Differentiable Model Predictive Safety for Heterogeneous Mobility at Urban Intersections

Wenzhe Song
School of Business
Stevens Institute of Technology
Hoboken, 07030, USA
wsong14@stevens.edu

Hao Zhang*
Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, 15213, USA
haoz4@andrew.cmu.edu

***Abstract*—The imminent integration of autonomous vehicles and mobile robots in urban settings presents a critical safety challenge for future intelligent transportation systems. This paper addresses the complex problem of coordinating heterogeneous agents with disparate dynamics at unregulated intersections. We introduce a novel framework, differentiable model predictive safety (DMPS), which embeds the foresight of model-predictive control into a data-driven, end-to-end reinforcement learning architecture. DMPS agents learn a latent dynamics model to predict future trajectories contingent on their actions. A learned, differentiable safety critic then evaluates the risk of these trajectories. Crucially, by leveraging backpropagation through the entire unrolled predictive model, agents can efficiently compute the gradient of future safety with respect to their current action, enabling a minimal and precise online safety correction. Integrated into a multi-agent training scheme, DMPS virtually eliminates collisions to less than 5.6% in high-density, mixed vehicle-robot traffic simulations, demonstrating state-of-the-art safety without compromising energy and traffic efficiency.**

Keywords*—autonomous vehicles; mobile robots; multi-agent coordination; differentiable model predictive safety; safety-critical reinforcement learning

I. INTRODUCTION

The convergence of autonomous systems and urban infrastructure is paving the way for smart cities, envisioned as seamlessly integrated ecosystems of intelligent agents [1]. A cornerstone of this vision is the transformation of public spaces, particularly urban intersections, into bustling hubs of autonomous activity [2]. These nexuses will no longer serve homogeneous vehicular traffic alone, but will host a complex, heterogeneous ensemble of agents, including high-speed autonomous vehicles (AVs) for passenger transport and low-speed autonomous mobile robots (AMRs) for logistics and services [3]. The safe, efficient, and scalable coordination of this mixed traffic presents a frontier scientific challenge, demanding a paradigm shift in how we approach autonomous decision-making.

This emerging reality of heterogeneous traffic introduces complexities that transcend the scope of current research, as depicted in Fig. 1. The problem is characterized by three fundamental challenges. First, the disparate dynamics are profound, also, multi-ton vehicles with 2D freedom of movement share pathways with lightweight robots whose motion is constrained to predefined 1D corridors. This invalidates the common assumption of agent homogeneity that underpins many multi-agent coordination algorithms [4]. Second, the asymmetric vulnerability is extreme. A collision between two vehicles is an accident; a collision between a vehicle and a robot is a catastrophic failure for the robot, imposing a non-uniform, high-stakes safety requirement [5]. Third, the interaction complexity is combinatorial. The introduction of a network of robot transit paths that intersect with vehicular lanes creates a dense web of conflict points, leading to a combinatorial explosion in potential interaction scenarios that must be safely navigated [6].

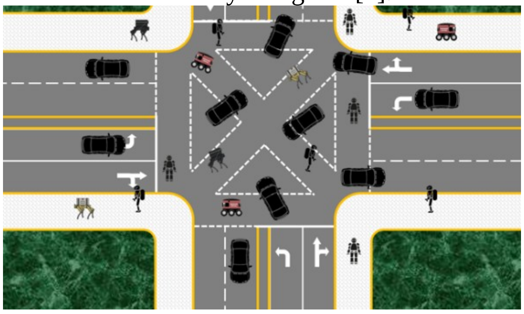

Fig. 1. Illustration of a mixed-traffic scenario with autonomous vehicles and mobile robots.

Addressing this multifaceted challenge reveals a fundamental tension between the two dominant paradigms in safe autonomy. On one hand, control-theoretic and formal methods champion rigor [7]. Techniques like control barrier functions (CBFs) offer provable safety guarantees by enforcing constraints within a known state space [8]. State-of-the-art implementations can handle complex, non-analytic road boundaries by formulating the problem as quadratic programing to be solved online. However, this entire paradigm is predicated on the availability of an accurate, low-dimensional analytical model of the system's dynamic [9]. This assumption is untenable for perception-based agents whose state is a high-dimensional sensor reading, and crafting a single, coherent set of CBF constraints for multiple agent classes with fundamentally different control modalities is a problem of immense complexity [10].

On the other hand, the paradigm of deep reinforcement learning (DRL) offers unparalleled adaptability, learning complex behaviors directly from high-dimensional inputs [11]. However, RL's inherent reliance on trial-and-error exploration is fundamentally unsafe, and the resulting policies are opaque black boxes that lack performance guarantees, especially in scenarios not encountered during training [12]. While some work has explored integrating RL with safety mechanisms, these often take the form of simple post-hoc filters or reward penalties that fail to address the core issue of proactive safety [13].

This paper argues that the path forward requires a principled synthesis of these two worlds. We introduce differentiable model predictive safety (DMPS), a framework that integrates the predictive foresight of model-based control with the adaptive power of deep learning. Inspired by model

Accepted version of the paper published in: W. Song and H. Zhang, "Differentiable Model Predictive Safety for Heterogeneous Mobility at Urban Intersections," 2025 IEEE 5th International Conference on Industrial Automation, Robotics and Control Engineering (IARCE), Chongqing, China, 2025, pp. 1-6. DOI: 10.1109/IARCE68366.2025.11485680. The final, published version is available at: https://ieeexplore.ieee.org/document/11485680

predictive control (MPC), DMPS endows each agent with an internalized, learned model of the world to predict the future consequences of its actions. A learned, differentiable cost function then evaluates the safety of these predicted futures. The core innovation lies in leveraging the end-to-end differentiability of this entire predict-and-evaluate pipeline. Using standard backpropagation, an agent can efficiently compute the analytical gradient of its future safety with respect to its current action. This allows for a single, optimal gradient descent step on the action itself, effectively performing an online trajectory optimization in a single, efficient pass. By embedding this mechanism within a multi-agent RL framework, we create a system where agents can learn the nuanced, cooperative strategies required to safely share the road.

II. PROBLEM FORMULATION

We formalize the complex challenge of heterogeneous traffic coordination as a decentralized partially observable Markov decision process (Dec-POMDP). This section provides the abstract mathematical foundation, with specific instantiations detailed in the methodology.

A. *Environment and Agent Specification*

The operational environment consists of two superimposed domains. The primary domain is the 2D plane, accessible to vehicles. The secondary domain is a 1D network of paths, accessible only to robots. The environment geometry is defined by the spaces accessible to each agent type. 1) The vehicle-accessible space: , is the subset of the 2D plane representing the road surface; 2) The robot transit network, , consists of a finite set of predefined, parameterized paths. Each path is a function that maps a distance along the path to a 2D pose (position and orientation). The robot-accessible space, , is the union of all these path geometries. The conflict points at the designed mixed-traffic urban intersection is modeled as shown in Fig. 2.

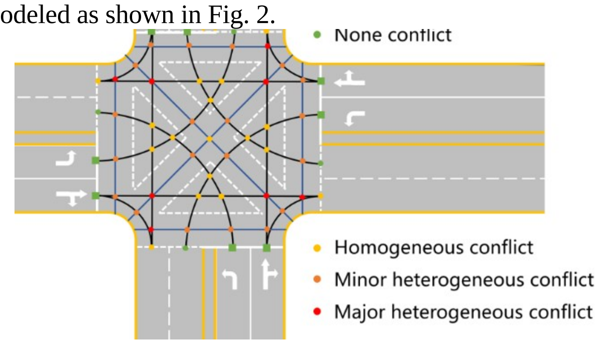

Fig. 2. Conflict point distribution at the designed mixed-traffic urban intersection with autonomous vehicles and mobile robots.

The system is constituted by a set of heterogeneous agents, , which is the union of vehicle agents, , and robot agents, . The global state is the composite of all individual agent states. The action of an agent is from its action space , and the joint action of all agents is . The system evolves according to the transition dynamics , and each agent receives a local observation based on the global state.

B. *Objective and Constraints*

The objective for the collective of agents is to learn a set of decentralized policies, , that maximizes the joint expected long-term return, defined by a reward function and a discount factor. The central research problem is thus to find an optimal set of policies that solves the following high-dimensional, partially observable, constrained optimal control problem:

(1)

The DMPS framework is architected to find a high-quality, tractable solution to this problem by learning the safety aspect of the constraints implicitly from data. For the system to be considered valid at all times t, the following constraints must hold.

(2)

(3)

where two fundamental sets of hard constraints are built. 1) Path adherence constraint for robots: Eq. (2) constraints that the pose of any robot must remain on its designated path network. 2) Safety constraint for all agents: the physical footprints of any two agents must not intersect. Let be the physical area occupied by agent . Then, the set of safe states is defined in Eq. (3).

III. DIFFERENTIABLE MODEL PREDICTIVE SAFETY

To find a tractable, high-performance solution to the constrained optimal control problem articulated in Eq. (1), we introduce our framework DMPS. This framework is conceived as a data-driven, learnable approximation of classical MPC. It eschews the need for analytical models and computationally expensive online solvers by embedding a predictive optimization process directly into the agent's policy architecture. This design transforms the agent from a purely reactive decision-maker into a proactive, deliberative one. The methodology is structured as a cohesive pipeline, beginning with the concrete specification of our heterogeneous agent models, followed by the development of a differentiable predictive engine, and culminating in a principled online action optimization mechanism.

A. *System Instantiation: Heterogeneous Agent Modeling*

Our first methodological step is to instantiate the abstract definitions from Section II with the concrete mathematical representations for our specific multi-agent system. This framing of the agents and their objectives is itself a contribution to modeling complex urban environments.

The state-action tuple for a vehicle agent , which operates with two degrees of freedom in the plane, is defined as:

(4)

where the state is specified by its 2D pose and velocity vector. The action is a 2D vector that controls its longitudinal acceleration and steering rate. In contrast, the state-action tuple for a robot agent , which is kinematically constrained to a predefined network of 1D paths, is defined as:

(5)

Here, the robot's state is defined by its assigned path index, its scalar distance traveled along that path, and its longitudinal velocity. Its 2D pose is deterministically calculated from its path index and distance. Consequently, its control is simplified to a scalar action representing its longitudinal acceleration along the fixed path.

The objective driving the learning process for each agent is to maximize a multi-component reward function Ri, specified as:

(6)

Each component serves a distinct purpose: a goal component provides a large sparse reward for task completion; a time component applies a small penalty per timestep to encourage efficiency; a comfort component penalizes high jerk to promote smooth behavior; and a proximity component acts as a dense, soft safety signal by penalizing small distances to other agents, thereby aiding the learning process and complementing the hard safety constraint imposed by the DMPS optimization.

Energy Efficiency:

DMPS reduces energy use through two mechanisms. The com fort term in Eq. (6) adds a penalty on jerk—the derivative of ac celeration. High jerk means the vehicle is changing its accelera tion rapidly, like when a driver floors the gas pedal or slams the brakes. Our agents learn to avoid this. The second mechanism comes from Eq. (11). The gradient gives us the direction to mo ve the action for better safety, but we only take a small step. W e do not override the original action completely. So if the polic y says "accelerate at 2 m/s²" and there is some risk ahead, mayb e we reduce it to 1.5 m/s² instead of forcing a full stop. The pre diction horizon H also matters here. With H=10 steps of looka head, an agent can see a robot entering its path 1 second before they actually meet. It starts decelerating at maybe 1 m/s² rather than waiting and then braking at 4 m/s². After the robot passes, the agent speeds up slowly. Compare this to AM-MAPPO: that method uses hard action masking, so agents often stop complet ely even when slowing down would be enough. Full stops follo wed by re-acceleration waste significant energy.

B. *Differentiable Predictive Modeling and Risk Assessment*

This stage forms the agent's deliberative core, providing a solution to the challenge of partial observability and unknown dynamics. We construct a learned, differentiable mapping from the current observation and a candidate action to a future risk assessment. This is achieved by training two coupled neural networks: a latent dynamics model and a safety critic.

First, an encoder network, , maps a high-dimensional observation to a compact latent state vector. This latent space is learned such that it forms a sufficient statistic for both prediction and control. The core predictive engine is a latent forward dynamics model, . This network functions as the agent's learned, implicit model of the world's physics, predicting the next latent state:

(7)

To enable genuine foresight, we recursively unroll this single-step model over a prediction horizon of timesteps, generating an entire predicted future trajectory of latent states contingent on the current action. Having predicted a future trajectory, a differentiable safety critic, , assigns a non-negative scalar safety cost to any given latent state. The theoretical advantage of a learned critic is its capacity to approximate a highly complex, non-linear risk function from data—a function that would be intractable to define analytically, especially given the asymmetric vulnerabilities in our heterogeneous setting. The total cost for a trajectory is then defined using a risk-averse max-pooling operation over the horizon, corresponding to the -norm of the future cost sequence, which makes the agent sensitive to the single most dangerous predicted moment:

(8)

These predictive components are trained jointly by minimizing a combined loss function, which is a weighted sum of a dynamics loss and a critic loss. The dynamics loss is a self-supervised objective that ensures temporal consistency in the latent space, while the critic loss trains the critic to regress to a ground-truth cost based on simulator data.

C. *Action Optimization via Differentiable Programming*

This final stage constitutes a principled, online control synthesis mechanism that directly addresses the safety constraint. Given a nominal action from the base RL policy, our objective is to solve the per-timestep optimal control problem:

(9)

This is a high-dimensional, non-convex optimization problem that is impossible to solve exactly in real- time. Our key insight is to approximate the solution by taking a single step of a first-order optimization method: gradient descent. The power of our framework lies in the fact that the gradient of the objective function can be computed analytically and efficiently.

This is achieved by leveraging the end-to-end differentiability of our predictive pipeline. We perform a single backpropagation pass, where the scalar total cost is differentiated with respect to the action vector. This process, known as backpropagation through time (BPTT), implicitly applies the chain rule through the entire unrolled computational graph of the dynamics model and the critic. The resulting action gradient is formally expressed as:

(10)

The first term in the sum is the gradient of the cost function with respect to a future state, obtained via the safety critic. The second term is the total derivative of the future state with respect to the current action, which encapsulates how a change in today's control input propagates through the learned dynamics to affect the state steps into the future. Thus, the action gradient represents the sensitivity of the maximum future risk to an infinitesimal change in the current action.

To improve safety, we perform a projected gradient descent step on the action. The final, safety- corrected action is given by:

(11)

where is the optimization step size, a hyperparameter controlling the magnitude of the safety correction. This mechanism provides a principled, minimal, and highly effective online correction. Furthermore, it forms a tight feedback loop where the safety module not only corrects unsafe actions at runtime but also provides a rich, structured gradient signal that regularizes the base RL policy towards inherently safer, more proactive behaviors.

Computational Efficiency

DMPS adds minimal overhead to the control pipeline. The key insight is that computing the safety gradient requires just one f orward rollout and one backward pass—no iterative solver is n eeded. We use a 256-dimensional latent space, which is suffici ent to capture the relevant dynamics while keeping matrix oper ations fast. In practice, the dynamics model and critic are small MLPs that run efficiently on a GPU. The entire safety correctio n step takes a small fraction of the 100ms control period, leavin g ample time for perception and other modules.

Multi-Scale Coordination

We did not explicitly design three separate coordination mechanisms, but looking at how trained agents behave, we can identify different time scales at work. The safety critic operates fastest. It looks at predicted states and computes a gradient when something dangerous shows up. This is basically reactive, happening within the prediction horizon. What surprised us more was the medium-term behavior. Agents start adjusting well before a conflict becomes critical. A vehicle at 50 km/h notices a slow robot (they only go up to 5 km/h) and begins decelerating early. No need for the full -8.0 m/s^2 braking capacity. Why does this happen? The dynamics model gives foresight beyond immediate danger. Then there is a slower pattern we only noticed after training finished. Over 5×10^7 steps, certain coordination habits emerged. Vehicles yield to robots at specific conflict points. Faster agents tend to go first. Nobody coded these rules. The reward function just penalizes collisions and delays. The rest came from learning.

IV. EXPERIMENTAL EVALUATION

We conduct a comprehensive experimental evaluation to rigorously validate the efficacy and efficiency of the DMPS framework. Our evaluation is centered on a single, highly challenging, and representative urban intersection scenario. Our analysis focuses on two key aspects: (1) a direct performance comparison of our proposed method against strong, established baselines across critical safety and efficiency metrics, both during and after training, and (2) a targeted ablation study to isolate and understand the contributions of the core components within the DMPS framework itself.

A. *Experimental Setup*

All experiments are conducted within the CARLA simulator, chosen for its realistic physics and sensor models. Our evaluation is centered on a complex, signal-free urban intersection scenario adapted from CARLA's map. This specific environment was chosen for its high degree of interaction complexity, featuring a four-lane, two-way arterial road for vehicles that intersects with a network of six predefined, one-dimensional transit corridors for mobile robots (four perpendicular and two diagonal), creating a challenging conflict zone. To ensure the robustness of our findings, we evaluate all methods under varying traffic densities within this scenario, ranging from light (5-10 agents) to dense and highly interactive conditions, with up to 10 vehicles and 15 robots simultaneously active.

The system comprises two distinct agent classes. The AVs are modeled as standard sedans, possessing a 2D action space defined by continuous longitudinal acceleration between -8.0 and 4.0 m/s^2 and a steering angle between -0.6 and 0.6 radians, with a maximum speed of 50 km/h. The AMRs are modeled as small, wheeled delivery platforms. Their movement is constrained to the predefined path network, and their action space is 1D, defined solely by longitudinal acceleration between -1.0 and 1.0 m/s2 along their designated path, with a maximum speed of 5 km/h. All agents, regardless of type, are equipped with an identical sensor suite, including a 360-degree, 32-channel LiDAR for environmental perception and proprioceptive sensors for egomotion data.

To rigorously evaluate the performance of our DMPS-MAPPO framework, we benchmark it against three strong and representative baselines. The first, Vanilla MAPPO, serves as a fundamental performance benchmark. The second, reward shaping (RS-MAPPO), represents a common heuristic approach using dense proximity penalties. Finally, we compare against a robust rule-based method, action masking (AM-MAPPO), which uses a constant-velocity prediction model to forbid any discrete action predicted to cause a collision. Our evaluation methodology relies on a portfolio of key performance indicators. Learning stability and efficiency are assessed via the mean reward per Episode throughout the training process. Final system safety is measured by the collision rate (%) dissected into total collisions and critical vehicle-robot incidents. Overall task effectiveness is gauged by the success rate (%) and operational efficiency is quantified by the average delay (s).

B. *Comparative Analysis of Safety Frameworks*

We first analyze the learning dynamics of all frameworks, with the training curves for mean episodic reward shown in Fig. 3. The plot clearly highlights the superior learning stability and efficiency of our DMPS-MAPPO method. It achieves a smoother and more monotonically increasing reward curve, converging to the highest final reward level. This indicates that the DMPS module, by preventing catastrophic failures during exploration, provides a more stable learning signal for the policy, mitigating the challenge of credit assignment after sparse negative events.

In contrast, Vanilla MAPPO exhibits highly volatile training with frequent collapses in performance, which is characteristic of learning in safety-critical environments where fatal errors lead to episode termination and poor policy updates. RS- MAPPO shows improved stability over the vanilla version, but its convergence is slower and it plateaus at a lower reward, suggesting that the dense negative reward penalties, while helpful, may discourage some optimal, efficient behaviors. AM-MAPPO, while also stable, converges to a significantly lower reward level, likely because its conservative masking mechanism frequently limits the policy's ability to explore and find the most efficient solutions. The superior learning process of DMPS-MAPPO translates directly into class-leading final performance. The analysis of the converged policies, presented in Fig. 4 and Fig. 5, further solidifies these findings.

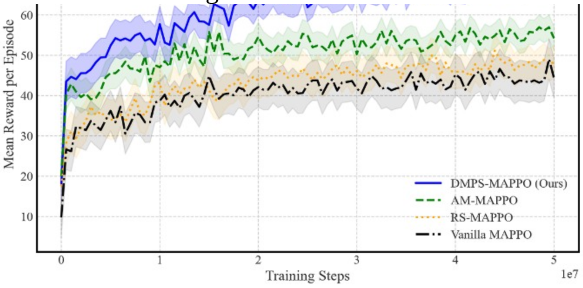

Fig. 3. Training progression of DMPS-MAPPO and baseline frameworks, showing mean reward per episode over training steps.

TABLE I. FINAL PERFORMANCE COMPARISON IN THE DENSE, HETEROGENEOUS TRAFFIC SCENARIO

Method	**Performance**			
	Collisions (%)	**Vehicle-Robot Collisions (%)**	**Success Rate (%)**	**Average Delay (s)**
Vanilla MAPPO	18.5 ± 2.4	12.1 ± 1.8	25.4 ± 3.1	12.5 ± 1.3
RS-MAPPO	11.2 ± 1.9	6.8 ± 1.1	52.3 ± 3.4	14.8 ± 1.5
AM-MAPPO	7.1 ± 1.3	3.5 ± 0.9	61.5 ± 2.6	18.2 ± 2.1
DMPS-MAPPO	4.8 ± 0.8	2.9 ± 0.6	81.6 ± 2.2	14.2 ± 1.2

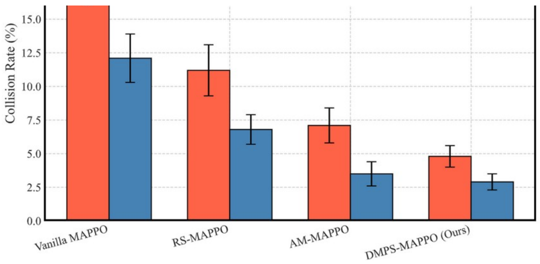

Fig. 4. Final collision performance comparison of DMPS-MAPPO and baseline frameworks.

In the critical dimension of safety, DMPS-MAPPO sets a new state-of-the-art, reducing the total collision rate to 4.8%, as shown in Table I. This represents a 32.7% relative improvement over the strongest safety baseline, AM-MAPPO. The framework is particularly effective at mitigating the most severe outcomes, lowering the critical vehicle-robot collision rate to 2.9%. This strong safety record is a direct result of the DMPS's core mechanisms: the learned safety critic correctly identifies high-risk situations, and the predictive model provides the necessary foresight to allow the optimizer to steer agents clear of danger. Crucially, this substantial gain in safety is achieved while maintaining high operational efficiency.

With a task success rate of 81.6%, our method significantly outperforms all baselines in achieving the overall objective. The average delay for DMPS-MAPPO agents is 14.2s. While this reflects a modest and expected cost for the added safety layer compared to the unsafe Vanilla MAPPO (12.5s), it is substantially more efficient than the overly conservative AM-MAPPO (18.2s). This highlights a key advantage: our framework makes minimal, precise adjustments rather than resorting to the inefficient "stop-and-go" behaviors induced by hard constraints.

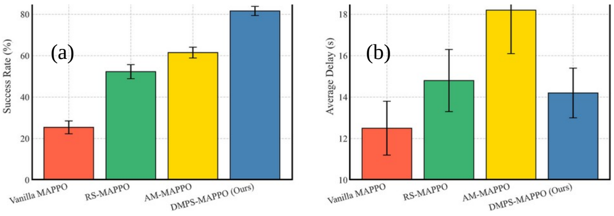

Fig. 5. Performance metrics with final success rate comparison (a) and final operational efficiency comparison (b).

C. *Ablation Study on DMPS Mechanisms*

To isolate and verify the contributions of the core components within the DMPS framework itself, we performed a targeted ablation study as shown in Fig. 6. We tested two variants: (1) DMPS w/o prediction, a reactive version where the safety critic only evaluates the safety of the current latent state, removing the foresight capability. (2) DMPS w/o gradient, a version that uses the full predictive model but replaces the precise gradient-based correction with a fixed, heuristic push.

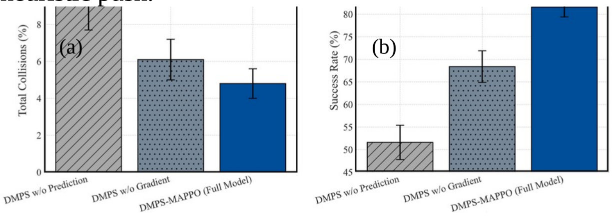

Fig. 6. Ablation study of DMPS core components with (a) for collision rate and (b) for success rate.

TABLE II. BLATION STUDY OF DMPS COMPONENTS

Ablation Variant	**Performance**	
	Total Collisions (%)	**Success Rate (%)**
DMPS w/o Prediction	9.2 ± 1.5	51.6 ± 3.8
DMPS w/o Gradient	6.1 ± 1.1	68.4 ± 3.5
DMPS-MAPPO	4.8 ± 0.8	81.6 ± 2.2

The ablation results presented in Table II are clear. Removing the predictive capability (w/o prediction) causes the collision rate to nearly double to 9.2%, proving that foresight is critical for navigating complex, dynamic interactions. Similarly, replacing the precise gradient-based optimization with a simpler heuristic (w/o gradient) is also highly detrimental, increasing collisions to 6.1%. This demonstrates that the minimal and contextually-aware nature of the learned gradient is key to achieving a high degree of safety without crippling efficiency. We conclude that both predictive modeling and differentiable optimization are essential and synergistic components of the DMPS framework's success.

V. CONCLUSIONS

This paper addressed the critical challenge of safe coordination in heterogeneous traffic systems. We introduced DMPS, a framework that integrates the foresight of model predictive control into a fully differentiable deep learning architecture. By learning latent dynamics and a differentiable safety critic, our framework provides a principled way to compute minimal, precise safety corrections within an end-to-end learning loop. Our results show that this approach can effectively resolve complex multi-agent conflicts between agents with fundamentally different motion constraints, yielding a 32.7% reduction in collisions over baselines while maintaining high operational through- put. DMPS represents a significant step towards creating robustly safe AI systems, paving the way for the complex and dynamic autonomous ecosystems of our future cities.

REFERENCES

[1] Sun H, Li B, Zhang H, et al. Ecological electric vehicle platooning: an adaptive tube-based distributed model predictive control approach. IEEE Trans. on Transportation Electrification, 2024,11:1048-1060.

[2] Zhang H, Chen B, Lei N, et al. Coupled velocity and energy management optimization of connected hybrid electric vehicles for maximum collective efficiency. Applied Energy, 2024,360:122792.

[3] Al-Aawar, N. and Arkadan A.-R.A., "Optimal Control Strategy for Hybrid Electric Vehicle Powertrain," IEEE Journal of Emerging and Selected Topics in Power Electronics, 2015, 03:362-370.

[4] Zhang H, Lei N, Chen B, et al. Bi-level transfer learning for lifelong intelligent energy management of electric vehicles. IEEE Transactions on Intelligent Transportation Systems, 2025, Early Access.

[5] Chen C, Cai M, Wang J et al. Cooperation method of connected and automated vehicles at unsignalized intersections: lane changing and arrival scheduling. IEEE Trans Veh Technol 2022; PP:1–16.

[6] Zhang H, Chen B, Lei N, et al. Integrated thermal and energy management of connected hybrid electric vehicles using deep reinforcement learning. IEEE Trans. on Transportation Electrification, 2024,10:4594-4603.

[7] Li B, Zhuang W, Zhang H, et al. A comparative study of energy-oriented driving strategy for connected electric vehicles on freeways with varying slopes. Energy, 2023,289:129916.

[8] Zhang H, Lei N, Chen B, et al. Data-driven predictive energy consumption minimization strategy for connected plug-in hybrid electric vehicles. Energy, 2023,283:128514.

[9] Chen C, Wang J, Xu Q, et al. Mixed platoon control of automated and human-driven vehicles at a signalized intersection: dynamical analysis and optimal control. Transp Res Part C. 2021;127:103138.

[10] M. Noaeen, A. Naik, L. Goodman, et al. Reinforcement learning in urban network traffic signal control: A systematic literature review, Expert Systems with Applications. 2022, 199:116830.

[11] Zhang H, Lei N, Chen B, et al. Modeling and control system optimization for electrified vehicles: A data-driven approach. Energy, 2024,311:133196.

[12] Lei N, Zhang H, Wang H, et al. Theory-Constrained Neural Network with Modular Interpretability for Fuel Cell Vehicle Modelling. IEEE Trans. on Vehicular Technology, 2025, Early Access.

[13] Yazdani M., Parineh H., Sarvi M., et al. Intelligent vehicle pedestrian light (IVPL): A deep reinforcement learning approach for traffic signal control, Transportation Research Part C: Emerging Technologies, 2023, 125: 102942.