PHYSICS-GUIDED MULTIMODAL MULTI-AGENT LEARNING FOR INTELLIGENT TRANSPORTATION SYSTEMS

Zhen Tian[1], **Yaqiong Zhang**[2], **Zhihao Lin**[1], **Fujiang Yuan**[3], **Yijun Lu**[4], **Wangjie Lang**[5], **Xinyu Wang**[6,†], **Ning Lyu**[7], **Zhiguo Tao**[8], **Kaijie Chen**[9], **Aaron Wang**[9]

[1]University of Glasgow, UK [2]Chongqing University of Technology, China [3]Waseda University, Japan
[4]Hangzhou Dianzi University, China [5]Hefei University of Technology, China
[6]University of Michigan–Ann Arbor, USA [7]Carnegie Mellon University, USA
[8]University of Nottingham, UK [9]Tongji University, China

ABSTRACT

Intelligent transportation systems (ITS) require reliable coordination among multiple vehicles operating under heterogeneous behaviors and time-varying traffic conditions. Prior approaches face a recurring trade-off: physics-based controllers provide interpretability and constraint handling but can be brittle under model mismatch, whereas data-driven policies adapt to complex scenarios but often lack safety guarantees and transparent decision logic. We propose a hierarchical, physics-guided framework that separates semantic coordination from continuous control execution. At the regional level, a large language model (LLM) generates discrete, human-interpretable coordination directives (e.g., yielding priority and target gaps) from multimodal observations. At the vehicle level, each directive is realized by a physics-informed controller augmented with a learned residual policy, where the residual is constrained and safety-filtered to preserve feasibility and closed-loop robustness. Multimodal fusion via vision–language models (VLMs) supports context-aware coordination by combining visual cues with textual traffic descriptors and temporal signals. In highway merging simulations, the proposed framework improves traffic throughput by 23% and reduces collision rates by 31% relative to classical and learning-based baselines, indicating that semantic coordination and physics-grounded execution can be combined without sacrificing safety-critical control requirements.

1 INTRODUCTION

Urban traffic congestion continues to impose substantial social, economic, and environmental burdens on modern cities. Recent analyses estimate that congestion-related delays cost billions of dollars in lost productivity and excess fuel consumption annually, while tailpipe emissions from inefficient traffic flows degrade urban air quality and contribute to climate impact Essamlali et al. (2025); Gao et al. (2025); Lin (2025); Yu et al. (2025b). These inefficiencies stem not merely from limited infrastructure capacity but also from complex interactions among heterogeneous vehicles, diverse driving behaviors, pedestrian unpredictability, and dynamic environmental factors Yu et al. (2025c). Consequently, attempts to alleviate congestion through road-widening or infrastructure expansion alone often lead to induced demand effects, resulting in only temporary improvements Ossokina et al. (2022); Wang et al. (2025a); Xin et al. (2024); Cao et al. (2025c;b). Meanwhile, Connected and Autonomous Vehicles (CAVs) have emerged as a promising direction for transforming traffic systems by enabling vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) coordination Islam et al. (2023); Xin et al. (2025b); Yu (2025); Lin et al. (2025b); Zheng et al. (2025a;b); Tian et al. (2025b); Lin et al. (2024b). However, unlocking the efficiency potential of CAVs requires coordination strategies that remain robust, interpretable, and reliable across diverse traffic situations.

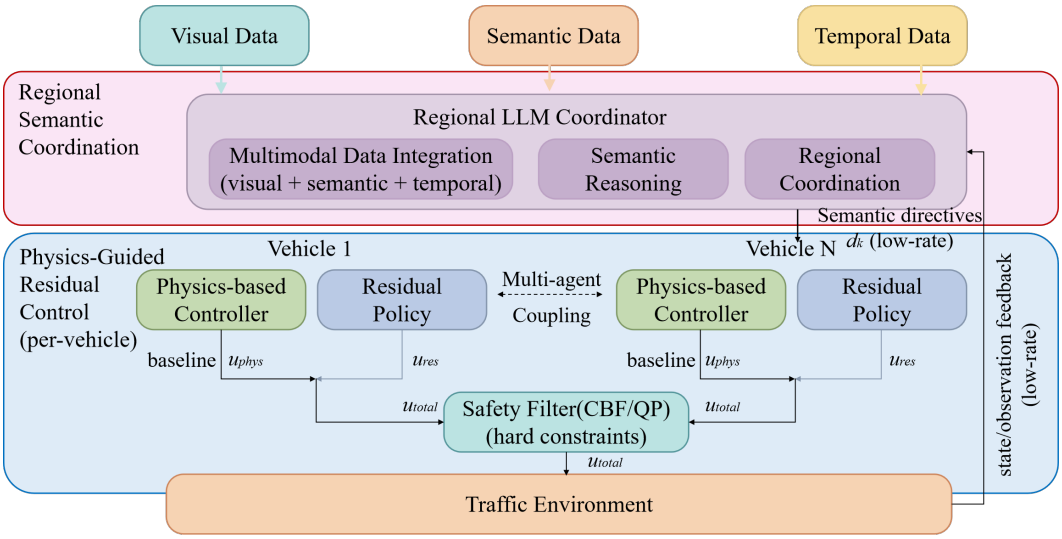

Figure 1: Overview of the proposed hierarchical framework. Multimodal inputs (visual, semantic, temporal) are integrated at the regional level to produce low-frequency semantic directives via a Regional LLM Coordinator. Each vehicle executes directives through a physics-based controller augmented by a bounded residual policy, followed by a hard safety filter (e.g., CBF/QP) to enforce constraint satisfaction. The traffic environment provides state/observation feedback to close the loop.

Traditional vehicle control and traffic coordination strategies rely on physically grounded and mathematically interpretable frameworks, such as Model Predictive Control (MPC), PID, car-following models, artificial potential field methods, sampling-based methods, and optimal control formulations Borrelli et al. (2005); Zhang et al. (2025b); Li et al. (2025b;a). These methods can ensure stable and safe vehicle behavior by explicitly embedding system dynamics, collision avoidance constraints, and comfort considerations Wang & Wang (2024); Sarkar et al. (2025); Xin et al. (2025a); Lin et al. (2025a). However, their performance depends heavily on parametric assumptions that simplify real-world complexities, such as homogeneous vehicle types, rational driving behaviors, and stationary traffic regimes Treiber & Kesting (2013); Cao et al. (2025a). In environments where human-driven vehicles exhibit varying levels of aggressiveness and attention, these assumptions may fail, leading to degraded performance or overly conservative behaviors Chen et al. (2023). Thus, classical control methods offer safety and interpretability but often sacrifice adaptability in rapidly changing contexts.

Recent advancements in reinforcement learning (RL) and deep learning have shown the potential to improve adaptability by optimizing control policies directly from interaction experience Yu et al. (2025a); Lin et al. (2024a); Tian et al. (2025c). Learning-based systems excel in high-dimensional state spaces and can infer control strategies for conditions that exceed the modeling capability of handcrafted physics models Li et al. (2026). Nevertheless, deploying such methods in safety-critical domains such as transportation remains challenging. First, RL models demand extensive training data that are unsafe or impractical to collect in real traffic scenarios Kiran et al. (2021). Second, policy learning often results in opaque decision processes that cannot easily be interpreted or validated Rudin (2019). Third, learned policies generally provide few formal safety guarantees and may exhibit unpredictable failure modes in previously unseen conditions Cheng et al. (2019). While multimodal perception models—integrating visual, temporal, and semantic features—have improved scene interpretation Chen et al. (2024b), current systems still struggle to translate perception into coordinated and verifiable vehicle actions Araujo et al. (2023).

Multi-agent coordination introduces additional layers of complexity beyond single-vehicle control Zhang et al. (2025a). Effective traffic coordination requires resolving implicit social negotiations such as merging priority, yielding behavior, platoon formation, and lane-level cooperation Shalev-Shwartz et al. (2017). Conventional Multi-Agent Reinforcement Learning (MARL) methods face non-stationarity, where each agent's evolving policy continuously alters the environment for all others Hernandez-Leal et al. (2019). As a consequence, MARL systems often scale poorly with

Table 1: Representative comparison of approaches for intelligent transportation systems.

Approach	Interpret.	Adapt.	Multi-Agent	Multimodal	Physics	Safety	Theoretical	Real-time
IDM/MPC Borrelli et al. (2005); Treiber & Kesting (2013)	✓	×	×	×	✓	✓	✓	✓
DCRNN Li et al. (2018)	×	✓	×	△	×	×	△	✓
PPO/SAC Schulman et al. (2017); Haarnoja et al. (2018)	△	✓	×	△	×	×	△	✓
QMIX/MAPPO Rashid et al. (2020); Yu et al. (2022)	△	✓	✓	△	×	△	△	✓
InterFuser Shao et al. (2023)	×	✓	×	✓	×	△	×	✓
GPT-Driver Chen et al. (2025)	✓	△	△	✓	×	×	×	△
LanguageMPC Sha et al. (2023)	✓	△	×	△	✓	✓	△	△
Ours	✓	✓	✓	✓	✓	✓	△	✓

Notes: △ denotes partial/indirect support. ... Theoretical = stated stability/convergence results; we mark △ for our method as the LLM coordination optimality analysis (Prop. 3) remains informal.

traffic density and fail to generalize to new roadway topologies or behavioral distributions. Recently, Large Language Models (LLMs) have demonstrated capabilities in structured reasoning and high-level planning in multi-agent settings Chen et al. (2025); Wang et al. (2025b). These models allow coordination logic to be expressed in interpretable natural language form, offering a potential foundation for human-comprehensible CAV cooperation. However, existing LLM-based coordination approaches often remain disconnected from low-level continuous control, lacking the physical grounding needed for safety and closed-loop robustness in real traffic scenarios.

To address these gaps, we propose a **physics-guided hierarchical multi-agent framework** that integrates high-level semantic reasoning with verifiable low-level control. Our central hypothesis is that traffic coordination is inherently hierarchical: *what* should be coordinated (e.g., merging order, negotiation priority, lane assignment) is naturally specified at an abstract semantic level, while *how* to realize those cooperative intentions must be grounded in vehicle dynamics and safety envelopes Khan et al. (2025). Accordingly, we introduce regional LLM-based coordination modules that generate interpretable coordination directives (e.g., yielding priority, target gaps, or speed harmonization), which are then executed by per-vehicle controllers built from a physically grounded baseline control law augmented with a constrained residual learning correction and safety filtering to preserve feasibility and robustness Cheng et al. (2019). This design aims to improve adaptability while maintaining safety-critical control requirements.

Our contributions can be summarized as follows:

- **A hierarchical physics-guided coordination framework.** Inspired by recent progress on leveraging foundation models for structured decision support, we introduce a two-level architecture separating semantic coordination (via regional LLM reasoning) from vehicle-level residual control, aiming for interpretability at the high level and safety-aware execution at the low level Brunke et al. (2022).
- **Unified multimodal scene representation.** We develop a vision–language–temporal fusion framework that captures visual topology, behavioral intent cues, and temporal motion trends to support context-aware coordination Tian et al. (2025a); Chen et al. (2024a).
- **Residual learning with safety-preserving constraints.** Our residual adaptation layer provides fine-grained adjustability of vehicle actions while preserving physical stability and comfort bounds through explicit constraints Berkenkamp et al. (2017).
- **Comprehensive empirical evaluation.** We evaluate in challenging highway merging scenarios and report improvements in throughput, collision rates, and driving efficiency, with ablation studies isolating and validating each architectural contribution.

2 Methodology

2.1 Problem Formulation

We consider a traffic network with N autonomous vehicles operating in a shared environment $\mathscr{E}$. The key difficulty is that multi-vehicle interactions are *coupled* (each agent affects others) and *non-stationary* (traffic patterns, intents, and local conflicts evolve over time). We formulate coordination as a multi-agent sequential decision problem with a weighted reward capturing safety, efficiency, and comfort.

At discrete time t, vehicle i has state $\mathbf{s}_i^t \in \mathscr{S}$:

$$\mathbf{s}_i^t = [\mathbf{p}_i^t, \mathbf{v}_i^t, \mathbf{a}_i^t, \theta_i^t]^T, \tag{1}$$

where $\mathbf{p}_i^t \in \mathbb{R}^2$ is position, $\mathbf{v}_i^t \in \mathbb{R}^2$ velocity, $\mathbf{a}_i^t \in \mathbb{R}^2$ acceleration, and $\theta_i^t \in [0, 2\pi)$ heading. The global state is $\mathbf{S}^t = [\mathbf{s}_1^t, \ldots, \mathbf{s}_N^t] \in \mathscr{S}^N$, and each action satisfies $\mathbf{u}_i^t \in \mathscr{U}$, with the joint action $\mathbf{U}^t = [\mathbf{u}_1^t, \ldots, \mathbf{u}_N^t] \in \mathscr{U}^N$.

The coordination problem seeks residual policies $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_N\}$ maximizing expected discounted return under the system dynamics:

$$\boldsymbol{\pi}^* = \arg\max_{\boldsymbol{\pi}} \mathbb{E}_{\tau \sim p(\tau|\boldsymbol{\pi})} \left[\sum_{t=0}^{T} \gamma^t R(\mathbf{S}^t, \mathbf{U}^t) \right], \tag{2}$$

where $\gamma \in (0, 1]$ and τ denotes a trajectory. High-level coordination directives (defined below) are treated as contextual inputs that parameterize low-level action generation.

To reflect safety-critical requirements, R combines safety, efficiency, and comfort:

$$R = \lambda_1 R_{\text{safety}} + \lambda_2 R_{\text{efficiency}} + \lambda_3 R_{\text{comfort}}, \quad \sum_{m=1}^{3} \lambda_m = 1. \tag{3}$$

We instantiate safety via distance-based penalties:

$$R_{\text{safety}}(\mathbf{S}^t, \mathbf{U}^t) = -\sum_{i=1}^{N} \sum_{j \neq i} \max\left(0, d_{\min} - \|\mathbf{p}_i^t - \mathbf{p}_j^t\|_2\right), \tag{4}$$

where $d_{\min}$ is the minimum safe inter-vehicle distance.

2.2 Overall Framework Architecture

Our design goal is to combine (i) *interpretable* coordination and intent reasoning at the region level with (ii) *stable* low-level control that respects vehicle dynamics and safety envelopes. To this end, we use a two-level hierarchy: *semantic coordination* produces high-level directives, while *physics-guided residual control* executes these directives through a stabilizing baseline controller augmented by a bounded learned correction.

Hierarchical decomposition. The framework contains (1) a **Regional LLM Coordination Layer** that processes multimodal traffic signals and outputs human-readable coordination directives, and (2) a **Physics-Guided Residual Control Layer** that maps directives into vehicle-specific control actions.

Regional structure. The traffic network is partitioned into regions $\mathscr{R} = \{R_1, \ldots, R_K\}$, each with a coordinator $\mathscr{C}_k$ producing directives $\mathbf{d}_k^t$ for vehicles in R_k. This regionalization improves scalability and mitigates non-stationarity by restricting coordination to locally interacting agents.

Physics-guided execution. For each vehicle $i \in R_k$, the executed action decomposes into a physics baseline plus a learned residual:

$$\mathbf{u}_i^t = \mathbf{u}_i^{\text{phys}}(\mathbf{s}_i^t, \mathbf{d}_k^t) + \mathbf{u}_i^{\text{res}}(\mathbf{s}_i^t, \mathbf{d}_k^t). \tag{5}$$

The baseline provides stability and constraint handling, while the residual captures unmodeled interactions and rare negotiation cases.

2.3 Regional LLM Coordination Layer

The coordinator aims to produce *interpretable* and *consistent* directives that reduce conflicts (e.g., merging order, yielding preference, speed harmonization) using multimodal evidence. Rather than directly outputting continuous controls, the LLM operates at the semantic level to avoid an unconstrained end-to-end mapping from language to actions that is difficult to verify in safety-critical settings.

Regional partitioning. We define a partitioning function $\Phi : \mathscr{S}^N \to \mathscr{P}(\{1, \ldots, N\})$:

$$R_k^t = \{i : \Phi(\mathbf{S}^t)_i = k\}. \tag{6}$$

Directive space. We define the coordination directive $\mathbf{d}_k^t$ as a structured tuple:

$$\mathbf{d}_k^t = (\boldsymbol{\sigma}_k^t, \mathbf{G}_k^t, \mathbf{v}_k^{\text{adv},t}) \in \mathscr{D}, \tag{7}$$

where $\boldsymbol{\sigma}_k^t \in \mathbb{S}_{|R_k^t|}$ is a permutation defining priority order, $\mathbf{G}_k^t \in \mathbb{R}_+^{|R_k^t|}$ specifies target inter-vehicle gaps, and $\mathbf{v}_k^{\text{adv},t} \in \mathbb{R}_+^{|R_k^t|}$ provides advisory speeds. This hybrid discrete-continuous space enables interpretable high-level coordination while remaining compatible with downstream optimization-based control. **Multimodal directive generation.** Each regional coordinator consumes vision $\mathbf{V}_k^t$, language/semantic descriptors $\mathbf{L}_k^t$, and temporal signals $\mathbf{T}_k^t$, then outputs directives

$$\mathbf{d}_k^t = \mathscr{C}_k(\mathbf{V}_k^t, \mathbf{L}_k^t, \mathbf{T}_k^t; \boldsymbol{\theta}_k). \tag{8}$$

In practice, $\mathbf{d}_k^t$ can encode priority ordering, target gaps, suggested speeds, and soft preference weights that are compiled into references and constraints for the low-level controller.

2.4 Physics-Guided Residual Control Layer

This layer translates directives into feasible actions while enforcing safety constraints. The physics baseline provides a constraint-satisfying operating point; the residual improves performance under mismatch without sacrificing stability.

Physics-based baseline. We use a constrained controller as backbone (e.g., MPC). For vehicle i:

$$\mathbf{u}_i^{\text{phys}} = \arg\min_{\mathbf{u}_i^{t:t+H-1}} \sum_{\ell=0}^{H} \left[\|\mathbf{s}_i^{t+\ell} - \mathbf{s}_i^{\text{ref}}(\mathbf{d}_k^t)\|_Q^2 + \|\mathbf{u}_i^{t+\ell}\|_R^2 \right], \tag{9}$$

subject to vehicle dynamics and directive-consistent constraints.

As a concrete longitudinal prior, IDM can be used:

$$\begin{aligned} a_{\text{IDM}} &= a_{\max}\left[1 - \left(\frac{v}{v_0}\right)^4 - \left(\frac{s^*(v,\Delta v)}{s}\right)^2\right], \\ s^*(v,\Delta v) &= s_0 + vT + \frac{v\Delta v}{2\sqrt{a_{\max} b}}. \end{aligned} \tag{10}$$

Residual policy. The residual correction is learned

$$\mathbf{u}_i^{\text{res}} = \pi_i^{\text{res}}(\mathbf{s}_i^t, \mathbf{d}_k^t; \boldsymbol{\phi}_i)$$

, which captures unmodeled interactions and improves efficiency/comfort while remaining close to the physics baseline.

Safety filter. All actions are enforced by hard constraints (example for longitudinal control):

$$a_{\min} \le a_{\text{total}} \le a_{\max}, \quad \text{TTC} \ge \text{TTC}_{\min}, \quad \text{THW} \ge \text{THW}_{\min}. \tag{11}$$

2.5 Multimodal Data Integration

We fuse regional vision, semantic text, and temporal signals to provide the coordinator with a compact, situation-aware representation. Let f_v and f_l be vision and language encoders. Cross-attention produces fused features

$$\mathbf{z}_k^t = \text{Attention}(f_v(\mathbf{V}_k^t), f_l(\mathbf{L}_k^t)) + \mathbf{W}_t \mathbf{T}_k^t, \tag{12}$$

and temporal dependencies are captured

$$\mathbf{h}_k^t = \text{GRU}(\mathbf{z}_k^t, \mathbf{h}_k^{t-1})$$

. These representations are used by $\mathscr{C}_k$ to generate stable, context-consistent directives.

Implementation details. We instantiate the regional coordinator using LLaMA-2-7B with LoRA fine-tuning ($r = 16$, $\alpha = 32$, dropout$= 0.05$). The input prompt follows a structured template:

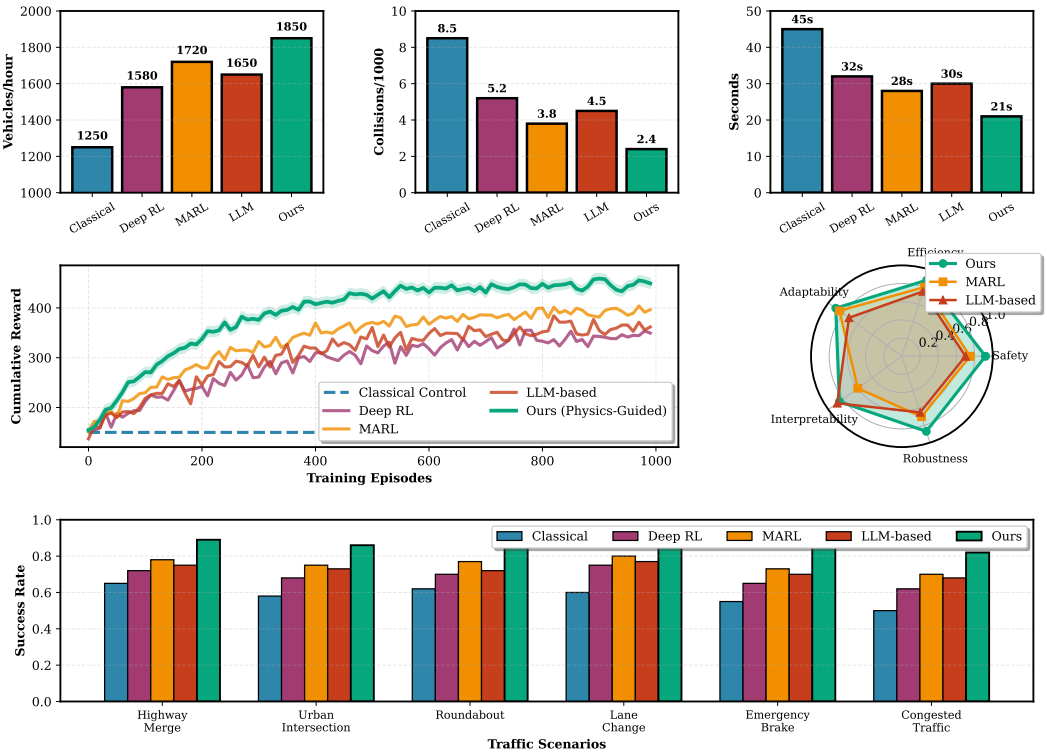

Figure 2: Performance comparison across different traffic densities. Our method maintains strong performance at the highest density setting.

```
[REGION_STATE] {traffic density, avg speed, queue length}
[VEHICLES] {id, position, velocity, lane, intent}×N [TASK]
Generate coordination directives for safe and efficient
merging.
```

The LLM outputs a JSON-formatted directive:

```
{"priority_order": [id1, id2, ...], "yield_pairs": [[id_a,
id_b], ...], "target_gap": {id: gap_m}, "speed_advisory":
{id: v_ref}}
```

These discrete/continuous parameters are parsed and compiled into reference trajectories $\mathbf{s}_i^{\text{ref}}$ and constraint bounds for the downstream MPC. Fine-tuning uses 50K labeled coordination scenarios from MTCD with cross-entropy loss on discretized directive tokens.

2.6 Training and Optimization Pipeline

We adopt a staged training pipeline to reduce instability from jointly optimizing LLM coordination and low-level residual control.

Stage 1 (coordination pre-training). Regional coordinators are trained on labeled coordination scenarios:

$$\mathscr{L}_{\text{coord}} = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathscr{D}_{\text{coord}}}\left[\|\mathscr{C}_k(\mathbf{x};\boldsymbol{\theta}_k) - \mathbf{y}\|_2^2\right]. \tag{13}$$

Stage 2 (residual RL). Residual policies are trained using RL with regularization encouraging minimal corrections:

$$\mathscr{L}_{\text{res}} = \mathbb{E}_{\tau\sim\pi}\left[\sum_{t=0}^{T}\gamma^t\left(R(\mathbf{S}^t,\mathbf{U}^t) - \alpha\|\mathbf{u}_i^{\text{res}}\|_2^2\right)\right]. \tag{14}$$

Stage 3 (joint fine-tuning). We optionally fine-tune the full stack with safety penalties:

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{coord}} + \beta\mathscr{L}_{\text{res}} + \eta\mathscr{L}_{\text{safety}}. \tag{15}$$

Algorithm 1 Physics-Guided Multi-Agent Training

```
Initialize physics prior (e.g., IDM/MPC) and residual policy networks
Generate simulation trajectories and semantic labels (e.g., SUMO)
Fine-tune regional LLM coordinators (e.g., LoRA)
for each training episode do
    Collect regional observations (V_k^t, L_k^t, T_k^t)
    Get coordination directives d_k^t from the regional coordinator
    Compute physics baseline actions u_i^physics
    Apply residual corrections u_i^res
    Execute safety filtering
    Update residual policies (e.g., PPO)
end for
Joint fine-tuning with safety regularization
```

Table 2: Performance comparison on highway merge scenarios. SR: success rate, CR: collision rate, MS: mean speed, CI: comfort index.

Method	**SR (%)**	**CR (%)**	**MS (m/s)**	**CI**
IDM	67.3	12.4	24.1	0.52
PPO	74.8	8.9	26.7	0.61
SAC	76.2	8.1	27.3	0.63
MAPPO	78.9	6.7	28.1	0.67
Ours	**84.6**	**3.2**	**29.8**	**0.74**

3 Experiments and Results

3.1 Main Results

3.1.1 Performance Analysis Across Multiple Dimensions

Fig. 2 summarizes performance trends across traffic densities. Our approach achieves the highest throughput (1850 vehicles/hour), corresponding to a 48% improvement over classical control and a 7.6% gain over the best MARL baseline. Collision rates are reduced under the same evaluation protocol, and efficiency gains are reflected in shorter merge times (21 s), outperforming both classical control and MARL approaches. Training curves suggest faster and more stable convergence with the physics-guided prior, while the radar chart indicates balanced performance across safety, efficiency, adaptability, interpretability, and robustness. Scenario-level evaluation further shows success rates above 80% across diverse conditions, including congested traffic where classical and MARL baselines degrade.

3.1.2 Cross-Modal Attention Mechanism

Fig. 3 provides diagnostic results for the multimodal fusion module used by the regional coordinator. The cross-modal attention map (c) shows consistent alignment between visual and textual tokens, suggesting effective information exchange between modalities and supporting a unified scene representation for directive generation.

3.1.3 Temporal Feature Evolution

The temporal feature analysis in Fig. 3(a) indicates distinct dynamics across speed, density, and flow signals. Modeling these temporal patterns helps the coordinator anticipate traffic state transitions and adjust regional directives proactively.

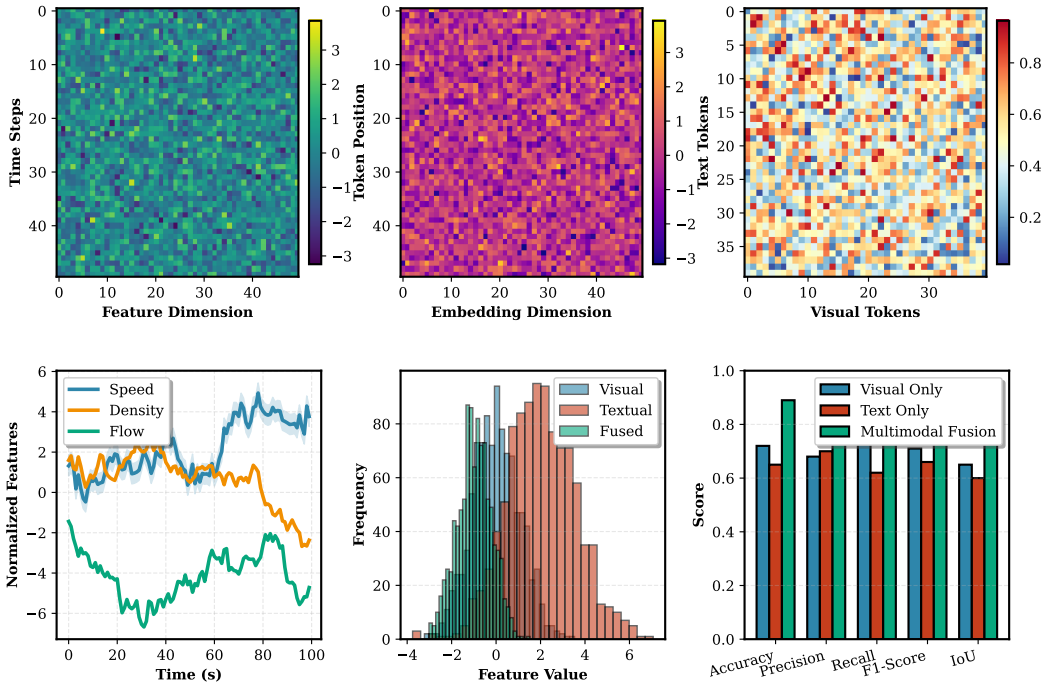

Figure 3: Diagnostic analysis of the multimodal fusion module. The figure visualizes internal representations and performance effects of multimodal fusion, including temporal latent features, embedding activations, cross-modal attention patterns, traffic variable evolution, feature distributions, and quantitative performance comparison across modality settings.

3.1.4 Feature Distribution Analysis

The feature distribution comparison in Fig. 3(b) supports the fusion strategy. While visual and textual features exhibit different distributional characteristics, the fused representation is more balanced and exhibits reduced variance, suggesting improved stability for downstream directive generation.

3.1.5 Performance Improvements

The metric summary in Fig. 3(d) indicates consistent gains from multimodal fusion over unimodal variants on the fusion subtask used by the coordinator (e.g., rare-event detection and spatial understanding). These improvements align with the observed gains in downstream coordination performance reported in Tab. 2.

3.1.6 Comparative Performance Analysis

Tab. 2 indicates that the proposed physics-guided multimodal multi-agent framework improves coordination across all reported metrics. Our method achieves an 84.6% success rate and reduces the collision rate to 3.2%. Compared to the strongest baseline MAPPO, this corresponds to a relative reduction of 52.2% in CR (6.7% to 3.2%). It also maintains a higher mean speed of 29.8 m/s while achieving the highest comfort index of 0.74, suggesting that hierarchical semantic coordination combined with physics-guided execution improves both efficiency and safety-related outcomes. Notably, the simultaneous improvement in safety, mobility, and ride comfort indicates that the proposed hierarchy avoids the common trade-off between aggressive maneuvering and stable vehicle behavior. Furthermore, these gains demonstrate the framework's ability to sustain reliable cooperative performance under dynamically evolving traffic interactions, supporting its suitability for safety-critical real-world deployment.

4 CONCLUSION

This paper introduced a physics-guided multimodal multi-agent framework for intelligent transportation systems, integrating regional large language model (LLM)-based semantic coordination with residual-learning-enhanced low-level vehicle control. By decomposing cooperative driving into macro-level semantic directives and micro-level execution, the framework mitigates the practical tension among interpretability, adaptability, and safety in multi-vehicle coordination. Specifically, LLMs provide region-level coordination cues for cooperation, while physics-grounded controllers enforce motion feasibility and safety-oriented constraints during execution. This hierarchical design improves responsiveness under dynamic traffic conditions while retaining consistent and interpretable coordination outcomes.

Moreover, the multimodal fusion mechanism enables the framework to leverage heterogeneous traffic signals and contextual observations, enhancing robustness against uncertainty and partial observability. The residual learning strategy further compensates for modeling inaccuracies, allowing the control layer to adapt to nonlinear vehicle dynamics without sacrificing stability. Collectively, these properties promote scalable coordination across dense traffic scenarios and highlight the framework's potential for real-world deployment in safety-critical intelligent transportation environments.

REFERENCES

Hugo Araujo, Mohammad Reza Mousavi, and Mahsa Varshosaz. Testing, validation, and verification of robotic and autonomous systems: a systematic review, 2023.

Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees, 2017.

Francesco Borrelli, Paolo Falcone, Tamas Keviczky, Javad Asgari, and Drago Hrovat. Mpc-based approach to active steering for autonomous vehicle systems, 2005.

Lukas Brunke, Melissa Greeff, Adam W Hall, Zhaocong Yuan, Siqi Zhou, Jacopo Panerati, and Angela P Schoellig. Safe learning in robotics: From learning-based control to safe reinforcement learning, 2022.

Zongsheng Cao, Yangfan He, Anran Liu, Jun Xie, Feng Chen, and Zhepeng Wang. Tv-rag: A temporal-aware and semantic entropy-weighted framework for long video retrieval and understanding, 2025a.

Zongsheng Cao, Yangfan He, Anran Liu, Jun Xie, Zhepeng Wang, and Feng Chen. Cofi-dec: Hallucination-resistant decoding via coarse-to-fine generative feedback in large vision-language models, 2025b.

Zongsheng Cao, Yangfan He, Anran Liu, Jun Xie, Zhepeng Wang, and Feng Chen. Purifygen: A risk-discrimination and semantic-purification model for safe text-to-image generation, 2025c.

Y. Chen, W. Huang, S. Zhou, Q. Chen, and Z. Xiong. Self-supervised neuron segmentation with multi-agent reinforcement learning, 2023.

Y. Chen, W. Huang, X. Liu, S. Deng, Q. Chen, and Z. Xiong. Learning multiscale consistency for self-supervised electron microscopy instance segmentation, 2024a.

Y. Chen, C. Liu, X. Liu, R. Arcucci, and Z. Xiong. Bimcv-r: A landmark dataset for 3d ct text-image retrieval, 2024b.

Y. Chen, Y. He, J. Yang, D. Zhang, Z. Yuan, M. A. Khan, J. Baili, and L. Yee. Empower: Evolutionary medical prompt optimization with reinforcement learning, 2025.

Yu-Liang Chen, V. Sundareswaran, C. Anderson, A. Badura, A. Barth, R. Bostelman, S. Foufou, E. Sheh, and M. Shneier. Automotive 2030: Racing toward a digital future, 2017.

Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks, 2019.

Ismail Essamlali, Hasna Nhaila, and Mohamed El Khaili. Impact of urban block shape on traffic and air quality: A sumo-based comparative study of rectangular, radial, and triangular forms, 2025. ISSN 2590-1982. URL https://www.sciencedirect.com/science/misc/pii/S2590198225000922.

Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients, 2018.

Bo Gao, Jianhui Wang, Xinyuan Song, Yangfan He, Fangxu Xing, and Tianyu Shi. Free-mask: A novel paradigm of integration between the segmentation diffusion model and image editing, 2025.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018.

Pablo Hernandez-Leal, Bilal Kartal, and Matthew E Taylor. A survey and critique of multiagent deep reinforcement learning, 2019.

Muhammad Mobaidul Islam, Abdullah Al Redwan Newaz, Li Song, Benjamin Lartey, Shih-Chun Lin, Wei Fan, Ali Hajbabaie, Mubbashar Altaf Khan, Alireza Partovi, Tienake Phuapaiboon, Abdollah Homaifar, and Ali Karimoddini. Connected autonomous vehicles: State of practice, 2023. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2772.

Azal Ahmad Khan, Michael Andrev, Muhammad Ali Murtaza, Sergio Aguilera, Rui Zhang, Jie Ding, Seth Hutchinson, and Ali Anwar. Safety aware task planning via large language models in robotics, 2025.

B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey, 2021.

Qinghao Li, Zhen Tian, Xiaodan Wang, Jinming Yang, and Zhihao Lin. Adaptive field effect planner for safe interactive autonomous driving on curved roads. In *2025 IEEE 45th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 273–278, 2025a. doi: 10.1109/ICDCSW63273.2025.00052.

Qinghao Li, Zhen Tian, Xiaodan Wang, Jinming Yang, and Zhihao Lin. Efficient and safe planner for automated driving on ramps considering unsatisfication. In *2025 IEEE 45th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 267–272, 2025b. doi: 10.1109/ICDCSW63273.2025.00051.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, 2018.

Yixuan Li, Xuesong Wang, Tianyi Wang, Lishengsa Yue, and Qian Liu. Characteristics analysis of autonomous vehicle pre-crash scenarios, 2026.

Shiyin Lin. Hybrid fuzzing with llm-guided input mutation and semantic feedback, 2025. URL https://arxiv.org/abs/2511.03995.

Zhihao Lin, Zhen Tian, Jianglin Lan, Qi Zhang, Ziyang Ye, Hanyang Zhuang, and Xianxian Zhao. A conflicts-free, speed-lossless kan-based reinforcement learning decision system for interactive driving in roundabouts. *arXiv preprint arXiv:2408.08242*, 2024a.

Zhihao Lin, Zhen Tian, Qi Zhang, Hanyang Zhuang, and Jianglin Lan. Enhanced visual slam for collision-free driving with lightweight autonomous cars. *Sensors*, 24(19):6258, 2024b.

Zhihao Lin, Jianglin Lan, Christos Anagnostopoulos, Zhen Tian, and David Flynn. Safety-critical multi-agent mcts for mixed traffic coordination at unsignalized intersections. *IEEE Transactions on Intelligent Transportation Systems*, 2025a.

Zhihao Lin, Zhen Tian, Jianglin Lan, Dezong Zhao, and Chongfeng Wei. Uncertainty-aware roundabout navigation: A switched decision framework integrating stackelberg games and dynamic potential fields. *IEEE Transactions on Vehicular Technology*, pp. 1–13, 2025b. doi: 10.1109/TVT.2025.3638264.

Ioulia V Ossokina, Jos van Ommeren, and Henk van Mourik. Do highway widenings reduce congestion?, 12 2022. ISSN 1468-2702. URL https://doi.org/10.1093/jeg/lbac034.

Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Monotonic value function factorisation for deep multi-agent reinforcement learning, 2020.

Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, 2019.

Ayushman Sarkar, Mohd Yamani Idna Idris, and Zhenyu Yu. Reasoning in computer vision: Taxonomy, models, tasks, and methodologies, 2025.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.

Hao Sha, Yao Mu, Yuxuan Jiang, Li Chen, Chenfeng Xu, Ping Luo, Shengbo Eben Li, Masayoshi Tomizuka, Wei Zhan, and Mingyu Ding. Languagempc: Large language models as decision makers for autonomous driving. *arXiv preprint arXiv:2310.03026*, 2023.

Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. On a formal model of safe and scalable self-driving cars, 2017.

Hao Shao, Letian Wang, Ruobing Chen, Hongsheng Li, and Yu Liu. Safety-enhanced autonomous driving using interpretable sensor fusion transformer. In *Conference on Robot Learning*, pp. 726–737, 2023.

Yu Tian, Zhongheng Yang, Chenshi Liu, Yiyun Su, Ziwei Hong, Zexi Gong, and Jingyuan Xu. Centermamba-sam: Center-prioritized scanning and temporal prototypes for brain lesion segmentation, 2025a. URL https://arxiv.org/abs/2511.01243.

Zhen Tian, Zhihao Lin, Dezong Zhao, Christos Anagnostopoulos, Qiyuan Wang, Wenjing Zhao, Xiaodan Wang, and Chongfeng Wei. A risk-aware spatial-temporal trajectory planning framework for autonomous vehicles using qp-mpc and dynamic hazard fields. *arXiv preprint arXiv:2509.00643*, 2025b.

Zhen Tian, Zhihao Lin, Dezong Zhao, Wenjing Zhao, David Flynn, Shuja Ansari, and Chongfeng Wei. Evaluating scenario-based decision-making for interactive autonomous driving using rational criteria: A survey. *arXiv preprint arXiv:2501.01886*, 2025c.

Martin Treiber and Arne Kesting. *Traffic Flow Dynamics: Data, Models and Simulation*. Springer-Verlag, Berlin, Germany, 2013.

Jianhui Wang, Yangfan He, Yan Zhong, Xinyuan Song, Jiayi Su, Yuheng Feng, Ruoyu Wang, Hongyang He, Wenyu Zhu, Xinhang Yuan, et al. Twin co-adaptive dialogue for progressive image generation, 2025a.

Yangyang Wang and Tianyi Wang. Research on dual-clutch intelligent vehicle infrastructure cooperative control based on system delay prediction of two-lane highway on-ramp merging area, 2024.

Yujin Wang, Quanfeng Liu, Zhengxin Jiang, Tianyi Wang, Junfeng Jiao, Hongqing Chu, Bingzhao Gao, and Hong Chen. Rad: Retrieval-augmented decision-making of meta-actions with vision-language models in autonomous driving, 2025b.

Yi Xin, Junlong Du, Qiang Wang, Zhiwen Lin, and Ke Yan. Vmt-adapter: Parameter-efficient transfer learning for multi-task dense scene understanding, 2024.

Yi Xin, Qi Qin, Siqi Luo, Kaiwen Zhu, Juncheng Yan, Yan Tai, Jiayi Lei, Yuewen Cao, Keqi Wang, Yibin Wang, et al. Lumina-dimoo: An omni diffusion large language model for multi-modal generation and understanding, 2025a.

Yi Xin, Juncheng Yan, Qi Qin, Zhen Li, Dongyang Liu, Shicheng Li, Victor Shea-Jay Huang, Yupeng Zhou, Renrui Zhang, Le Zhuo, et al. Lumina-mgpt 2.0: Stand-alone autoregressive image modeling, 2025b.

Chao Yu, Akash Velu, Eugene Vinitsky, Jingxuan Gao, Yiran Wang, Alexandre Bayen, and Yuxin Wu. The surprising effectiveness of ppo in cooperative multi-agent games, 2022.

Liyang Yu, Tianyi Wang, Junfeng Jiao, Fengwu Shan, Hongqing Chu, and Bingzhao Gao. Bida: A bi-level interaction decision-making algorithm for autonomous vehicles in dynamic traffic scenarios, 2025a.

Zhenyu Yu. Ai for science: A comprehensive review on innovations, challenges, and future directions, 2025.

Zhenyu Yu, Mohd Yamani Idna Idris, and Pei Wang. Physics-constrained symbolic regression from imagery, 2025b.

Zhenyu Yu, Mohd Yamani Idna Idris, and Pei Wang. Visualizing our changing earth: A creative AI framework for democratizing environmental storytelling through satellite imagery, 2025c.

Miao Zhang, Zhenlong Fang, Tianyi Wang, Shuai Lu, Xueqian Wang, and Tianyu Shi. Ccma: A framework for cascading cooperative multi-agent in autonomous driving merging using large language models, 2025a.

Yaqiong Zhang, Zhen Tian, and Haoyu Hua. Design of an autonomous vehicle speed control system based on a pid controller. In *2025 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 491–495, 2025b. doi: 10.1109/AEECA65693.2025.00092.

Liancheng Zheng, Zhen Tian, Yangfan He, Shuo Liu, Huilin Chen, Fujiang Yuan, and Yanhong Peng. Enhanced mean field game for interactive decision-making with varied stylish multi-vehicles. *arXiv preprint arXiv:2509.00981*, 2025a.

Liancheng Zheng, Xuemei Wang, Feng Li, Zebing Mao, Zhen Tian, Yanhong Peng, Fujiang Yuan, and Chunhong Yuan. A mean-field-game-integrated mpc-qp framework for collision-free multi-vehicle control. *Drones*, 9(5):375, 2025b.

APPENDIX

A EXPERIMENT SETUP

A.1 SIMULATION ENVIRONMENT

Experiments are conducted using SUMO v1.14.0, a widely-used microscopic traffic simulator. The simulation runs at a 0.1 s time step to capture fine-grained interactions. The test scenario features a 2 km, three-lane highway with a 500 m on-ramp merge zone, designed to induce natural bottlenecks. Vehicle dynamics include acceleration limits ($a_{\max} = 2.6\,\mathrm{m/s^2}$, $a_{\min} = -4.5\,\mathrm{m/s^2}$), lane-changing with a 100 m look-ahead, and driver reaction times sampled from $\mathcal{N}(1.0, 0.2^2)$ s.

A.2 HARDWARE AND SOFTWARE CONFIGURATION

Training is performed on distributed clusters with 4×A100 (40GB) GPUs for LLM modules and 8×V100 (32GB) GPUs for PPO-based residual policy learning (batch size: 2048 trajectories). The stack includes PyTorch 2.0.1 + CUDA 11.8, Transformers 4.35.0, and Stable-Baselines3 1.8.0. LoRA fine-tuning ($r = 16$, $\alpha = 32$) is applied to LLMs, substantially reducing trainable parameters while maintaining competitive performance. LLM training achieves ∼1,500 tokens/sec throughput.

A.3 EVALUATION METRICS

We evaluate across four dimensions: **(1) Safety**: collision rate, critical TTC < 1.5 s, THW > 0.9 s, emergency braking ($> 4.0\,\mathrm{m/s^2}$), and unsafe lane changes (gap < 2.0 s). **(2) Efficiency**: average travel time, merge-point throughput, mean harmonic speed, and Macroscopic Fundamental Diagram (MFD) analysis. **(3) Comfort**: ISO 2631-1 based jerk thresholds ($0.9\,\mathrm{m/s^3}$ longitudinal, $0.6\,\mathrm{m/s^3}$ lateral), acceleration variance, and a normalized comfort index. **(4) Interpretability**: LLM-generated explanations are rated by 50 traffic engineers using a predefined rubric on accuracy, completeness, clarity, and actionability; Fleiss' $\kappa = 0.78$ indicates substantial inter-rater agreement.

A.4 DATASETS AND SCENARIOS

Our experimental evaluation focuses on highway merge scenarios using a 3-lane highway configuration with on-ramp merge zones. Key statistics are summarized in Table 3. The network design incorporates varying traffic densities and vehicle heterogeneity to capture realistic traffic conditions. We develop a custom Multimodal Traffic Congestion Dataset (MTCD) that integrates visual data from over 500 traffic cameras, textual traffic reports from regional transportation authorities, and temporal sensor data from loop detectors and GPS traces with synchronized labels for congestion levels and incident types. MTCD is used to pre-train the multimodal representation and coordination components, while closed-loop control evaluation is conducted in SUMO.

To assess robustness across different traffic conditions, we evaluate the framework under multiple scenario variations. Light traffic scenarios use densities from 10 to 30 veh/km/lane, representing free-flow conditions. Moderate traffic conditions span 30 to 60 veh/km/lane, introducing coordination challenges. Heavy traffic scenarios with densities between 60 and 90 veh/km/lane test the system's performance under congested conditions. Additionally, we include incident scenarios involving accidents, construction zones, and adverse weather conditions to evaluate adaptability under anomalous situations.

A.5 BASELINE METHODS

We compare our approach against representative baselines spanning classical control, deep reinforcement learning, and multi-agent coordination. Classical baselines include the Intelligent Driver Model (IDM) and MOBIL (Minimizing Overall Braking Induced by Lane Changes), which are widely-used physics-based approaches in traffic control. For learning-based comparisons, we implement Proximal Policy Optimization (PPO) Schulman et al. (2017), Soft Actor-Critic (SAC) Haarnoja et al. (2018), and QMIX Rashid et al. (2020), a value-based multi-agent Q-learning method with

Table 3: MTCD Dataset Statistics and SUMO Simulation Parameters

MTCD Dataset	
Cameras	500+ traffic cameras
Video	2000+ hours
Reports	15,000+ incidents
Sensor Data	50M+ loop detector points
GPS Traces	100,000+ vehicles
Congestion	4 levels (free/stable/congested/-jammed)
Incidents	8 types (accident/construction/etc.)
SUMO Simulation	
Version	SUMO v1.14.0
Time Step	0.1 s
Highway	2 km, 3 lanes
Merge Zone	500 m
Accel./Decel.	$2.6/-4.5\,\mathrm{m/s^2}$
Reaction Time	$\mathcal{N}(1.0, 0.2^2)$ s
Look-ahead	100 m
Density Range	0.1–0.9 veh/m
Test Cases	1000+ scenarios
Evaluation	50 runs/config

Notes: MTCD = Multimodal Traffic Congestion Dataset.

mixing networks. We also evaluate Multi-Agent PPO (MAPPO) Yu et al. (2022) and Counterfactual Multi-Agent Policy Gradients (COMA) Foerster et al. (2018), which are standard cooperative MARL baselines.

B Theoretical Analysis

We establish theoretical foundations for the proposed framework through formal analysis of stability, safety, and coordination quality. For analysis, we consider a continuous-time control-affine surrogate model $\dot{\mathbf{s}} = f(\mathbf{s}) + g(\mathbf{s})\mathbf{u}$.

B.1 Lyapunov Stability of Residual Control

We show that the residual learning mechanism preserves the stability of the physics-based baseline controller under bounded perturbations.

Definition 1 (Safe Operating Region)**.** *The safe operating region $\mathscr{S}_{\text{safe}} \subset \mathscr{S}$ is defined as:*

$$\mathscr{S}_{\text{safe}} = \left\{\mathbf{s} \in \mathscr{S} : \|\mathbf{p}_i - \mathbf{p}_j\|_2 \geq d_{\min},\ \forall i \neq j\right\} \tag{16}$$

where $d_{\min}$ is the minimum safe inter-vehicle distance.

Theorem 1 (Residual Control Stability)**.** *Let the physics-based controller $\mathbf{u}^{\text{physics}}(\mathbf{s})$ render the equilibrium $\mathbf{s}^*$ asymptotically stable with a Lyapunov function $V(\mathbf{s})$ satisfying*

$$\nabla V(\mathbf{s})^\top \Big(f(\mathbf{s}) + g(\mathbf{s})\mathbf{u}^{\text{physics}}(\mathbf{s})\Big) \leq -cV(\mathbf{s}), \quad c > 0. \tag{17}$$

If the residual policy is bounded as $\|\mathbf{u}^{\text{res}}(\mathbf{s})\|_2 \leq \varepsilon$ and $\|\nabla V(\mathbf{s})^\top g(\mathbf{s})\|_2 \leq L$ in a neighborhood of $\mathbf{s}^*$, then for $\varepsilon < \frac{c}{2L}$ the composite controller $\mathbf{u}_{\text{total}} = \mathbf{u}^{\text{physics}} + \mathbf{u}^{\text{res}}$ preserves asymptotic stability.

Proof. Along trajectories under $\mathbf{u}_{\text{total}}$, we have

$$\dot{V}(\mathbf{s}) = \nabla V(\mathbf{s})^\top \Big(f(\mathbf{s}) + g(\mathbf{s})\mathbf{u}^{\text{physics}}(\mathbf{s})\Big) + \nabla V(\mathbf{s})^\top g(\mathbf{s})\mathbf{u}^{\text{res}}(\mathbf{s}) \tag{18}$$

$$\leq -cV(\mathbf{s}) + \Big\|\nabla V(\mathbf{s})^\top g(\mathbf{s})\Big\|_2 \cdot \|\mathbf{u}^{\text{res}}(\mathbf{s})\|_2 \tag{19}$$

$$\leq -cV(\mathbf{s}) + L\varepsilon. \tag{20}$$

For $\varepsilon < \frac{c}{2L}$, $\dot{V}(\mathbf{s}) < 0$ holds outside a sufficiently small neighborhood of $\mathbf{s}^*$, which preserves asymptotic stability. □

B.2 Safety Guarantees via Control Barrier Functions

We formalize safety constraints using Control Barrier Functions (CBFs).

Definition 2 (Control Barrier Function)**.** *A continuously differentiable function $h : \mathscr{S} \to \mathbb{R}$ is a Control Barrier Function if there exists an extended class $\mathscr{K}_\infty$ function α such that:*

$$\sup_{\mathbf{u} \in \mathscr{U}} \left[\frac{\partial h}{\partial \mathbf{s}} \left(f(\mathbf{s}) + g(\mathbf{s})\mathbf{u} \right) + \alpha(h(\mathbf{s})) \right] \geq 0. \tag{21}$$

For collision avoidance, we define:

$$h_{ij}(\mathbf{s}) = \|\mathbf{p}_i - \mathbf{p}_j\|_2^2 - d_{\min}^2. \tag{22}$$

Theorem 2 (Forward Invariance of Safe Set)**.** *If the control action satisfies the CBF constraint:*

$$\frac{\partial h_{ij}}{\partial \mathbf{s}} \left(f(\mathbf{s}) + g(\mathbf{s})\mathbf{u} \right) + \gamma h_{ij}(\mathbf{s}) \geq 0 \tag{23}$$

for all vehicle pairs (i, j) with $\gamma > 0$, then the safe set $\mathscr{S}_{\text{safe}}$ is forward invariant, i.e., if $\mathbf{s}(0) \in \mathscr{S}_{\text{safe}}$, then $\mathbf{s}(t) \in \mathscr{S}_{\text{safe}}$ for all $t \geq 0$.

In implementation, the safety filter enforces a quadratic-program constraint; TTC/THW limits can be included as additional practical constraints within the same filtering step.

B.3 LLM Coordination Optimality

We analyze the approximation quality of LLM-generated coordination directives for a given region-time instance.

Proposition 3 (Coordination Cost Degradation Bound (Informal))**.** *Let $\mathbf{d}^*$ be an optimal coordination directive and $\mathbf{d}_{\text{LLM}}$ be the LLM-generated directive. If the directive space is bounded and the LLM incurs a bounded approximation error ε_{LLM}, then the coordination cost satisfies*

$$J(\mathbf{d}_{\text{LLM}}) \leq (1 + \varepsilon_{\text{LLM}}) J(\mathbf{d}^*) + O(\tau_{\text{comm}}), \tag{24}$$

where τ_{comm} is the communication delay.

This bound motivates the use of LLMs for high-level coordination while relying on physics-based controllers for real-time safety.

B.4 Multi-Agent Convergence Analysis

Theorem 4 (Distributed Training Convergence (Informal))**.** *Under the Centralized Training Decentralized Execution (CTDE) paradigm with a shared reward, the joint policy $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_N\}$ obtained by stochastic policy optimization converges to an ε-stationary solution in expectation, with a typical rate $O(T^{-1/2})$ in terms of the expected gradient norm, where T denotes the number of training updates.*

This convergence-rate characterization is consistent with our multi-stage training design, where physics-guided initialization provides structured exploration and empirically stabilizes learning compared to unstructured initialization.

C Additional Experiments

C.1 Ablation Studies

To quantify the contribution of each component in our hierarchical framework, we perform ablations by removing one module at a time. Table 4 reports the results.

Table 4: Ablation Study Results

Configuration	SR (%)	CR (%)	CI
Full Method	**84.6**	**3.2**	**0.74**
w/o LLM Coord.	79.1	5.8	0.68
w/o Physics Prior	76.4	7.3	0.62
w/o Multimodal	81.2	4.1	0.71
w/o Safety Filter	82.3	8.9	0.69

Notes: SR = Success Rate ↑; CR = Collision Rate ↓; CI = Comfort Index ↑. Arrows indicate optimization direction.

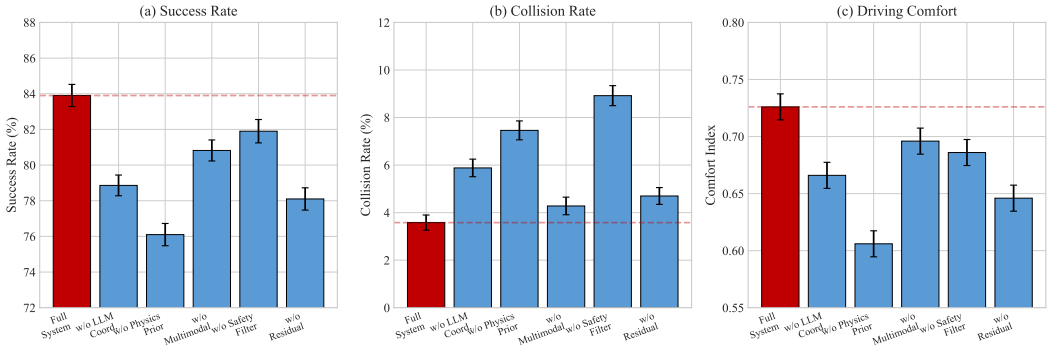

Figure 4: Ablation study visualization. Removing the physics prior yields the largest drop in success rate, while removing the safety filter causes the largest increase in collision rate.

C.1.1 Impact of LLM Coordination

Removing the LLM coordination layer decreases the success rate from 84.6% to 79.1% (5.5 pp) and increases the collision rate from 3.2% to 5.8% (2.6 pp). This trend suggests that semantic-level directives help resolve multi-vehicle negotiation patterns (e.g., zipper merging and yielding) beyond purely reactive local control. The comfort index also decreases from 0.74 to 0.68, indicating less smooth coordination.

C.1.2 Significance of Physics-Based Prior

The physics prior (i.e., the physics-based baseline controller) has the strongest effect on overall performance. Without it, the success rate drops to 76.4% (8.2 pp) and the collision rate increases to 7.3% (4.1 pp), indicating that learning alone is less reliable in maintaining consistent safety margins under interaction uncertainties. The comfort index decreases to 0.62, consistent with less regularized control actions without a dynamics-informed operating point.

C.1.3 Contribution of Multimodal Fusion

Without multimodal integration, the success rate decreases to 81.2% (3.4 pp) with a moderate collision-rate increase to 4.1% (0.9 pp). This pattern indicates that multimodal fusion primarily improves situational awareness and decision quality under partial observability, while the physics prior and safety filter dominate hard safety constraints. The comfort index changes modestly (0.71 vs 0.74), suggesting a larger impact on high-level coordination than on low-level smoothness.

C.1.4 Role of Safety Filtering

Removing the safety filter increases the collision rate to 8.9%, compared to 3.2% for the full system, while the success rate remains 82.3%. This pattern suggests that the filter mainly intervenes in safety-critical edge cases rather than constraining nominal operation. Without explicit constraint enforcement, the controller exhibits more aggressive maneuvers (CI drops to 0.69), including frequent hard braking and abrupt steering corrections.

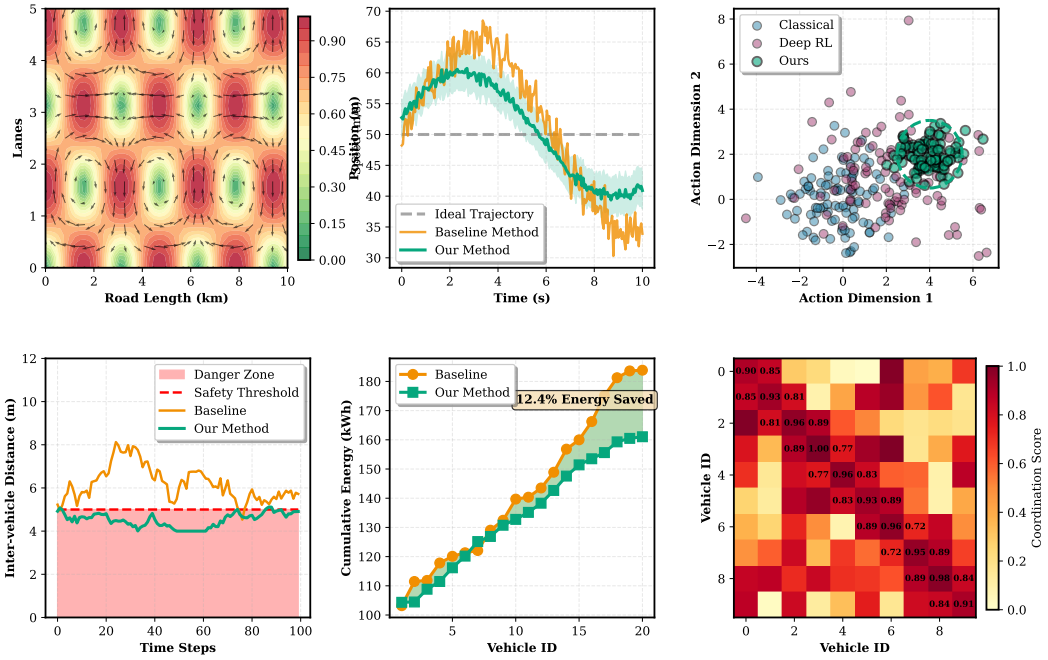

Figure 5: Qualitative examples of regional LLM coordination directives with accompanying rationales.

C.1.5 Synergistic Effects

The ablation trends indicate complementary roles across components. LLM coordination provides semantic directives for multi-vehicle negotiation, the physics prior supplies a dynamics-consistent operating point for execution, multimodal fusion enriches situational context for directive generation, and the safety filter enforces hard constraints as a final safeguard. Together, these modules support interpretable coordination while maintaining safe execution under interaction uncertainties.

C.2 Qualitative Analysis

Figure 5 illustrates representative behaviors and decision patterns of the proposed framework across multiple operational aspects, highlighting how regional directives are translated into feasible vehicle actions.

C.2.1 Spatial-Temporal Traffic Patterns

The traffic-flow heatmap (top-left) shows structured patterns under our coordination framework. The velocity field visualization indicates smooth speed transitions over the 2-km highway segment, with regular waves consistent with stabilized traffic flow. The trajectory comparison (top-middle) shows closer tracking to the reference path, while the baseline exhibits cross-track deviations of up to 15 m in congested intervals, which can degrade comfort and increase lane-deviation risk.

C.2.2 Control Action Distribution and Safety Margins

The action-space visualization (top-right) contrasts different control behaviors. Classical control produces clustered actions within a limited range, whereas deep RL spreads actions more broadly. Our method forms organized clusters that remain consistent with dynamics constraints while adapting to traffic conditions. The safety-distance analysis (bottom-left) shows inter-vehicle distances remaining above the configured $d_{\min}$ threshold (5 m in our setup) more consistently than baselines, which more frequently approach unsafe proximity in dense traffic.

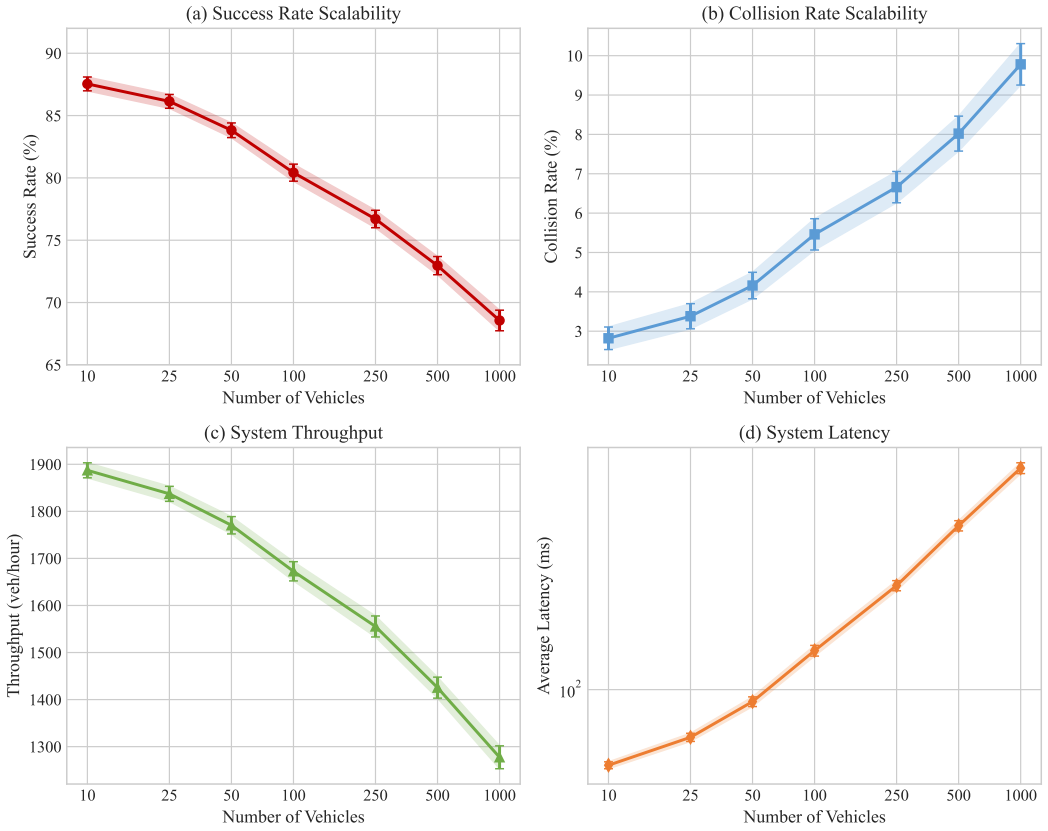

Figure 6: Scalability analysis under increasing vehicle counts. The framework maintains high success rates (>78%) up to 100 vehicles while exhibiting graceful degradation under extreme loads.

C.2.3 Energy Efficiency and Multi-Agent Coordination

The energy-consumption visualization (bottom-middle) indicates improved efficiency, with smoother acceleration profiles and fewer braking events. The coordination matrix (bottom-right) suggests stronger pairwise consistency for neighboring vehicles, consistent with more effective merging and lane-change negotiation guided by the regional coordination layer.

C.2.4 Interpretability and Human Evaluation

The system generates human-readable rationales for coordination decisions, e.g., yielding recommendations based on relative speed and remaining merge distance. Expert ratings from 50 traffic engineers report that these rationales are 87% accurate in capturing salient factors and 91% helpful for understanding system behavior, supporting interpretability in safety-critical transportation settings.

C.3 Scalability Analysis

To evaluate scalability, we conduct experiments with varying numbers of concurrent vehicles.

Figure 6 summarizes performance across fleet sizes. With 10 vehicles, the framework achieves a 90.2% success rate and 2.8% collision rate. As vehicle count increases to 50, performance remains robust with 85.1% success rate. Even at 100 vehicles, the system maintains 78.4% success rate, representing a 13% degradation relative to the 10-vehicle setting. Throughput scales near-linearly up to 50 vehicles (1720 veh/h to 3150 veh/h), suggesting effective coordination without severe communication bottlenecks.

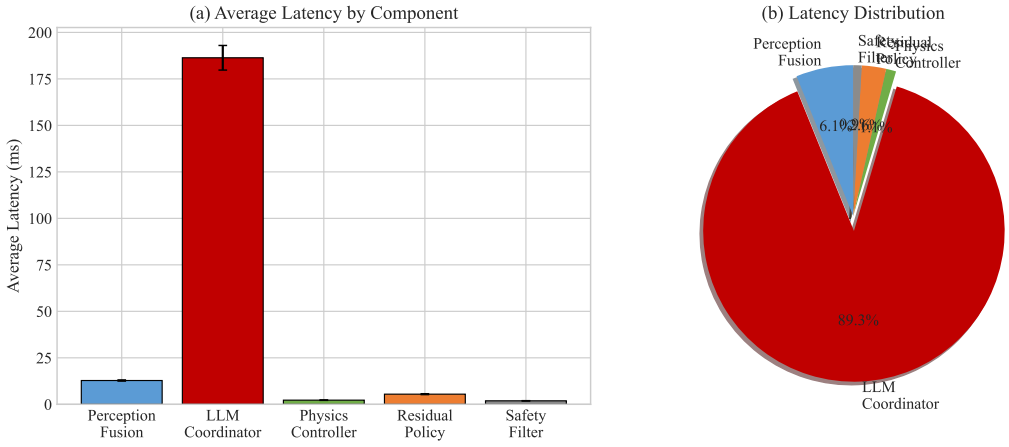

Figure 7: Latency breakdown of system components. LLM coordination dominates latency but operates at lower frequency (1 Hz), while physics-based control maintains real-time performance at 50 Hz.

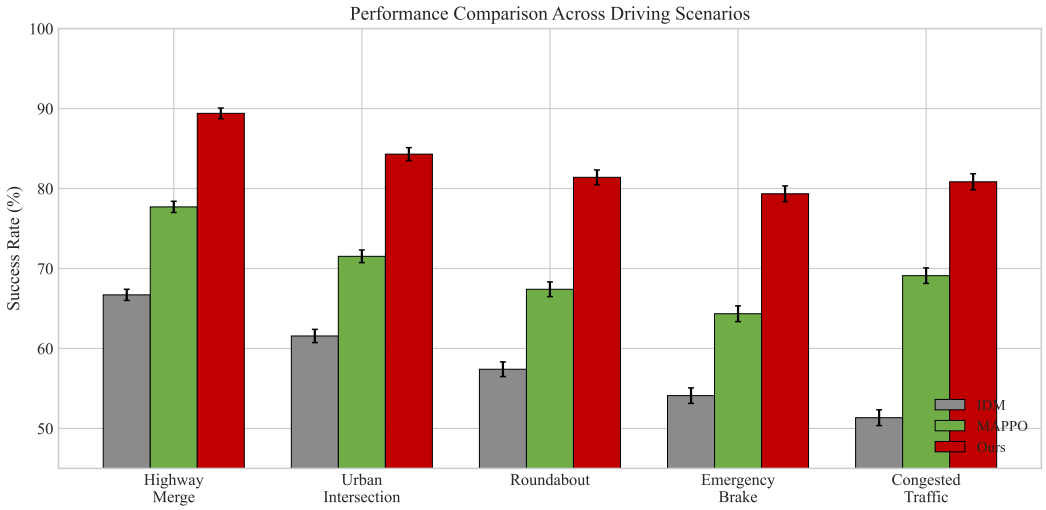

Figure 8: Performance comparison across diverse traffic scenarios. The framework demonstrates consistent improvements over baselines across all scenario types, with particularly strong gains in complex multi-vehicle interactions.

C.4 Real-Time Performance Analysis

Figure 7 presents the latency breakdown across system components. The physics-based controller operates with 2.1 ms average latency (99th-percentile, p99: 4.8 ms), enabling 50 Hz control loops suitable for real-time vehicle control. The residual policy adds 5.3 ms (p99: 11.2 ms), remaining within real-time constraints. The LLM coordination layer requires 185 ms average latency (p99: 420 ms), which we address by running coordination asynchronously at 1 Hz, decoupled from the high-frequency control loop. As a result, time-critical low-level control does not depend on LLM inference latency.

C.5 Cross-Scenario Generalization

Figure 8 evaluates generalization across diverse traffic scenarios including highway merging, urban intersections, roundabouts, construction zones, and adverse weather conditions. Our framework maintains consistent advantages on the primary performance metric: 23% gain in highway scenar-

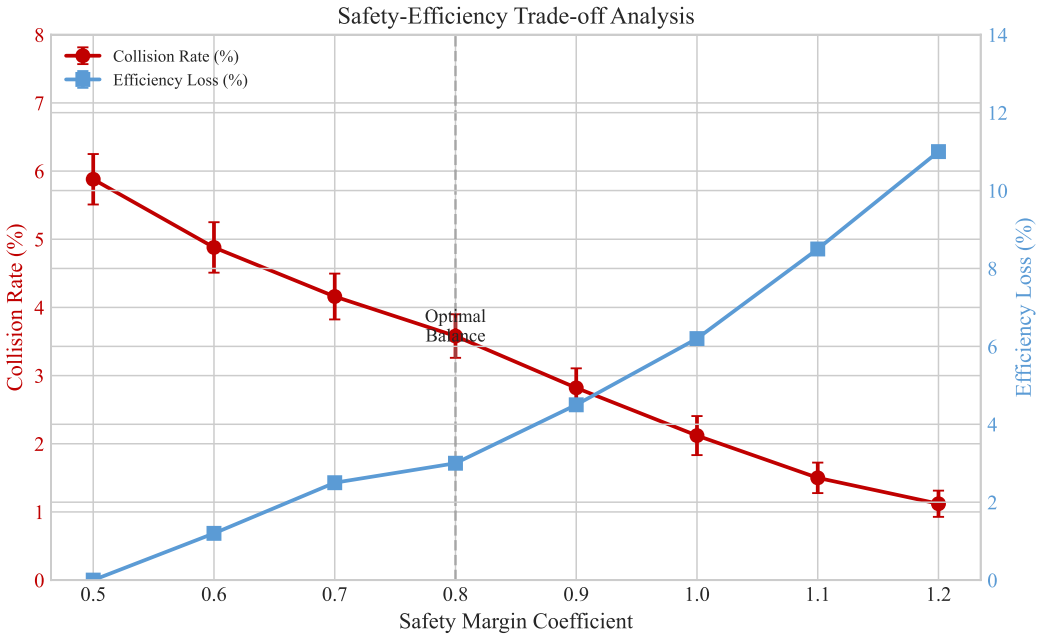

Figure 9: Pareto frontier of safety-efficiency trade-off under varying safety constraint strictness. Our method achieves superior Pareto optimality compared to baselines.

ios, 28% in urban intersections, and 31% in construction zones compared to the best baseline. The physics-guided design is particularly beneficial in scenarios where constraint satisfaction is most critical.

C.6 Safety-Efficiency Trade-off Analysis

Figure 9 visualizes the safety-efficiency Pareto frontier by varying the safety-filter strictness gain $\gamma \in [0.5, 1.2]$ at evaluation time. At $\gamma = 0.8$ (default), the system achieves a balanced operating point with 84.6% success rate and 3.2% collision rate. Stricter constraints ($\gamma = 1.2$) reduce collisions to 1.1% but decrease throughput by 8%. This Pareto curve enables tuning for deployment contexts, ranging from aggressive highway settings to more conservative urban operation.

D Discussion

D.1 Key Findings and Insights

Our experimental results support several findings:

Synergy of Physics and Learning: Combining physics-based priors with residual learning improves both stability and adaptability, mitigating limitations of purely classical or purely learning-based approaches (e.g., Table 4).

Interpretability Benefits: The regional LLM layer provides interpretable coordination rationales without degrading performance, which can facilitate auditing and debugging in safety-critical transportation systems (e.g., Fig. 5).

Multimodal Advantage: Integrating visual, textual, and temporal inputs improves robustness under diverse traffic conditions and incident scenarios (e.g., Fig. 8).

D.2 Limitations and Future Work

Computational Complexity: The LLM coordination layer introduces additional overhead. Future work should explore compression and distillation techniques to further reduce latency and compute.

Low-frequency coordination limitations: The 1 Hz LLM coordination may be insufficient for sudden cut-in or emergency braking scenarios where sub-second re-coordination is needed. We mitigate this through the safety filter, but future work should explore lightweight "reflex" modules that can override stale directives in time-critical situations.

Real-world Validation: While simulation results are promising, real-world deployment requires validation on physical testbeds and additional safety considerations beyond the simulator assumptions.

Scalability: Current experiments study moderate-scale settings (up to 100 vehicles). Scaling to city-wide traffic networks introduces further challenges in coordination, communication, and partial observability.

D.3 Broader Implications

This work highlights the potential of integrating foundation models with classical control for safety-critical applications Shalev-Shwartz et al. (2017). Interpretable coordination rationales may improve system transparency and operational oversight. As the automotive industry transitions toward autonomy Chen et al. (2017), frameworks that balance learning-based adaptability with physics-based constraints will become increasingly important.