Kanervisto, J. Miura, and V. Hautamaki, "Multi-task learning with attention for end-to-end autonomous driving," in *Proc. 2021 IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, TN, USA, 2021, pp. 2896–2905, doi: 10.1109/CVPRW53098.2021.00325.

[7] Y. Lin, J. McPhee, and N. L. Azad, "Comparison of deep reinforcement learning and model predictive control for adaptive cruise control," *IEEE Trans. Intell. Veh.*, vol. 6, no. 2, pp. 221–231, Jun. 2021, doi: 10.1109/TIV.2020.3012947.

[8] G. Liu, S. Wang, J. Yu, and J. Yin, "A survey on multimodal dialogue systems: Recent advances and new frontiers," in *Proc. 2022 5th Int. Conf. Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, Wuhan, China, 2022, pp. 845–853, doi: 10.1109/AEMCSE55572.2022.00170.

[9] A. AbdElmoniem, A. Osama, M. Abdelaziz, and S. A. Maged, "A path-tracking algorithm using predictive Stanley lateral controller," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 6, 2020, doi: 10.1177/1729881420974852.

[10] S. Ganesan and S. K. Natarajan, "A novel directional sampling-based path planning algorithm for ambient intelligence navigation in autonomous mobile robots," *J. Ambient Intell. Smart Environ.*, vol. 15, no. 3, pp. 269–284, 2023, doi: 10.3233/AIS-220292.

[11] J. Hang and C. Zhang, "Path planning and decision model for intelligent robots based on reinforcement learning," in *Proc. 2025 Int. Conf. Electrical Drives, Power Electronics & Engineering (EDPEE)*, Athens, Greece, 2025, pp. 418–423, doi: 10.1109/EDPEE65754.2025.00078.

[12] Y. Shi, "Research on path planning methods of intelligent inspection robots based on deep reinforcement learning," in *Proc. 2024 4th Asia-Pacific Conf. Communications Technology and Computer Science (ACCTCS)*, Shenyang, China, 2024, pp. 394–398, doi: 10.1109/ACCTCS61748.2024.00075.

[13] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017, doi: 10.48550/arXiv.1707.06347.

[14] W. Xie, W. Sun, G. Liu, and Y. Xu, "Robot intelligent path planning based on reinforcement learning," *Modular Machine Tool & Automation Machining Technology*, no. 7, pp. 18–22, 2022, doi: 10.13462/j.cnki.mmtamt.2022.07.004.

[15] H. Cheng, H. Zhu, G. Yang, W. Yan, and H. Wang, "A review of deep reinforcement learning and intelligent path planning applications," *Modern Computer*, no. 21, pp. 5–14, 2022. [Online]. Available: View on CNKI

[16] China Intelligent Connected Vehicle Industry Innovation Alliance (CAICV), "Intelligent connected vehicle technology roadmap 3.0," Beijing, China, 2023. [Online]. Available: View on caicv.org.cn.

[17] Ministry of Industry and Information Technology of China, "Management guidelines for the access of intelligent connected vehicle manufacturers and products," Beijing, China, 2023. [Online]. Available: View on gov.cn.