5G-enabled Safe and Robust Deep Multi-Agent Reinforcement Learning Framework for CAV Coordination

June 2025

Region 1:
New England University Transportation Center

Fei Miao

Song Han

University of Connecticut

In cooperation with U.S. Department of Transportation,
Office of the Assistant Secretary for Research and Technology (OST-R)

Grant #: 69A3552348301

TECHNICAL DOCUMENTATION

1. Project No. 161157	**2. Government Accession No.** 01904668	**3. Recipient's Catalog No.**

4. Title and Subtitle 5G-enabled Safe and Robust Deep Multi-agent Reinforcement Learning Framework for CAV Coordination	**5. Report Date** June 2025
	6. Performing Organization Code N/A
7. Author(s) **Fei Miao 0000-0003-0066-4379** **Song Han 0000-0002-1491-7675**	**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address New England University Transportation Center 181 Presidents Drive University of Massachusetts - Amherst Amherst, MA 01003	**10. Work Unit No. (TRAIS)**
	11. Contract or Grant No. 69A3552348301
12. Sponsoring Agency Name and Address United States Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590	**13. Type of Report and Period Covered** Final Research Report
	14. Sponsoring Agency Code USDOT OST-R

15. Supplementary Notes

Report uploaded and accessible at the NEUTC Website (www.umass.edu/neutc)

16. Abstract

We address the problem of coordination and control of Connected and Automated Vehicles (CAVs) in the presence of imperfect observations in mixed traffic environment. A commonly used approach is learning-based decision-making, such as reinforcement learning (RL). However, most existing safe RL methods suffer from two limitations: (i) they assume accurate state information, and (ii) safety is generally defined over the expectation of the trajectories. It remains challenging to design optimal coordination between multi-agents while ensuring hard safety constraints under system state uncertainties (e.g., those that arise from noisy sensor measurements, communication, or state estimation methods) at every time step. We propose a safety guaranteed hierarchical coordination and control scheme called Safe-RMM to address the challenge. Specifically, the high-level coordination policy of CAVs in mixed traffic environment is trained by the Robust Multi-Agent Proximal Policy Optimization (RMAPPO) method. Though trained without uncertainty, our method leverages a worst-case Q network to ensure the model's robust performances when state uncertainties are present during testing. The low-level controller is implemented using model predictive control (MPC) with robust Control Barrier Functions (CBFs) to guarantee safety through their forward invariance property. We compare our method with baselines in different road networks in the CARLA simulator. Results show that our method provides the best evaluated safety and efficiency in challenging mixed traffic environments with uncertainties.

17. Key Words Autonomous driving, reinforcement learning, multi-agent system, safe and robust control, simulation (http://trt.trb.org)	**18. Distribution Statement** No restrictions.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 10	**22. Price**

About NEUTC

The New England Regional University Transportation Center (NEUTC) is a diverse, multidisciplinary consortium committed to addressing the pressing issue of traffic safety. Our objective, in line with the Infrastructure Investment and Jobs Act (IIJA), is to drive transformative research, education, and technology transfer to address critical traffic safety needs in a time when roadway fatalities are distressingly high.

Our research and educational activities at NEUTC are guided by four principal safety themes, each addressing a critical challenge in transportation safety. These themes capture the various integral components of the transportation system, focusing on technology, infrastructure, vehicles, and users with a commitment to public engagement. Our overarching theme is promoting safety, with the common underlying science being the study of behavioral, systemic, environmental, and mobility-driven factors on safety.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under 69A3552348301 from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

All matching funds were University salaries and graduate tuition waivers.

Motivation

Machine learning models assisted by more accurate on-board sensors, such as camera and LiDARs, have enabled intelligent driving to a certain degree. Meanwhile, advances in wireless communication technologies also make it possible for information sharing beyond an individual's perception (Martín-Sacristán, et al., 2020; Mun, Seo, & Lee, 2021). Through vehicle-to-everything (V2X) communications, it has been shown that shared information can contribute to CAVs decision-making (Buckman, Pierson, Karaman, & Rus, 2020; Miller & Rim, 2020) and improve the safety and coordination of CAVs (Zhang, Han, Wang, & Miao, 2023). However, it remains challenging for reinforcement learning (RL) or multi-agent reinforcement learning (MARL)-based decision-making methods to guarantee the safety of CAVs in complicated dynamic environments containing human driven vehicles (HDVs) and optimize the joint behavior of the entire system. For real-world CAVs, state uncertainties from noisy sensor measurements, state estimation algorithms or the communication medium pose another challenge. There can be scenarios where safety is highly correlated with the correctness of state information, especially in the presence of HDVs/unconnected vehicles. Moreover, there is no hardware demonstration (small scale or full scale) of MARL for CAVs. The existing literature in this area generally falls into the following major categories.

Safe RL and Robust RL: Different approaches have been proposed to guarantee or improve safety of the system, such as defining a safety shield or barrier assisting RL or MARL algorithm in either training or execution stage (Brunke, et al., 2022; Cai, Cao, Lu, Zhang, & Xiong, 2021), constrained RL/MARL that learns a risk network (Wen, Duan, Li, Xu, & Peng, 2020), an expected cost function (Lu, Zhang, Chen, Başar, & Horesh, 2021), or cost constraints from language (Wang, Fang, Tomilin, Fang, & Du, 2024) that define the safety requirements. For MARL of CAV, safety-checking module with CBF-PID controller for each individual vehicle has been designed (Wang, et al., 2023; Zhang, Han, Wang, & Miao, 2023). However, the above works assume accurate state inputs to RL or MARL algorithm from the driving environment and cannot tolerate noisy or inaccurate state input. Meanwhile, robust RL and robust MARL that only considers to train a policy under state uncertainty or model uncertainty (Liang, Sun, Zheng, & Huang, 2022; Han, Wang, Su, Shi, & Miao, 2022; Salvato, Fenu, Medvet, & Pellegrino, 2021; Pinto, Davidson, Sukthankar, & Gupta, 2017) without explicitly considering the safety requirements have been proposed recently. However, in the multi-agent settings with imperfect observations, considering both safety requirements and robustness in a unified decision-making framework for CAVs still remains challenging.

Rule-Based Approaches: Unified optimization framework poses challenges that can be addressed by decomposing the problem into hierarchical structures. Specifically, the higher-level control is responsible for decision making and the lower-level control is responsible for safe execution. For the higher-level planner, heuristic rule-based methods can be employed in which a set of rules govern the behavior of each agent within the system. For instance, existing driving behavior models in mixed traffic can be found in (Treiber, Hennecke, & Helbing, 2000; Kesting, Treiber, & Helbing, 2007; Munigety, 2018). However, these models often lack robustness and make various assumptions about HDVs, which prevents generalization to all scenarios. MPC can be used for the lower-level controller due to its ability in reference tracking and handling hard constraints in real time. In situations where imperfect observations are present, robust MPC approaches may be used, such as tube MPC (Lopez, Slotine, & How, 2019; Mayne & Kerrigan, 2007; Sinha, Harrison, Richards, & Pavone, 2022). Nevertheless, tube-based MPC approaches require a feedback controller that can keep the actual system trajectory close to the nominal one. The calculation of such feedback controller is not trivial in multi-agent systems with nonlinear dynamics. Min-Max MPC (Raimondo, Limon, Lazar, Magni, & ndez Camacho, 2009) can also be adopted but it is often difficult to solve, and when it is approximated, the approximation can result in an overly conservative solution.

Executive Summary

Problem summary

We consider the robust cooperative policy-learning problem under uncertain state inputs for CAVs in mixed traffic environments including HDVs that do not communicate or coordinate with CAVs, and various driving scenarios such as multi-lane intersection and highway (as shown in below Figure 1). We assume that each CAV can get shared information from V2V and V2I communications. We consider that a CAV agent i has accurate self-observation of its driving state but potentially perturbed observations of the other vehicles. The two parts collectively constitute the state s_i in reinforcement learning and used by the MPC controller as inputs to the robust CBFs. We will demonstrate the proposed method in both simulator and F1/10th scale CAVs.

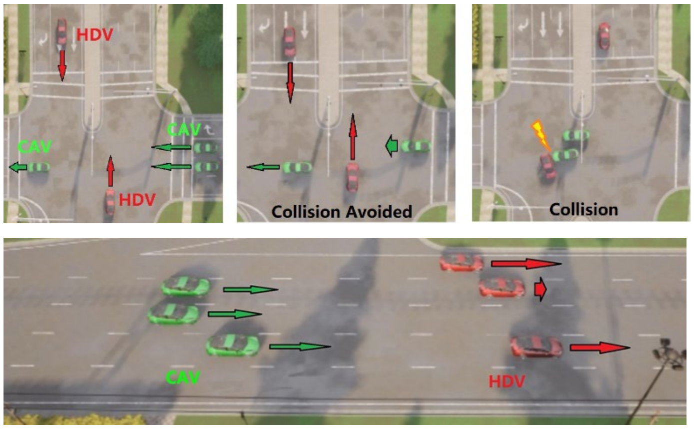

Figure 1: Intersection scenario (top 3) and highway scenario (below 1)

The problem of Multi-Agent Reinforcement Learning with State Uncertainty for CAVs is defined as a tuple $G = (S, A, P, r_i, \tilde{o}, G, \gamma)$ where $G = (N, E)$ is the communication network of all CAV agents. S is the joint state space of all agents. The state space of agent i: $S_i = \{o_i, o_{N_i}, o_{N_i^{UV}}\}$ contains self-observation o_i , observations $o_{N_i} = \{o_j | j \in N_i\}$ being the communicated message shared by neighbor connected agents N_i , observations $o_{N_i^{UV}}$ of unconnected vehicles N_i^{UV} either observed by agent i itself or shared by other agents or infrastructures. For example, self-observation o_i and shared observations o_{N_i} can contain location, velocity, acceleration and lane-detection; observations of unconnected vehicles N_i^{UV} can contain location and velocity.

Methodology Summary and Details We present our proposed safe MARL algorithm, Safe-RMM, demonstrated in the below Figure 2 with two major parts: Robust MAPPO (middle brown-and-blue block) and Robust CBF-based Model Predictive Control (right red block). It is a hierarchical decision-making framework design, it is begun by the design of our Robust MARL algorithm and followed by the details of the MPC controller using robust CBFs. Our Robust MARL algorithm augments MAPPO (Yu, et al., 2022) such that each PPO agent is equipped with a worst-case Q network (Liang, Sun, Zheng, & Huang, 2022). The worst-case Q estimates the potential impact of state perturbations on the policy's action selection and the resulting expected return. By incorporating it into the policy's training objective, we enhance the robustness of the trained policy against state perturbations. Consequently, the proposed algorithm improves both the safety and efficiency of CAVs under state uncertainties.

Robust MAPPO: The proposed Robust MARL algorithm, as shown in the Figure 2 above, uses centralized training and decentralized execution. We are inspired by the Worst-case-aware Robust RL framework (Liang, Sun, Zheng, & Huang, 2022) and designed the robust MAPPO. Each robust PPO agent maintains a policy network ("***actor***") $\pi_i(s_i)$, a value network ("***critic***") $V(s)$ and the second critic $Q_i(s_i, a_i)$ network approximating ***the worst-case action values***. To learn a safe cooperative policy for the CAVs, the MARL interacts with the MPC controller during the training rollout process. As the algorithm starts, agent i's policy takes initial state s_i and samples an action a_i; the MPC-controller with robust CBFs given a_i computes a safe control u_i for the vehicle to execute. The agent receives r_i and all agents synchronously move to the next time-step by observing the new state $s' = \prod_i s_i'$.

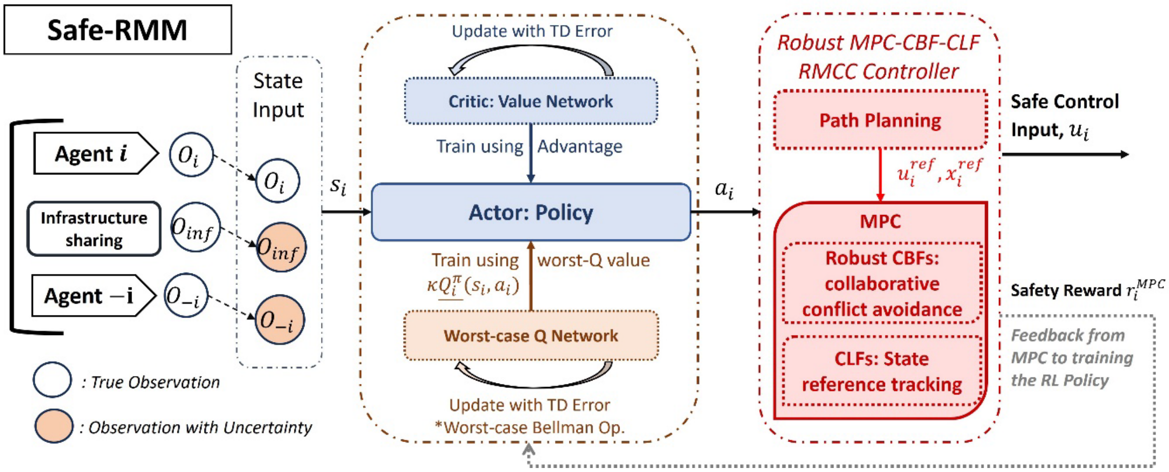

Figure 2 Safe-RMM Algorithm diagram

Robust Model Predictive Control: We adopt receding horizon control to implement a low-level controller for every agent i in the road network. The low-level controller maps the high-level plans/actions $a_i \sim \pi_i$ into primitive actions/control inputs for agent i. Firstly, a path planning function is used to map the high-level plans/actions into state and action references. Subsequently this information is fed to the MPC controller. To prevent collision between agents, safety constraints are incorporated into the controller using CBFs.

Experiment

We conduct our experiment in the CARLA Simulator environment (Dosovitskiy, Ros, Codevilla, Lopez, & Koltun, 2017), where each vehicle is configured with onboard GPS and IMU sensors and a collision sensor that detects the collision with other objects. We show two challenging scenarios at **Intersection** and on **Highway**, where we spawn multiple CAVs and some HDVs randomly (Figure 1). **Intersection** CAVs pass the intersection and HDVs from opposite sides cross the box at the same time. HDVs pose critical safety threats as they could either hit or be hit by a CAV from the side when driving fast ($\approx 10\ m/s$). CAVs aim to avoid collision and reach the preset destination after passing through the intersection.

Intersection: Training results in are shown in Figure 3; evaluations in both scenarios are presented in Table 1 below. For each entry in both tables, the left integer is the number of collisions happened during evaluation (in 50 episodes); the right value is the agents' mean discounted return considering only the rewards related to velocity and goal-achievement. We highlight the top performance across all methods with the least collisions and the highest efficiency return.

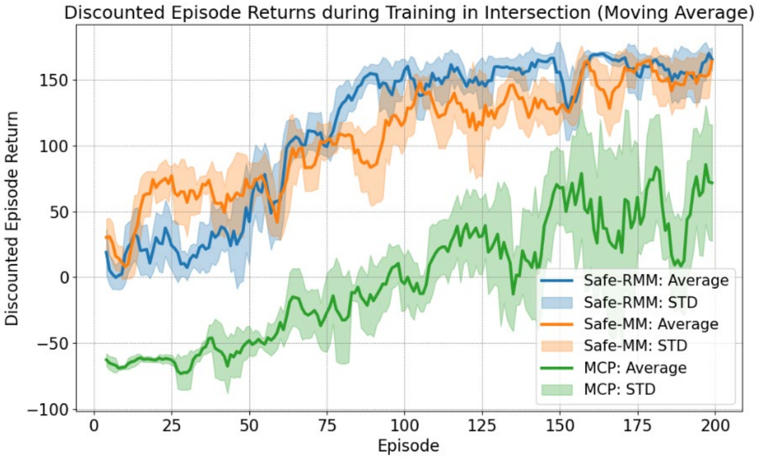

Figure 3 Agents' performances when training in Intersection

Evaluation Results in Intersection and Highway

	Uncertainty			
Method	**None**	$e^{\mathbf{rand}}$	$\mathrm{ERR}^{\mathcal{V}}$	ERR^{T}
	Intersection			
Safe-RMM[1]	**0, 162.9**	**0, 161.4**	**0, 162.2**	**0, 161.8**
Safe-MM[2]	**0**, 157.9	**0**, 155.7	**0**, 155.9	**0**, 155.7
MCP[3]	3, 65.7	2, 60.6	**0**, 66.2	2, 67.7
MP[4]	33, 148.4	41, 149.1	36, 145.9	30, 139.0
RULE[5]	2, 120.9	1, 113.9	3, 105.5	2, 112.3
	Highway			
Safe-RMM	**0, 162.0**	**0, 169.4**	**0**, 166.4	**0**, 161.8
Safe-MM	**0**, 161.3	**0**, 168.7	**0, 168.7**	**0, 163.0**
MCP	**0**, 56.8	2, 55.8	1, 60.7	2, 58.4
MP	35, 74.1	34, 74.5	38, 73.8	38, 74.5

Table 1 Evaluation Results of Safe-RMM compared with Baselines

Highway CAVs are spawned behind HDVs (red) on a multi-lane highway. During training and evaluation, HDVs keep in their lanes at random speed from $[7 \sim 9]\ m/s$ except one random HDV simulates a stop and go scenario. CAVs aim to avoid any collision and drive at the speed limit of the road to arrive at the destination. We trained three models: our Safe-RMM method, the "non-robust" Safe-MM that also adopts our framework, and the MCP method adopting MARL-PID controller with CBF safety shield (Zhang, Han, Wang, & Miao, 2023). Each model is trained on two scenarios respectively for 200 episodes. In evaluation, aside from the three trained models, we have "MP", MARL with PID controller, as an example of learning-based method without safety shielding; and we also implemented the benchmark "RULE" adopting a rule-based planner and a robust MPC controller. The "RULE" benchmark has been implemented based on the method proposed in (Sabouni, Ahmad, Cassandras, & Li, 2023) which introduces a safety-guaranteed rule for managing vehicle merging on roadways. This rule ensures safe interactions between vehicles arriving from different roads converging at a common point. Methods are evaluated for 50 episodes in both scenarios under four uncertainty configurations: None (uncertainty-free), random error e^{rand}, and two targeting errors ERR^V and ERR^T.

Hardware demonstration

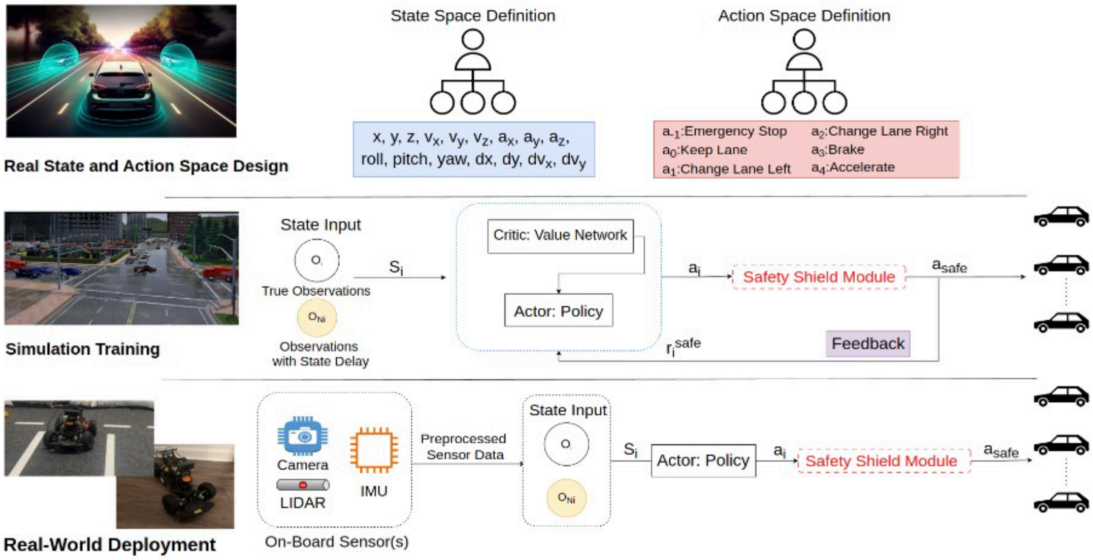

Figure 4 Hardware Demonstration of Robust MARL with Safety Guarantees and Communication

This paper introduces RSR-RSMARL as shown in Figure 4, a novel Robust and Safe MARL framework that supports Real-Sim-Real (RSR) policy adaptation for multi-agent systems with communication among agents, with both simulation and hardware demonstrations. Deep multi-agent reinforcement learning (MARL) has been demonstrated effectively in simulations for many multi-robot problems. For autonomous vehicles, the development of vehicle-to-vehicle (V2V) communication technologies provide opportunities to further enhance safety of the system. However, zero-shot transfer of simulator-trained MARL policies to hardware dynamic systems remains challenging, and how to leverage communication and shared information for MARL has limited demonstrations on hardware. This problem is challenged by discrepancies between simulated and physical states, system state and model uncertainties, practical shared information design, and the need for safety guarantees in both simulation and hardware. RSR-RSMARL leverages state (includes shared state information among agents) and action representations considering real system complexities for MARL formulation. The MARL policy is trained with robust MARL algorithm to enable zero-shot transfer to hardware considering the sim-to-real gap. A safety shield module using Control Barrier Functions (CBFs) provides safety guarantee for each individual agent. Experiment results on F1/10th-scale autonomous vehicles with V2V communication demonstrate the ability of RSR-RSMARL framework to enhance driving safety and coordination across multiple configurations. These findings emphasize the importance of jointly designing robust policy representations and modular safety architectures to enable scalable, generalizable RSR transfer in multi-agent autonomy.

Outcomes

This project focuses on designing a 5G-enabled safe and robust deep multi-agent reinforcement learning (MARL) framework for CAV coordination, to prove rigorous safety guarantees and the benefit of V2X communication in improving system safety and efficiency. The novelty of the achievement of the result spans fundamental theory and algorithm principles, model designs, and validation methodologies that will emerge to form a new integrated framework for communication, machine learning, and control of future safe AVs. In particular, the novel contributions of this proposal are as follows. (1) Given enhanced observation capability by 5G-based V2X communication, we design a safe multi-agent reinforcement learning (MARL) framework with a computationally tractable algorithm. It makes operation decisions that rigorously guarantee safety of AVs according to the requirements, and robust to various types of driving scenarios. (2) Based on the quantitative communication requirements from MARL, we develop an effective real-time flow scheduling framework for 5G-based V2X communications, featuring per-flow real-time schedulability guarantee through time-frequency-space resource allocation to provide desired communication quality for safety-critical decision-making and control functions on AVs. (3) We validate the proposed approach using simulators, small-scale testbeds such as F1/10th vehicles.

Papers that already published and submitted for this project:

1. Zhili Zhang, H M Sabbir Ahmad, Ehsan Sabouni, Yanchao Sun, Furong Huang, Wenchao Li, and Fei Miao, “Safety Guaranteed Robust Multi-Agent Reinforcement Learning with Hierarchical Control for Connected and Automated Vehicles”, ICRA 2025, Internation Conference on Robotics and Automation, https://arxiv.org/abs/2309.11057, May 2025.
2. Keshawn Smith, Zhili Zhang, H M Sabbir Ahmad, Ehsan Sabouni, Maniak Mondal, Song Han, Wenchao Li, and Fei Miao, “Robust and Safe Multi-Agent Reinforcement Learning Framework with Communication for Autonomous Vehicles”, under review, https://arxiv.org/abs/2506.00982, June 2025.

Impacts

This project has great potential for technology commercialization and will lead to significant societal impact to improve the safety of CAVs. The PIs take the following steps to pursue technology transfer. We demonstrate the success of the proposed methodology to CTSRC/CTI leadership team, in the future we hope to also demonstrate the hardware performance to CT DOT representatives using the small-scale CAV testbed mentioned above. Based on the received feedback, the research team will refine the methodology framework and deploy them on the electric autonomous buses on UConn Depot campus. The hardware demonstrations have also been shown for many K-12 outreach activities at UConn. The PIs work with Verizon to incorporate the proposed real-time 5G resource management solution in their 5G base stations install the enhanced 5G modules on the small-scale autonomous vehicles. The PIs work with Nvidia and Qualcomm to practical autonomous driving scenarios. In the future, the PIs will work with the CTSRC/CTI leadership, to plan demonstration of the developed technologies on full-scale CAVs testbed if possible, to representatives from different DOTs in New England area, government officials, and AV manufacturers to explore commercialization opportunities.

References

Brunke, L., Greeff, M., Hall, A. W., Yuan, Z., Zhou, S., Panerati, J., & Schoellig, A. P. (2022). Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems, 5*, 411–444.

Buckman, N., Pierson, A., Karaman, S., & Rus, D. (2020). Generating Visibility-Aware Trajectories for Cooperative and Proactive Motion Planning. 3220–3226.

Cai, Z., Cao, H., Lu, W., Zhang, L., & Xiong, H. (2021). Safe Multi-Agent Reinforcement Learning through Decentralized Multiple Control Barrier Functions. *Safe Multi-Agent Reinforcement Learning through Decentralized Multiple Control Barrier Functions*.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An Open Urban Driving Simulator. 1–16.

Han, S., Wang, H., Su, S., Shi, Y., & Miao, F. (2022). Stable and Efficient Shapley Value-Based Reward Reallocation for Multi-Agent Reinforcement Learning of Autonomous Vehicles. 8765–8771.

Kesting, A., Treiber, M., & Helbing, D. (2007). General Lane-Changing Model MOBIL for Car-Following Models. *Transportation Research Record, 1999*, 86-94. doi:10.3141/1999-10

Liang, Y., Sun, Y., Zheng, R., & Huang, F. (2022). Efficient adversarial training without attacking: Worst-case-aware robust reinforcement learning. *Advances in Neural Information Processing Systems, 35*, 22547–22561.

Lopez, B. T., Slotine, J.-J. E., & How, J. P. (2019). Dynamic tube MPC for nonlinear systems. *2019 American Control Conference (ACC)*, (pp. 1655–1662).

Lu, S., Zhang, K., Chen, T., Başar, T., & Horesh, L. (2021). Decentralized policy gradient descent ascent for safe multi-agent reinforcement learning. *35*, 8767–8775.

Martín-Sacristán, D., Roger, S., Garcia-Roger, D., Monserrat, J. F., Spapis, P., Zhou, C., & Kaloxylos, A. (2020). Low-Latency Infrastructure-Based Cellular V2V Communications for Multi-Operator Environments With Regional Split. *IEEE Trans. Intell. Transp. Syst., 22*, 1052–1067.

Mayne, D. Q., & Kerrigan, E. C. (2007). TUBE-BASED ROBUST NONLINEAR MODEL PREDICTIVE CONTROL. *IFAC Proceedings Volumes, 40*, 36-41. doi:https://doi.org/10.3182/20070822-3-ZA-2920.00006

Miller, A., & Rim, K. (2020). Cooperative Perception and Localization for Cooperative Driving. 1256–1262.

Mun, H., Seo, M., & Lee, D. H. (2021). Secure Privacy-Preserving V2V Communication in 5G-V2X Supporting Network Slicing. *IEEE Trans. Intell. Transp. Syst.*

Munigety, C. R. (2018). Modelling behavioural interactions of drivers' in mixed traffic conditions. *Journal of Traffic and Transportation Engineering (English Edition), 5*, 284-295.

Pinto, L., Davidson, J., Sukthankar, R., & Gupta, A. (2017). Robust Adversarial Reinforcement Learning. In D. Precup, & Y. W. Teh (Ed.), *Proceedings of the 34th International Conference on Machine Learning. 70*, pp. 2817–2826. PMLR. Retrieved from https://proceedings.mlr.press/v70/pinto17a.html

Raimondo, D. M., Limon, D., Lazar, M., Magni, L., & ndez Camacho, E. F. (2009). Min-max Model Predictive Control of Nonlinear Systems: A Unifying Overview on Stability. *European Journal of Control, 15*, 5-21. doi:https://doi.org/10.3166/ejc.15.5-21

Sabouni, E., Ahmad, H. M., Cassandras, C. G., & Li, W. (2023). Merging Control in Mixed Traffic with Safety Guarantees: A Safe Sequencing Policy with Optimal Motion Control. *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, (pp. 4260-4265). doi:10.1109/ITSC57777.2023.10422265

Salvato, E., Fenu, G., Medvet, E., & Pellegrino, F. A. (2021). Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning. *IEEE Access, 9*, 153171–153187.

Sinha, R., Harrison, J., Richards, S. M., & Pavone, M. (2022). Adaptive Robust Model Predictive Control with Matched and Unmatched Uncertainty. *American Control Conference.* Retrieved from https://arxiv.org/abs/2104.08261

Treiber, M., Hennecke, A., & Helbing, D. (2000, February). Congested Traffic States in Empirical Observations and Microscopic Simulations. *Physical Review E, 62*, 1805-1824. doi:10.1103/PhysRevE.62.1805

Wang, J., Yang, S., An, Z., Han, S., Zhang, Z., Mangharam, R., . . . Miao, F. (2023). Multi-Agent Reinforcement Learning Guided by Signal Temporal Logic Specifications. *arXiv preprint arXiv:2306.06808*.

Wang, Z., Fang, M., Tomilin, T., Fang, F., & Du, Y. (2024). Safe Multi-agent Reinforcement Learning with Natural Language Constraints. *Safe Multi-agent Reinforcement Learning with Natural Language Constraints*. Retrieved from https://arxiv.org/abs/2405.20018

Wen, L., Duan, J., Li, S. E., Xu, S., & Peng, H. (2020). Safe reinforcement learning for autonomous vehicles through parallel constrained policy optimization. 1–7.

Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems, 35*, 24611–24624.

Zhang, Z., Han, S., Wang, J., & Miao, F. (2023). Spatial-temporal-aware safe multi-agent reinforcement learning of connected autonomous vehicles in challenging scenarios. 5574–5580.