Rationale-aware Autonomous Driving Policy utilizing Safety Force Field implemented on CARLA Simulator

Ho Suk* **Taewoo Kim*** **Hyungbin Park** **Pamul Yadav** **Junyong Lee** **Shiho Kim**
Yonsei University
{sukho93, boratw, phb88, pamul, jjunilee, shiho}@yonsei.ac.kr

Abstract

Despite the rapid improvement of autonomous driving technology in recent years, automotive manufacturers must resolve liability issues to commercialize autonomous passenger car of SAE J3016 Level 3 or higher. To cope with the product liability law, manufacturers develop autonomous driving systems in compliance with international standards for safety such as ISO 26262 and ISO 21448. Concerning the safety of the intended functionality (SOTIF) requirement in ISO 26262, the driving policy recommends providing an explicit rational basis for maneuver decisions. In this case, mathematical models such as Safety Force Field (SFF) and Responsibility-Sensitive Safety (RSS) which have interpretability on decision, may be suitable. In this work, we implement SFF from scratch to substitute the undisclosed NVIDIA's source code and integrate it with CARLA open-source simulator. Using SFF and CARLA, we present a predictor for claimed sets of vehicles, and based on the predictor, propose an integrated driving policy that consistently operates regardless of safety conditions it encounters while passing through dynamic traffic. The policy does not have a separate plan for each condition, but using safety potential, it aims human-like driving blended in with traffic flow.

1 Introduction

Guaranteeing the safety of self-driving is one of the essential factors in the autonomous driving system. According to the international standard, ISO 21448, titled Road vehicles – Safety of the intended functionality, the autonomous driving system can be divided into three subsystems: perception, planning, and actuation. As an agent which takes care of safe driving, the driving policy that implements the vehicle level safety strategy (VLSS) at the decision-making level, and the planning subsystem containing the policy are essential. To complete the core of the planning subsystem, developers design and implement driving policies reflecting their own strategy. In 2017, Intel Mobileye introduced Responsibility-Sensitive Safety (RSS) [3], and NVIDIA introduced Safety Force Field (SFF) [4] respectively. Those mathematical models guarantee to avoid going into unsafe situation, ultimately assuring safe driving. The models have two main advantages. First, serving as add-ons, they are compatible with existing driving policy. Second, they are mathematically proven, as a result, actions for safety of vehicle agent are explainable to human robustly.

From the 2010s, ADAS of SAE J3016 Level 2 began to spread widely in the passenger car market, and in 2018, Waymo launched a self-driving taxi service. However, automotive manufacturers in the private car market are struggling with liability issues of autonomous vehicles to achieve Level 3 or higher. It is because a decision failure of autonomous driving system higher than Level 3 could be a responsibility taken by not the driver but the automotive manufacturer who failed to detail the thorough specification for its autonomous driving system. Therefore, in order to respond to product

*These authors contributed equally to this work.

Machine Learning for Autonomous Driving Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), New Orleans, USA.

liability claims related to defects of the autonomous driving system, automotive manufacturers voluntarily comply with international standards for vehicle safety, such as ISO 26262 and ISO 21448, in the development process of autonomous vehicles. If an autonomous vehicle complying with these standards causes an accident, the automotive manufacturer could be freed from the product liability because it would be justified that the accident cannot be prevented even with state-of-the-art technologies which satisfy the safety standards. In the context of commercialization of autonomous vehicles, RSS and SFF, which are interpretable and mathematically rigorous, could be a candidate for the robust model substantiating that the safety of the intended functionality (SOTIF) of the planning subsystem is guaranteed in compliance with ISO 21448. Otherwise, without mathematical support [27], the reason for decision like "The trained network said so" cannot provide enough rational logic.

Intel released an open-source library called ad-rss-lib [25] that implements the RSS partially. Also, NVIDIA provides a software development kit called DriveWorks SDK that includes the SFF implementation [11] for approved users. Intel ad-rss-lib does not cover the whole scope of its paper, but it provides Python binding and CARLA [2] integration [33]. However, NVIDIA DriveWorks SDK is a non-public property, and it is implemented to integrate with its NVIDIA DRIVE platform equipped with NVIDIA DRIVE OS [18], so researchers have utilized it relatively little. In the basic example architectures suggested by Intel and NVIDIA to integrate the RSS and SFF with existing autonomous driving systems, RSS and SFF play a role of the last resort to prevent collisions of the autonomous vehicle by overriding a decision from the planning subsystem.

Based on the NVIDIA's conception, this work presents the SFF model implemented from scratch, integrating with open-source simulator CARLA. Furthermore, using the concept of claimed set and safety potential in the SFF implementation, we propose a method that integrates SFF into the planning subsystem to make a human-like driving policy that operates consistently whether safe or unsafe conditions regardless, eventually trying not to hinder the smooth traffic flow. Our work is different from NVIDIA's example, which adds SFF, a separate 'panic button' module, into the existing system architectures to prevent collisions.

2 Related Works

Responsibility-Sensitive Safety (RSS), devised by Intel Mobileye in 2017 and released as an open-source library in 2019, is a mathematical model formalizing the safety abiding by 5 common sense rules among drivers. Basically, RSS is built on the time-to-collision (TTC) measure.

Safety Force Field (SFF), devised by NVIDIA in 2017 and released only to permitted users as a software development kit in 2019, is also a mathematical model to guarantee the safety by trying to avoid unsafe situations. SFF is based on the measure for intersection of trajectories [14].

As explained in Intel's comparison table [21], the concepts of Intel RSS and NVIDIA SFF are quite similar, and even example architectures presented by the two companies are identical in larger scheme. In the example, RSS or SFF is an upper-level add-on module aiming for vehicular safety and it can work regardless of established driving policy structure. Having compatibility, the module operates in parallel with planning subsystem, and acts as a restrictor that limits actions derived from the driving policy. See Appendix A for more information about RSS, SFF, and an example architecture.

However, in this architecture, the performance of the RSS or SFF module entirely depends on the completeness of the driving policy that developers implement. If the driving policy is not designed elaborately, the RSS or SFF will suddenly intervene as a contingency plan only when an accident is imminent, while driving normally with existing driving policy under safe conditions. This dichotomous system has an advantage of being relatively easy to design, but it makes itself being hard to expect consistent and smooth driving like a human driver in continuously changing circumstance.

3 Claimed Set Predictor Learning and SFF Implementation

As described in Section 1, by virtue of its interpretability and mathematical rigor, SFF can provide an obvious mathematical basis for maneuver decision of planning subsystem. The SFF module included in the NVIDIA DriveWorks SDK depends on the NVIDIA DRIVE OS, and the source code is not publicly disclosed. To utilize with CARLA simulator and TensorFlow platform, we implemented the

SFF method from scratch. The details of SFF are described in NVIDIA's whitepaper. In Appendix B, we explain the core concept and definitions of SFF in a nutshell.

As explained in Appendix B, calculating the safety potential is the core of SFF, but the safety potential is derived from the claimed set, and the claimed set is derived from safety procedure. Therefore, according to NVIDIA's intention, the driving policy should be designed by developers in their own way first. Only after driving policy is completely implemented, SFF could be applied on autonomous driving system. In this context, to apply SFF, implementing the driving policy has the top priority.

Instead, in a different context, we tried to implement a driving policy using the safety potential of SFF. However, to obtain the claimed set required to calculate the safety potential, we fell into the contradiction that the safety procedure must already be implemented. To dispel the need to derive from a safety procedure, we propose a method learning the predictor for claimed sets. Taking advantage of CARLA simulator providing the ground truth of vehicle states, we trained the predictor network by supervised learning method. The predictor uses state vectors of vehicles and bird-eye view map image as inputs, and it outputs all vehicles' 2D actions: x-axis acceleration and y-axis acceleration that make up the claimed set. By covering the claimed set with a smooth function like mollifier, we make the claimed set differentiable and as a result, get the safety potential by calculating the intersection area between the two actors' claimed sets. In this way, without existing implemented safety procedure, we can obtain the claimed set of each actor vehicle and the safety potential. Our proposed system is described in Figure 1. See Appendix C for more claimed set prediction results.

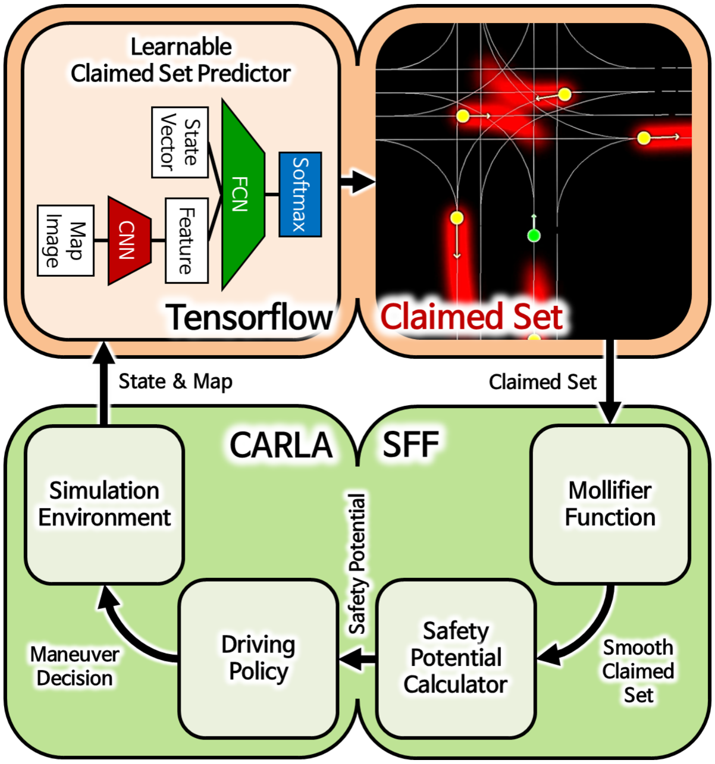


Figure 1: Simulation loop composed of proposed SFF-embedded CARLA. SFF integrates the claimed set predictor network for implementing the rationale-aware autonomous driving policy. (Red areas on upper-right image: other vehicles' claimed sets predicted by trained predictor network.)

4 SFF-Based Driving Policy Implementation

As described in Section 2, with the dichotomous system based on NVIDIA's basic example, the smoothness of maneuver comes from the completeness of driving policy. If the autonomous driving system has a crude driving policy, for instance, while cruising depending on the existing policy, even though there is a potential risk ahead, the SFF will intervene and take an abrupt action only at the moment when the risk is imminent, which is not ideal driving from human perspective [35].

Table 1: Experiment results. The number of arrivals represents efficiency, and the accident-free time represents safety. Each value is an average of 50 iterations. The higher the value, the better the performance of the corresponding model. (Bold: our result value.)

Aggression	Driving Policy	Number of arrivals	Accident-free time
No	No	0.26	406
	CARLA autopilot	0.92	1690
	RSS-CARLA implementation	0.30	2115
	SFF-CARLA implementation **(Ours)**	**1.40**	**2495**
Low	No	0.18	333
	CARLA autopilot	0.86	1506
	RSS-CARLA implementation	0.40	2427
	SFF-CARLA implementation **(Ours)**	**1.36**	**2287**
Intermediate	No	0.18	337
	CARLA autopilot	0.78	1436
	RSS-CARLA implementation	0.26	2173
	SFF-CARLA implementation **(Ours)**	**1.24**	**2050**
High	No	0.06	318
	CARLA autopilot	0.52	1332
	RSS-CARLA implementation	0.44	2594
	SFF-CARLA implementation **(Ours)**	**1.00**	**1886**

In Section 3, we proposed the learning of SFF claimed set predictor represented in a deep neural network and the implementation of safety potential calculation. Based on our implementations, we can get claimed sets of all vehicles on CARLA simulation, and using those claimed sets, we can calculate the safety potential of the ego-vehicle. To verify that our proposed model is implemented and trained correctly, using the calculated safety potential, we can eventually make an integrated driving policy (safety procedure) that operates consistently without having two separate modules dealing with safe condition and unsafe condition respectively. Our integrated driving policy ultimately aims to be human-like, without being too dogmatic or passive to go with traffic flow smoothly [22, 28].

To evaluate whether our SFF implementation is equivalent to existing models, we compare the performance of our driving policy with CARLA autopilot and RSS implementation on CARLA. We test an autonomous driving agent with our claimed set predictor to drive on road where other vehicles with random aggression are running. The random aggression means that sometimes other vehicles do not care about their surroundings and traffic signals according to probability hyperparameters. We divide the random aggression into 4 levels. In this experiment, we test how quickly and safely the agent arrives at destination without interfering with traffic flow on CARLA. In given period, we count the number of arrivals at randomly designated destinations and the accident-free timesteps of ego-vehicle. The results are represented in Table 1. See Appendix D for experiment details.

5 Discussion

In this work, we implement SFF, a mathematically proven model that could comply with ISO 21448 SOTIF, from scratch, and integrate it with CARLA simulator. Our method using the claimed set predictor does not need to care about an implementation method of existing driving policy to get claimed sets of vehicles. Using the claimed set predictor, we present an integrated driving policy that does not utilize an extra safety module. We verify that our driving policy based on predictor network shows competitiveness on safety compared with RSS by experiments at CARLA.

In future work, it might be possible to train not only the claimed set predictor but also the driving policy [37] by reinforcement learning (RL) [13, 29, 43], using the safety potential as a reward. The meta RL and the multi-agent RL [1, 23, 38] should also be considered to improve the driving policy for full automation. Furthermore, to specify the operational design domain (ODD) of autonomous driving system, it is essential to verify the system [20, 39] by detailed scenarios [6, 12, 24, 30, 31, 36], not randomly generated destination scripts.

Acknowledgments and Disclosure of Funding

This work was supported by the Institute for Information Communications Technology Planning Evaluation (IITP) grant funded by the Ministry of Science and ICT (MSIT, Korea, No.2021-0-01352, Development of technology for validating autonomous driving services in perspective of laws and regulations). And graduate student activities in this work were partly supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE)(P0020535, The Competency Development Program for Industry Specialist).

References

[1] S. Shalev-Shwartz, S. Shammah, and A. Shashua. *Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving*. arXiv preprint arxiv:1610.03295, 2016.

[2] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. *CARLA: An Open Urban Driving Simulator*. 2017 Conference on Robot Learning (CoRL). PMLR, 2017.

[3] S. Shalev-Shwartz, S. Shammah, and A. Shashua. *On a Formal Model of Safe and Scalable Self-driving Cars*. arXiv preprint arXiv:1708.06374, 2017.

[4] D. Nister, HL. Lee, J. Ng, and Y. Wang. *The Safety Force Field*. 2017. URL https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/the-safety-force-field.pdf

[5] P. Junietz, W. Wachenfeld, K. Klonecki, and H. Winner. *Evaluation of Different Approaches to Address Safety Validation of Automated Driving*. 2018 International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.

[6] A. Shashua, S. Shalev-Shwartz, and S. Shammah (Intel). *Implementing the RSS Model on NHTSA Pre-Crash Scenarios*. 2018. URL https://static.mobileye.com/website/corporate/rss/rss_on_nhtsa.pdf

[7] B. Yi, P. Bender, F. Bonarens, and C. Stiller. *Model Predictive Trajectory Planning for Automated Driving*. IEEE Transactions on Intelligent Vehicles, 4(1): 24-38, 2018.

[8] Y. Hu, W. Zhan, and M. Tomizuka. *Probabilistic Prediction of Vehicle Semantic Intention and Motion*. 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018.

[9] C. E. Tuncali, G. Fainekos, H. Ito, and J. Kapinski. *Simulation-based Adversarial Test Generation for Autonomous Vehicles with Machine Learning Components*. 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018.

[10] A. Corso, P. Du, K. Driggs-Campbell, and M. J. Kochenderfer. *Adaptive Stress Testing with Reward Augmentation for Autonomous Vehicle Validation*. 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.

[11] D. Nister, HL. Lee, J. Ng, and Y. Wang. *An Introduction to the Safety Force Field*. 2019. URL https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/an-introduction-to-the-safety-force-field-v2.pdf

[12] T. Ponn, C. Gnandt, and F. Diermeyer. *An Optimization-Based Method to Identify Relevant Scenarios for Type Approval of Automated Vehicles*. Proceedings of the ESV—International Technical Conference on the Enhanced Safety of Vehicles, Eindhoven, The Netherlands. 2019.

[13] S. Nageshrao, E. Tseng, and D. Filev. *Autonomous Highway Driving using Deep Reinforcement Learning*. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 2019.

[14] N. Alarcon. *DRIVE Labs: Eliminating Collisions with Safety Force Field*. 2019. URL https://developer.nvidia.com/blog/drive-labs-eliminating-collisions-with-safety-force-field/

[15] I. Bae, J. Moon, and S. Kim. *Driving Preference Metric-Aware Control for Self-Driving Vehicles*. International Journal of Intelligent Engineering and Systems 12(6): 157-166, 2019.

[16] A. Sadat, M. Ren, A. Pokrovsky, YC. Lin, E. Yumer, and R. Urtasun. *Jointly Learnable Behavior and Trajectory Planning for Self-Driving Vehicles*. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019.

[17] S. Vaskov, H. Larson, S. Kousik, M. Johnson-Roberson, and R. Vasudevan. *Not-at-Fault Driving in Traffic: A Reachability-Based Approach*. 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.

[18] NVIDIA. *NVIDIA DRIVE*. 2019. URL https://on-demand.gputechconf.com/gtc-cn/2019/pdf/CN9618/presentation.pdf

[19] W. Ding, J. Chen, and S. Shen. *Predicting Vehicle Behaviors Over An Extended Horizon Using Behavior Interaction Network*. 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.

[20] C. E. Tuncali, G. Fainekos, D. Prokhorov, H. Ito, and J. Kapinski. *Requirements-Driven Test Generation for Autonomous Vehicles with Machine Learning Components*. IEEE Transactions on Intelligent Vehicles, 5(2): 265-280, 2019.

[21] Intel. *RSS Concept SFF - Nvidia RSS – Mobileye - Intel*. 2019. URL https://newsroom.intel.com/wp-content/uploads/sites/11/2019/03/Intel-SFFvsRSS-table.pdf

[22] M. Naumann, H. Konigshof, M. Lauer, and C. Stiller. *Safe but not Overcautious Motion Planning under Occlusions and Limited Sensor Range*. 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019.

[23] Y. C. Tang. *Towards Learning Multi-Agent Negotiations via Self-Play*. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. 2019.

[24] S. Vaskov, S. Kousik, H. Larson, F. Bu, J. Ward, S. Worrall, M. Johnson-Roberson, and R. Vasudevan. *Towards Provably Not-at-Fault Control of Autonomous Robots in Arbitrary Dynamic Environments*. arXiv preprint arXiv:1902.02851, 2019.

[25] B. Gassmann, F. Oboril, C. Buerkle, S. Liu, S. Yan, M. S. Elli, I. Alvarez, N. Aerrabotu, S. Jaber, P. van Beek, D. Iyer, and J. Weast. *Towards Standardization of AV Safety: C++ Library for Responsibility Sensitive Safety*. 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019.

[26] M. Barbier, A. Renzaglia, J. Quilbeuf, L. Rummelhard, A. Paigwar, C. Laugier, A. Legay, J. Ibanez-Guzman, and O. Simonin. *Validation of Perception and Decision-Making Systems for Autonomous Driving via Statistical Model Checking*. 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019.

[27] L. Li, N. Zheng, and FY. Wang. *A Theoretical Foundation of Intelligence Testing and Its Application for Intelligent Vehicles*. IEEE Transactions on Intelligent Transportation Systems, 22(10): 6297-6306, 2020.

[28] R. Emuna, A. Borowsky, and A. Biess. *Deep Reinforcement Learning for Human-Like Driving Policies in Collision Avoidance Tasks of Self-Driving Cars*. arXiv preprint arXiv:2006.04218, 2020.

[29] A. Baheri, S. Nageshrao, HE. Tseng, I. Kolmanovsky, A. Girard, and D. Filev. *Deep Reinforcement Learning with Enhanced Safety for Autonomous Highway Driving*. 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2020.

[30] C. Neurohr, L. Westhofen, T. Henning, T. de Graaff, E. Mohlmann, and E. Bode. *Fundamental Considerations around Scenario-Based Testing for Automated Driving*. 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2020.

[31] A. Calo, P. Arcaini, S. Ali, F. Hauer, and F. Ishikawa. *Generating Avoidable Collision Scenarios for Testing Autonomous Driving Systems*. 2020 IEEE International Conference on Software Testing, Validation and Verification (ICST). IEEE, 2020.

[32] L. Wang, C. F. Lopez, and C. Stiller. *Generating Efficient Behaviour with Predictive Visibility Risk for Scenarios with Occlusions*. 2020 IEEE International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2020.

[33] B. Gassmann. *INTEGRATION OF RSS (Responsible Sensitive Safety)*. 2020. URL https://drive.google.com/file/d/1whREmrCv67fOMipgCk6kkiW4VPODig0A/view

[34] T. Stahl, M. Eicher, J. Betz, and F. Diermeyer. *Online Verification Concept for Autonomous Vehicles – Illustrative Study for a Trajectory Planning Module*. 2020 IEEE International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2020.

[35] I. Bae, J. Moon, J. Jhung, H. Suk, T. Kim, H. Park, J. Cha, J. Kim, D. Kim, and S. Kim. *Self-Driving like a Human driver instead of a Robocar: Personalized comfortable driving experience for autonomous vehicles*. arXiv preprint arXiv:2001.03908, 2020.

[36] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick, and F. Diermeyer. *Survey on Scenario-Based Safety Assessment of Automated Vehicles*. IEEE Access 8: 87456-87477, 2020.

[37] C. Zhao, L. Li, Z. Li, FY. Wang, and X. Wu. *A comparative study of state-of-the-art driving strategies for autonomous vehicles*. Accident Analysis & Prevention 150: 105937, 2021.

[38] Z. Zhu, and H. Zhao. *A Survey of Deep RL and IL for Autonomous Driving Policy Learning*. IEEE Transactions on Intelligent Transportation Systems, 2021.

[39] X. Xu, X. Wang, X. Wu, O. Hassanin, and C. Chai. *Calibration and evaluation of the Responsibility-Sensitive Safety model of autonomous car-following maneuvers using naturalistic driving study data*. Transportation Research Part C: Emerging Technologies 123: 102988, 2021.

[40] T. Nyberg, C. Pek. L. Dal Col, C. Noren, and J. Tumova. *Risk-aware Motion Planning for Autonomous Vehicles with Safety Specifications*. 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021.

[41] F. Oboril, and KU. Scholl. *RSS+: Pro-Active Risk Mitigation for AV Safety Layers based on RSS*. 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021.

[42] M. Won, and S. Kim. *Simulation Driven Development Process Utilizing Carla Simulator for Autonomous Vehicles*. 2022 International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH). 2022.

[43] Z. Cao, S. Xu, X. Jiao, H. Peng, and D. Yang. *Trustworthy safety improvement for autonomous driving using reinforcement learning*. Transportation Research Part C: Emerging Technologies 138: 103656, 2022.

A RSS and SFF

Responsibility-Sensitive Safety (RSS) compares a dangerous time t of the ego vehicle with the danger threshold time t_b^{long}, t_b^{lat} on both longitudinal and lateral side. If the threshold is reached, RSS judges it as a dangerous situation and applies a proper response that follows the constraint on the speed with either a longitudinal or lateral acceleration. In other words, the threshold could be represented in a trajectory set polygon. If the trajectory sets between ego vehicle and other road user are intersected, RSS chooses one of the following three decisions to restore the safe condition: brake or continue forward or drive away.

Safety Force Field (SFF) says that if actors follow the safety procedure, which is a family of control policies, the safety potential ρ_{AB} that quantifying the risk does not increase anymore, so it can be guaranteed that the actors will not cause unsafe situation eventually. This can be proved mathematically by a chain rule for the safety potential. In short, the method of RSS is to minimize the intersection between actors' claimed sets which is an union of trajectories resulting from the each actor's safety procedure. Key definitions are described in Appendix B.

Intel RSS or NVIDIA SFF is added as an add-on module harmonizing with existing subsystems. It receives both the world reconstruction data from perception subsystem and the maneuver decision from planning subsystem. For ego-vehicle's safety, as an upper-level restrictor, it could override received decision and pass a limited decision to actuation subsystem. A basic example architecture with RSS or SFF suggested by Intel and NVIDIA is described in Figure 2.

B Definitions in SFF

State $\boldsymbol{x_A(t)}$ is a vector containing position (2D or 3D), direction, and velocity of the vehicle actor A.

Control Policy $\boldsymbol{\frac{dx_A}{dt} = f(x_w, t)}$ is a smooth and bounded function of differentiating the state $x_A(t)$ with respect to time t.

Safety Procedure $\boldsymbol{S_A = \{\frac{dx_A}{dt}\}}$ is family of control policies $\frac{dx_A}{dt}$.

Claimed Set $\boldsymbol{C_A(x_A)}$ is a union of trajectories acquired by the safety procedure S_A. It is covered by a smooth function.

Safety Potential $\boldsymbol{\rho_{AB}}$ is a non-negative measure of intersection between the claimed sets $C_A(x_A)$ and $C_B(x_B)$ of actor A and B. It is a bump function where $\rho_{AB} = 0$ if there is no intersection

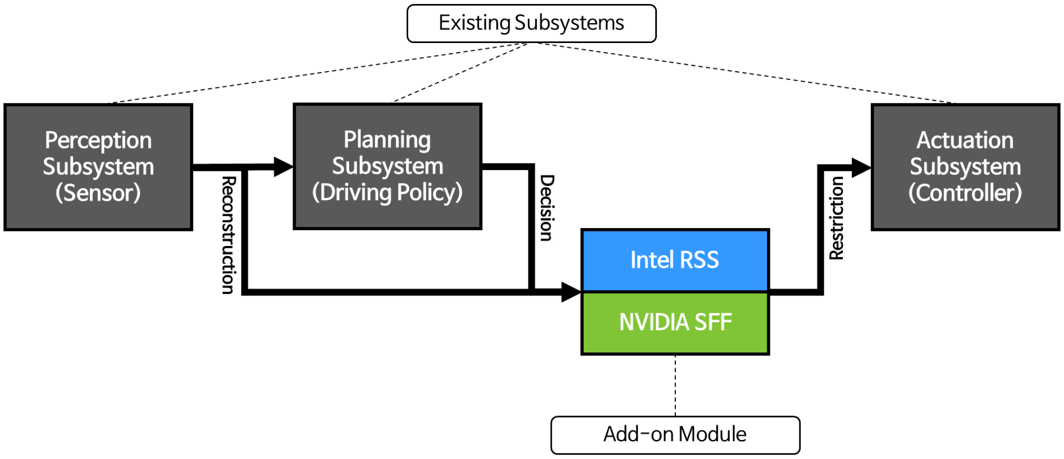
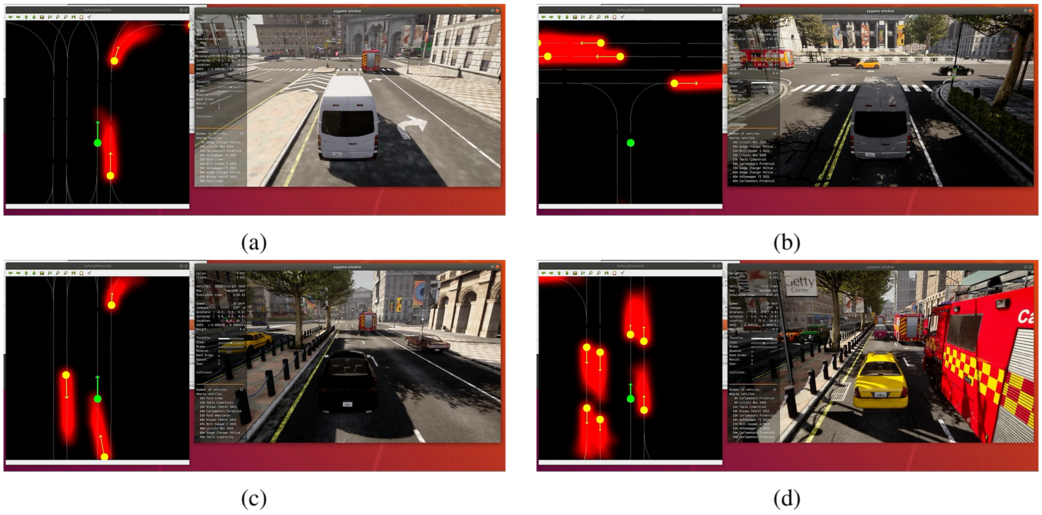

Existing Subsystems
Perception Subsystem (Sensor)
Planning Subsystem (Driving Policy)
Actuation Subsystem (Controller)
Reconstruction
Decision
Restriction
Intel RSS
NVIDIA SFF
Add-on Module

Figure 2: Basic example architecture with RSS or SFF. (Grey: existing subsystems. Blue/Green: RSS/SFF implementation as an add-on module.)

between claimed sets, and $\rho_{AB} > 0$ if intersection occurs. The safety potential can be defined as a dot product between claimed sets, which are smooth functions, and the dot product between real number functions is computed as an integral. In simple terms, safety potential is an area where the claimed sets from road users intersect.

Safety Potential $\boldsymbol{F_{AB} = -\frac{d\rho_{AB}}{dx_A}}$ is a negative gradient of safety potential ρ_{AB}. If actor A and B follow their respective safety procedures S_A and S_B, the safety potential ρ_{AB} does not increase anymore.

C Claimed Set Prediction Results


(a)

(b)

(c)

(d)

Figure 3: Visualization of claimed set prediction and corresponding scene of CARLA simulator. (Green dot: ego-vehicle's position. Yellow dots: other vehicles' positions. Red areas: other vehicles' claimed sets predicted by trained predictor network and smoothed by mollifier function.)

D Experiment Details

Random Aggression The aggression level is divided into 4 levels: No, Low, Intermediate, and High. Each level has different probability hyperparameters deciding the disregard level for two situations: surroundings during lane changing, and traffic signals during crossroad passing. The “No” level has the lowest probability, 0 exactly, and the “High” level has the highest.

Environment As an environment, we use the Town10 map provided by CARLA, a small town with 9 crossroads described in Figure 4. There are 50 other vehicles as non-player characters (NPC) on the road. The vehicle models include sedan, SUV, van, and fire engine, which have different dimensions and dynamics. Each vehicle drives to a random destination. While driving, they decide to change their lane with a fixed probability, and while lane changing, sometimes they ignore their surroundings according to the aggression level hyperparameter. They also ignore traffic lights occasionally according to the aggression level.

Figure 4: CARLA Town10 environment used for training and test in this work.

Iteration We run 50 iterations for each random aggression level environment for each decision model. Each iteration consists of 5000 timesteps, but if the ego-vehicle is involved in a collision, the iteration terminates immediately. Every iteration is initialized randomly.

Driving Policy We implement a driving policy using the safety potential, which depends on the learnable claimed set predictor. The longitudinal controller decides the acceleration based on the safety potential of ego-vehicle, and the lateral controller decides the steering angle trying to follow an imaginary center line.