Learning from Demonstrations of Critical Driving Behaviours Using Driver's Risk Field

Du, Yurui; Acerbo, Flavia Sofia; Kober, Jens; Son, Tong Duy

DOI
10.1016/j.ifacol.2023.10.1376

Publication date
2023

Document Version
Final published version

Published in
IFAC-PapersOnLine

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 2774–2779

Learning from Demonstrations of Critical Driving Behaviours Using Driver's Risk Field ⋆

Yurui Du [*,**] **Flavia Sofia Acerbo** [*] **Jens Kober** [**] **Tong Duy Son** [*]

[*] *Siemens Digital Industries Software, Leuven, Belgium, E-mail: flavia.acerbo@siemens.com; son.tong@siemens.com*

[**] *Department of Cognitive Robotics, Delft University of Technology, Delft, the Netherlands; E-mail: duyurui10@gmail.com; j.kober@tudelft.nl*

Abstract: In recent years, imitation learning (IL) has been widely used in industry as the core of autonomous vehicle (AV) planning modules. However, previous IL works show sample inefficiency and low generalisation in safety-critical scenarios, on which they are rarely tested. As a result, IL planners can reach a performance plateau where adding more training data ceases to improve the learnt policy. First, our work presents an IL model using the spline coefficient parameterisation and offline expert queries to enhance safety and training efficiency. Then, we expose the weakness of the learnt IL policy by synthetically generating critical scenarios through optimisation of parameters of the driver's risk field (DRF), a parametric human driving behaviour model implemented in a multi-agent traffic simulator based on the Lyft Prediction Dataset. To continuously improve the learnt policy, we retrain the IL model with augmented data. Thanks to the expressivity and interpretability of the DRF, the desired driving behaviours can be encoded and aggregated to the original training data. Our work constitutes a full development cycle that can efficiently and continuously improve the learnt IL policies in closed-loop. Finally, we show that our IL planner developed with less training resource still has superior performance compared to the previous state-of-the-art.

Keywords: Autonomous vehicles, Learning and adaptation in autonomous vehicles, Trajectory and path planning

1. INTRODUCTION

Today, autonomous vehicles (AVs) worldwide are undergoing extensive road tests in the real world, and some have already been put into active service. However, level 4+ autonomous driving still remains a significant challenge due to the "long tail" of real-world driving events, meaning AVs can be unsafe in rarely occurring safety-critical scenarios (Jain et al., 2021). In the AV application stack, the motion planning module is one of the keys to solving this bottleneck as it determines the AV's driving policy. By learning from large-scale driving datasets of expert demonstrations, imitation learning (IL) has been exploited as the core planner in real-world traffic scenarios, such as unsigned rural roads (Pomerleau, 1989), highways (Bojarski et al., 2016), and urban driving (Hawke et al., 2020; Bansal et al., 2018; Scheel et al., 2021).

However, despite the growing use of IL in AVs' planning module, efficiently improving its safety in long-tail events is a difficult task. In our view, it can be decomposed into three sub-tasks, namely, data-efficient learning, generation of safety-critical scenarios, and data augmentation of critical driving behaviours. Our motivations are given as follows:

It is observed that IL models require an excessive amount of training resource in order to achieve capable, but sometimes unsafe driving behaviours due to the distributional shift between the training and validation distributions (Bansal et al., 2018; Scheel et al., 2021). To enhance training efficiency and driving safety, we utilise the spline parameterisation for the IL model's predicted trajectory as proposed by our previous work (Acerbo et al., 2021) and an offline expert query approach to mitigate the distributional shift (Scheel et al., 2021).

Validation of IL models under critical traffic scenarios is often missing in published research. Most IL models are validated with log-replay data, where the traffic agents' trajectories are logged, and the dynamic interactions between traffic agents are not considered. To address this problem, recent research proposed to build reactive simulations with traffic agents that respond to others (Bergamini et al., 2021; Suo et al., 2021; Tan et al., 2021; Igl et al., 2022). These works however mainly focused on generating traffic scenarios similar to the ones from the original dataset,

⋆ This work was carried out within the thesis of Yurui Du at Siemens. This project was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 956123, and the Flanders Innovation & Entrepreneurship – VLAIO funded project BECAREFUL.

10.1016/j.ifacol.2023.10.1376

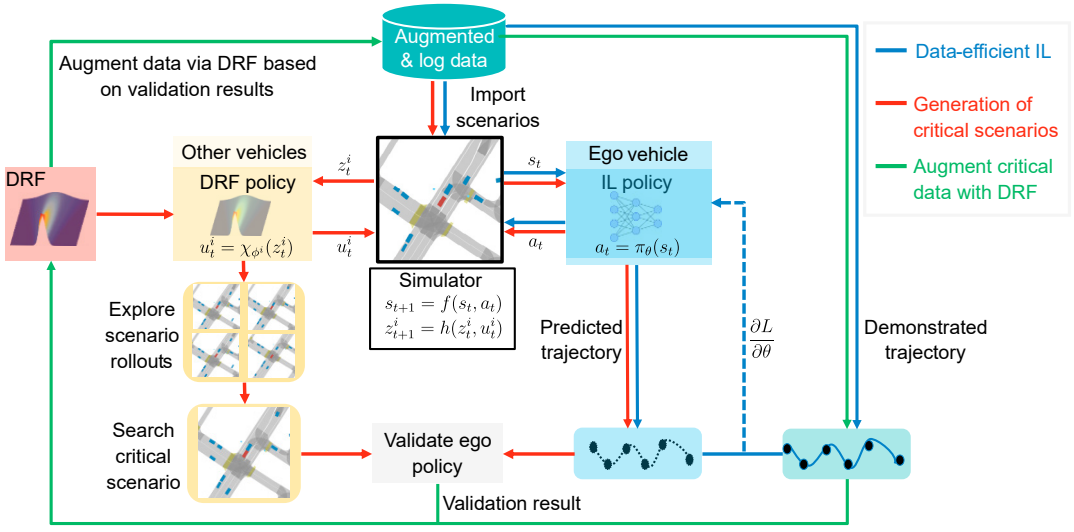

Fig. 1. Overview of our three-part work that respectively addresses a safe, efficient IL method, generation of critical scenarios for validation, and data augmentation encoding desired driving behaviours via DRF. All three parts constitute a development cycle that allows us to continuously improve IL policies in closed-loop.

which are not representative of critical traffic scenarios. Critical scenarios can also be manually designed by human experts, but this approach is not scalable to generate highly complex urban traffic scenarios with multiple traffic agents (Abeysirigoonawardena et al., 2019). Furthermore, the diverse driving styles of traffic agents in the real world are not considered, making the obtained critical scenarios not satisfactorily capture the diversity and complexity of real-world driving. To generate realistic, complex critical scenarios that can help us discover weak driving policies in validation, we employ driver's risk field (DRF) (Kolekar et al., 2020), a parametric model that represents human driving behaviours using the driver's subjective perceived risk of the environment. Compared to other driver models discussed in Sec. 2.3, DRF is a unified theory that allows us to represent different driving behaviours by tuning few parameters without switching between different models. We use DRF as traffic agents in a model-based multi-agent simulator based on the Lyft Prediction Dataset. By optimising the DRF parameters, critical traffic scenarios can be generated with realistic and diverse agents on a large scale.

Another bottleneck for IL is that increasing size of dataset does not necessarily improves IL models' robustness and safety (Scheel et al., 2021). This may indicate that IL models reach a performance plateau during training and stop learning from normal traffic data. To continuously improve the performance of IL models, we present a novel and flexible data augmentation method in which the DRF is exploited to encode desired driving behaviours in the original training data to purposefully improve poorly trained IL policies exposed in the validation results from critical scenarios.

Our contributions are three-fold, as summarised in Fig. 1:

(1) An IL method combining the spline coefficient parameterisation with the closed-loop offline expert query approach for efficient training. We demonstrate its superior performance over existing methods by validating it in large urban driving datasets and our generated critical traffic scenarios.
(2) Scalable generation of realistic and critical traffic scenarios in an interactive DRF-based traffic simulator to test ego driving policies in validation.
(3) A novel data augmentation method encoding desired driving behaviours with DRF to continuously improve IL performance in both recorded scenarios from logged data and our generated critical scenarios.

2. RELATED WORKS

In this section, we first discuss imitation learning and useful approaches to improve performance. Second, limitations of previous works on critical scenario generation are presented. Third, our motivations for modelling traffic agents using DRF are given via a comparison of different human driving behaviour models.

2.1 Imitation learning

Compared to optimisation-based motion planners, IL is attractive for its scalability to integrate new functionalities by learning from expert demonstrations rather than optimising human engineered objective functions. Additionally, with the availability of large-scale driving datasets and continuously produced data during testing, IL is becoming a popular method for motion planning in AV industry. However, one of the major challenges is that IL often suffers from distributional shift, which is caused by the compounding error in the sequential decision-making process, such as motion planning for AVs. It leads the ego vehicle to unfamiliar scenarios that are not included in the training distribution. Eventually, the behaviour of the ego vehicle becomes completely unpredictable and unsafe due to large deviations from the demonstration.

In practice, many approaches have been proposed to mitigate the distributional shift and significantly improve IL performance. While these approaches may seem different,

they mostly mitigate the distributional shift by providing corrective actions during training so that the ego vehicle learns to recover from earlier deviations in the sequential decision-making process. One approach (Bansal et al., 2018) leverages simple behaviour cloning with data augmentation by adding perturbation noise to provide more robust driving policies. Similarly, another approach (Bojarski et al., 2016) tries to directly label the perturbed camera images with corrective actions to avoid drifting. However, these approaches generally depend on empirical experiences to engineer noise mechanisms before training. A more theoretically satisfying approach is the dataset aggregation (DAgger) (Ross et al., 2010), (Scheel et al., 2021), which generates the training distribution of corrective actions on the run and guarantees an ideal linear regret bound to mitigate the distributional shift. However, it also exerts a heavier computation burden. To improve training efficiency, spline parameterisation as a powerful representation of predicted trajectories for IL has been proposed (Acerbo et al., 2021).

2.2 Critical scenario generation

Prior to our work, critical scenario generation has been studied in (Abeysirigoonawardena et al., 2019). By manually assigning waypoints and multiple available actions for each agent to choose from during the rollout of their policies, the most critical scenario can be found using search algorithms. However, the excessive manual labour required and the heavy computation burden greatly impair its scalability because the time complexity of this approach grows exponentially with the number of traffic agents, designed waypoints and actions. Other works on traffic scenario generation and reactive simulation with interactive agents mostly used generative methods such as latent variable models (Suo et al., 2021), autoregressive models (Tan et al., 2021), and generative adversarial imitation learning (Bergamini et al., 2021), in order to capture the possibility of multiple futures. However, these works mainly focused on generating similar traffic scenarios or agents with similar driving policies as demonstrated in the original dataset. Therefore, the generated scenarios are not critical scenarios that are purposefully designed to challenge weak driving policies in validation.

2.3 Realistic traffic agent modelling

Human driving behaviour models can be categorised into non-parametric and parametric ones (Lefèvre et al., 2014). For non-parametric approaches, the driver model relies on a large amount of data to learn policies that behave like human driving. Whereas for the parametric models, the driver behaviour model is often built from prior expert knowledge to capture human driving features in mathematically analytical forms, in which the parameters can be identified by fitting the model to the given data.

The decision-making process of non-parametric models is often considered a blackbox, meaning their driving behaviours cannot be easily adjusted with guaranteed legality. Moreover, as diverse behaviours for different traffic agents need to be modelled individually to make realistic simulations, it will be extremely expensive to train non-parametric models for all driving behaviours.

Compared to non-parametric models, building parametric human driving behaviour models requires considerably smaller amount of parameters, and these parameters often have clear physical and mathematical meanings, making it easier to interpret and control behaviours of the model. However, most parametric models only work for specific driving scenarios, such as car following in free roads (Treiber et al., 2000) or multi-lane highways (Kesting et al., 2007). These fragmented methods are by nature flawed because real driving scenarios are highly complex. Therefore, it is difficult to identify all possible traffic scenarios and design smooth transitions for them.

For these reasons, the driver's risk field (DRF) (Kolekar et al., 2020), a parametric human driving behaviour model, is especially suitable for realistic agent modelling because:

(1) It provides the driver's subjective view of the driving risk in any given scenario.
(2) It can explain diverse driving behaviours with a unified theory.
(3) It has interpretable parameters tunable to mimic diverse driving behaviours.

3. METHODS

In this section, we first specify the formulation of our IL method for the ego vehicle. Then, the parametric modelling of other traffic agents using DRF is discussed. We propose our method to generate critical scenarios with DRF agents that act adversarially to challenge the IL policy in validation. Finally, we present a novel data augmentation method that encodes demonstrations of critical driving behaviours to purposefully improve weak IL policies exposed in critical scenarios.

3.1 Efficient IL with spline coefficient parameterisation and closed-loop offline expert query

IL is a supervised learning method that aims to directly mimic driving behaviours from expert demonstrations. In the context of IL, the expert policy is defined as $a_t^* = \pi^*(s_t)$, i.e., the mapping from an agent's states to its actions. We adopt a similar approach to the one proposed in (Scheel et al., 2021), which is similar to DAgger (Ross et al., 2010), but with better computational efficiency owing to an offline synthetic expert query rather than an active expert policy to aggregate training datasets. This offline expert query approach is achieved by a closed-loop training scheme. Assuming that the dataset D^* consists of N expert trajectories and each trajectory τ_i has the length of T steps, namely $D^* = \{\tau_i\}_{i=1}^N$, $\tau_i = \{(s_{i,j}, a_{i,j})\}_{j=1}^T$, we first sample the ego vehicle's current policy for K steps, which will lead the ego vehicle to unfamiliar scenarios due to the distributional shift. Then, the current policy is updated by minimising the above loss function in the remaining $T-K$ steps so the ego vehicle learns to recover from mistakes caused by the distributional shift. The parameter of the policy network is denoted by θ, which can be learnt by minimising the discounted cumulative expected loss with a discount factor of γ:

$$\hat{\theta} = \arg\min_{\theta} \mathbb{E}_{\tau\sim\pi^*} \sum_{t=K}^{T} \gamma^{t-K} L(\pi_\theta(s_t), a_t^*)\,. \tag{1}$$

In this work, $L(\pi_\theta(s_t), a_t^*) = ||\pi_\theta(s_t) - a_t^*||_1$ is the L1 distance between the learner's action $\pi_\theta(s_t)$ and the expert action a_t^*. Similarly to (Acerbo et al., 2021), spline coefficients are used to parameterise trajectories in the dataset D^* instead of using discrete waypoints for better safety, more stable long-horizon predictions and smoother trajectories. Furthermore, we show that this parameterisation greatly improves the training efficiency in Sec. 4.3.

The states s_t and actions a_t of the original expert trajectories are both denoted as a 3D vector (x, y, α) representing the position and orientation in the $SE(2)$ space. The corresponding n spline coefficients in all three directions can be expressed as a matrix $\mathrm{A}_{3\times n}$. By replacing a_t^* in (1) with $\mathrm{A}^*_{3\times n}$, our objective function can be rewritten as:

$$\hat{\theta} = \arg\min_\theta \mathbb{E}_{\tau\sim\pi^*} \sum_{t=K}^{T} \gamma^{t-K} L(\pi_\theta(s_t), \mathrm{A}^*_{3\times n}) \,. \quad (2)$$

3.2 Parametric agent modelling using DRF

The DRF builds the driver's subjective view of its surrounding environment as a 2D Gaussian distribution along the predicted path. The perceived risk is derived from DRF representing the driver's subjective view of the driving risk in traffic. It is a function of the ego vehicle's current velocity and steering angle $P_{risk}(v, \delta)$. Then, based on the risk threshold theory, the future velocity and steering angle are obtained by solving an optimisation problem to keep the perceived risk below the assigned threshold. For detailed formulations of the DRF, please refer to (Kolekar et al., 2020).

3.3 Critical scenario generation

In this part, we detail how to generate critical traffic scenarios with agents that follow DRF policies controlling their velocity profiles along their original trajectories. The agents are designed to react adversarially to the ego vehicle's driving policy by optimising the DRF parameters of agents. The traffic scenarios are initialised based on real-world urban driving data to improve the complexity and realism of the generated scenarios.

We assume $s_t = \{x_t, y_t, \alpha_t\}$ to be the state vector of the ego vehicle's pose at time t. This vector includes the 2D position and orientation of the vehicle w.r.t. the ego-centric reference frame at $t = 0$. Let $Z_t = \{z_t^i\}_{i=1}^M$ be the state vector that consists of the pose of all other M agent vehicles closest to the ego vehicle, where z_t^i is the state vector of the i^{th} agent vehicle's pose. Let us assume that $a_t = \pi_\theta(s_t)$ is the IL policy of the ego vehicle and $u_t^i = \chi_{\phi^i}(z_t^i)$ is the parametric policy of the i^{th} agent parameterised by ϕ^i, and also the dynamics model of the ego vehicle $s_{t+1} = f(s_t, a_t)$ and the agent $z_{t+1}^i = h(z_t^i, u_t^i)$. We can obtain ϕ^i for each agent's parametric policy leading to critical traffic scenarios by optimising the following objective:

$$\Phi^* = \arg\min_\Phi J(\theta, \Phi) \,, \quad (3)$$

where $\Phi = \{\phi^i\}_{i=1}^M$ is the vector of agents' parameters and $J(\theta, \Phi)$ is the cost-to-go function computed from the scenario via unrolling all vehicles' policies. The cost is computed from the L1 distance between the ego vehicle and other vehicles and the total number of accidents (collisions, off-road incidents) to encourage the formation of dense traffic and collisions:

$$J(\theta, \Phi) = \mathbb{E}_{s_t, Z_t} \sum_{t=0}^{T} L1(s_t, Z_t) - L_{accidents} \,. \quad (4)$$

To ease the computation burden, we assume that each agent can either drive aggressively or cautiously represented by different values of DRF parameters. For every scenario, DRF controls a total of M agents, which means that there are 2^M different combinations of agents' parameters that lead to 2^M possible futures. Therefore, the optimal combination corresponding to the most critical traffic scenarios can be obtained with an exhaustive search algorithm. To scalably generate critical scenarios, Simcenter HEEDS (Siemens, 2022), a high-performance, global design exploration and optimisation software, is used to optimise the parameters of DRF. The algorithm of generating critical scenarios is shown in Alg. 1.

Algorithm 1 Generate critical scenarios in a model-based multi-agent simulator with DRF

```
θ ← θ_0 // ego vehicle's policy
for k = 1, ..., N do
    // for each traffic scenario
    // Φ^k are agents' parameters of their DRF policies in the k^th scenario
    // Exhaustively search 2^M combinations of agents' DRF parameters {Φ_j^k}_{j=1}^{2^M} to get the optimal combination of agents' parameters leading to the critical sceanario
    Φ^k* = arg min_{Φ_j} J(θ, Φ_j^k)
    // J is computed via (4) by unrolling agents' policies
    // Validate ego policy θ in the k^th scenario with M DRF agents parameterised by Φ^k*
end for
return validation results from critical scenarios
```

3.4 Data augmentation for desired driving behaviours

Improving the performance of IL models for AVs is difficult, as adding more training data does not guarantee better performance. To address this problem, here we propose to augment the expert demonstrations by using DRF to control the ego vehicle's velocity profiles along their original trajectories, with different desired driving behaviours encoded in different DRF parameters. This method offers great flexibility to encode desired driving behaviours we wish the IL model to learn. The DRF ensures the new learnt policy is still applicable to the previous dataset because the DRF-augmented data distribution is similar to human demonstrations.

Other data augmentation methods for IL planning models, such as perturbing the original trajectory with noise (Bansal et al., 2018), or requiring an expert policy during training (Acerbo et al., 2021), although they can significantly improve IL performance, they do not guarantee further improvement by retraining with more data. Furthermore, since they cannot be used to learn desired driving behaviours that purposely improve previous weak IL policies, performance usually worsens in critical scenarios where other agents act adversarially. By comparison,

our data augmentation method encoding desired driving behaviours can be used to continuously improve poorly trained policies exposed in critical scenarios by learning from DRF-augmented demonstrations.

4. EXPERIMENTS

In this section, we evaluate the three contributions of this paper. In particular, we are interested in: the impact of spline trajectory parameterisation on the training efficiency of IL models; the ability of generated critical scenarios to help detect poorly trained policies; and the potential of learning desired driving behaviours via retraining with DRF-augmented demonstrations.

4.1 Data

We use the Lyft Prediction Dataset to train and validate our IL models. Both log-replay and generated critical scenarios are used in validation. In log-replay scenarios, the other agents are following their original trajectories. While in critical ones, the other agents are reactive and following the DRF policy, which controls their velocity profiles along their original trajectories.

4.2 Metrics

We evaluate all models in closed-loop, meaning that the IL policy takes full control throughout the entire duration of each scenario. For each scenario, we measure the following metrics to keep track of the number of violations and events to compare the performance of different IL models.

(1) **Safety metrics**: Record the number of *collisions* if the ego vehicle collides with other traffic agents.
(2) **Imitation metrics**: Record the number of *off-road events* if the ego vehicle deviates from its ground-truth trajectory by more than 4m in the lateral direction.
(3) **Subjective risk metrics**: Record the number of *aggressive driving behaviours* if the perceived risk (as specified in Sec. 3.2) of the ego vehicle is larger than 10^5. This is a comprehensive metric that large risk values can mean a very close distance to other vehicles, making the driver feel more at risk.

4.3 Data-efficient IL with spline parameterisation

In this experiment, we analyse the impact of spline parameterisation on the training efficiency by comparing our data-efficient IL model (trained 30h by us with 1 NVIDIA RTX A4000 laptop GPU, where 1h is equivalent to 10000 iterations, with a batch size of 6 2-second trajectories, i.e., approx. 9 epochs) with Lyft Urban Driver (trained 30h by Lyft with 32 Tesla V100 GPUs, for 61 epochs). To ensure a fair comparison, our IL model shares the same network architecture, except for the dimension of the output of the last layer, as Urban Driver. Both models are trained using the same dataset and simulator from Lyft (Scheel et al., 2021). In Table 1, we show that even with significantly less training resources, our model outperforms Urban Driver in all metrics, indicating better performance in safety and imitation. Additionally, our model has a less aggressive driving style compared to Urban Driver.

Table 1. Metrics for the baseline and our model from 2500 4-second log-replay scenarios.

Models	**Collision**			**Imitation**	**Aggressive**
	Front	Rear	Side	Off-road	**driving**
Urban Driver	1	3	**0**	4	140
Ours	**0**	2	**0**	**0**	**109**

Table 2. Metrics for the baseline and our (re-trained) model from 1250 4-second log-replay and critical scenarios.

Scenarios	**Models**	**Collision**			**Imitation**	**Aggressive**
		Front	Rear	Side	Off-road	**driving**
Log-replay	Urban Driver	**0**	**1**	**0**	**0**	71
	Ours	**0**	**1**	**0**	**0**	**56**
	Ours(Re)	**0**	**1**	**0**	**0**	100
Critical	Urban Driver	**0**	8	**0**	6	71
	Ours	**0**	7	1	**0**	**60**
	Ours(Re)	**0**	**5**	**0**	**0**	111

4.4 Generation of critical traffic scenarios

In this experiment, we compare our generated critical traffic scenarios with adversarial DRF agents to original traffic scenarios with log-replay agents. By unrolling our ego IL policy and Lyft Urban Driver in both kinds of scenarios and comparing their performance, it is shown in Table 2 that our generated critical scenarios are more challenging for both IL models to handle as the number of collisions increases in critical scenarios.

In addition, more aggressive driving is observed in critical scenarios. This is because the distance between agents is smaller in critical scenarios. Therefore, the ego vehicle subjectively "feels" more at risk driving in critical scenarios.

4.5 Data augmentation for desired driving behaviours

Passiveness due to causal confusion is a common mistake of IL models (de Haan et al., 2019; Vitelli et al., 2022), meaning that the IL model may learn to drive passively and unresponsive to approaching vehicles. To mitigate passiveness of the ego vehicle, we retrain the IL model with DRF-augmented data where the ego vehicle drives more aggressively when the rear vehicles are approaching.

In this experiment, we compare the performance of our retrained IL model (30h training + 2h retraining), our IL model (30h training), and Lyft Urban Driver in both log-replay and critical traffic scenarios.

In Table 2, it is shown that our retrained IL model performs better in critical scenarios and equally in log-replay scenarios regarding the collision and imitation metrics. Our retrained model is observed to grasp more aggressive driving behaviours. Also, our retrained model is sufficiently robust to handle both critical and log-replay scenarios. More importantly, we show that driving styles of IL models can be properly customised using DRF without compromising driving safety.

Fig. 2 presents qualitative results comparing our IL model before and after retraining with DRF-augmented data to alleviate passiveness. In the top row, we see that our IL model before retraining has a rear collision due to passive driving. By contrast, in the bottom row, the retrained model speeds up in time and safely passes the intersection without noticeable sign of passiveness, indicating that it

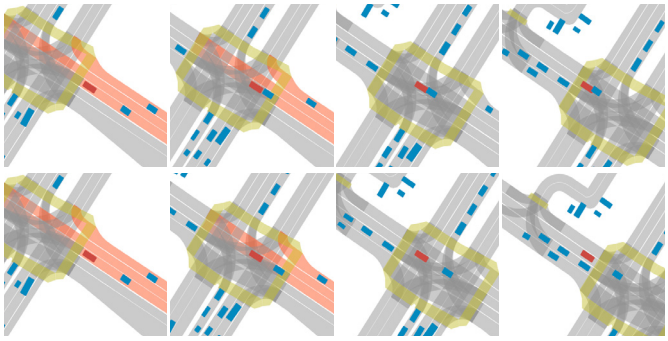

Fig. 2. Red rectangle is the ego vehicle controlled by our IL policy, and blue rectangles are other traffic agents controlled by DRFs leading to critical scenarios. Top row: Validation of our IL model in a critical scenario. Bottom row: Validation of our IL model retrained with augmented data in the same critical scenario.

has learnt the desired driving policy that drives slightly more aggressively if followed by a rear vehicle.

5. CONCLUSION

In this paper, we have demonstrated the potential of incorporating the DRF, a parametric human driving behaviour model, in a multi-agent traffic simulator to build a full development cycle that can continuously improve the performance of IL models. With the expressivity and interpretability of the DRF, we can generate critical scenarios with DRF-based agents that are parameterised to act adversarially to the ego IL policy. These generated critical scenarios are proven to be more challenging for the ego IL policy to handle than the recorded scenarios from the logged data. Moreover, weak policies are more easily detected from the validation with critical scenarios. To enhance the weak policy, we use DRF to encode desired driving behaviours to augment the expert demonstrations. By retraining the IL model with augmented data, the IL model achieves safer driving. The IL policy learnt via retraining is more robust as it is applicable to both critical scenarios with adversarial agents and recorded scenarios from logged data. We also show an improved (re)training efficiency by using the spline trajectory parameterisation. For future work, the theory and formulation of DRF should be studied more in-depth to further enhance the realism and diversity of generated critical scenarios.

REFERENCES

Abeysirigoonawardena, Y., Shkurti, F., and Dudek, G. (2019). Generating adversarial driving scenarios in high-fidelity simulators. In *2019 ICRA*, 8271–8277.

Acerbo, F.S., Alirezaei, M., der Auweraer, H.V., and Son, T.D. (2021). Safe imitation learning on real-life highway data for human-like autonomous driving.

Bansal, M., Krizhevsky, A., and Ogale, A.S. (2018). Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. *CoRR*, abs/1812.03079.

Bergamini, L., Ye, Y., Scheel, O., Chen, L., Hu, C., Pero, L.D., Osinski, B., Grimmett, H., and Ondruska, P. (2021). Simnet: Learning reactive self-driving simulations from real-world observations.

Bojarski, M., Testa, D.D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J., and Zieba, K. (2016). End to end learning for self-driving cars. *CoRR*, abs/1604.07316.

de Haan, P., Jayaraman, D., and Levine, S. (2019). Causal confusion in imitation learning. *CoRR*, abs/1905.11979.

Hawke, J., Shen, R., Gurau, C., Sharma, S., Reda, D., Nikolov, N., Mazur, P., Micklethwaite, S., Griffiths, N., Shah, A., and Kndall, A. (2020). Urban driving with conditional imitation learning. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 251–257.

Igl, M., Kim, D., Kuefler, A., Mougin, P., Shah, P., Shiarlis, K., Anguelov, D., Palatucci, M., White, B., and Whiteson, S. (2022). Symphony: Learning realistic and diverse agents for autonomous driving simulation. In *2022 ICRA*, 2445–2451.

Jain, A., Pero, L.D., Grimmett, H., and Ondruska, P. (2021). Autonomy 2.0: Why is self-driving always 5 years away?

Kesting, A., Treiber, M., and Helbing, D. (2007). General lane-changing model mobil for car-following models. *Transportation Research Record: Journal of the Transportation Research Board*, 1999, 86–94.

Kolekar, S., de Winter, J., and Abbink, D. (2020). Human-like driving behaviour emerges from a risk-based driver model. *Nature Communications*, 11, 4850.

Lefèvre, S., Sun, C., Bajcsy, R., and Laugier, C. (2014). Comparison of parametric and non-parametric approaches for vehicle speed prediction. In *2014 American Control Conference*, 3494–3499.

Pomerleau, D.A. (1989). *ALVINN: An Autonomous Land Vehicle in a Neural Network*, 305–313. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Ross, S., Gordon, G.J., and Bagnell, J.A. (2010). No-regret reductions for imitation learning and structured prediction. *CoRR*, abs/1011.0686.

Scheel, O., Bergamini, L., Wolczyk, M., Osiński, B., and Ondruska, P. (2021). Urban driver: Learning to drive from real-world demonstrations using policy gradients. In *5th Annual Conference on Robot Learning*.

Siemens (2022). www.plm.automation.siemens.com/global/en/products/simcenter/simcenter-heeds.html.

Suo, S., Regalado, S., Casas, S., and Urtasun, R. (2021). Trafficsim: Learning to simulate realistic multi-agent behaviors. In *2021 CVPR*, 10395–10404. IEEE Computer Society, Los Alamitos, CA, USA.

Tan, S., Wong, K., Wang, S., Manivasagam, S., Ren, M., and Urtasun, R. (2021). Scenegen: Learning to generate realistic traffic scenes. In *2021 CVPR*, 892–901.

Treiber, M., Hennecke, A., and Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations.

Vitelli, M., Chang, Y., Ye, Y., Ferreira, A., Wołczyk, M., Osiński, B., Niendorf, M., Grimmett, H., Huang, Q., Jain, A., and Ondruska, P. (2022). Safetynet: Safe planning for real-world self-driving vehicles using machine-learned policies. In *2022 ICRA*, 897–904.