Learning Safe and Human-like High-level Decisions for Unsignalized Intersections from Naturalistic Human Driving Trajectories

Lingguang Wang, Carlos Fernandez and Christoph Stiller

***Abstract*—Automated driving systems need to behave as human-like as possible, especially in highly interactive scenarios. In this way, the behavior can be better interpreted and predicted by other traffic participants, in order to prevent misunderstanding, and in the worst case, accidents. With this purpose, more and more human-driven trajectories in real traffic are recorded, making it possible to learn human-like driving styles. In this paper, we extend our previous behavior cloning approach, which has been successfully applied to highway driving, to generate high-level decisions for unsignalized intersections that are challenging during urban driving. Unlike many other approaches that utilize neural networks, either for end-to-end behavior cloning or for approximating Q-functions in reinforcement learning, where their decisions are intractable to understand, the output decisions of our approach are interpretable and easy to track. Meanwhile, the driving decisions are provably safe under reasonable assumptions by generalizing the Responsibility-Sensitive Safety (RSS) concept to complex intersections. Simulation evaluations show that our learned policy produces a more human-like behavior, and meanwhile, balances driving efficiency, comfort, perceived safety, and politeness better.**

***Index Terms*—Automated driving, decision making, Monte-Carlo Simulation, behavior cloning, Responsibility-Sensitive-Safety, unsignalized intersection**

I. Introduction & State of the Art

FUTURE transportation systems are the cornerstones to further improving the productivity of humans. Among them, autonomous driving aims to liberate human hands and has become one of the most important key technologies. However, self-driving systems are still expected to co-exist with human drivers on the road for decades before the street is fully automated. Therefore, it should also be one of the important goals of research to make autonomous vehicles imitate human driving behavior as much as possible. In this case, the passengers, other human drivers, and other human traffic participants can understand and better cooperate with autonomous vehicles. In addition, as the premise of the universal acceptance of automatic driving systems, providing provable safety is also crucial.

Many Reinforcement Learning (RL) approaches [1]–[3] have been applied to automated driving tasks. They assume an unpredictable or highly uncertain environment and try to learn a driving policy or a Q-function by letting the agent interact with this environment, with the goal of maximizing the accumulated future reward given one predefined reward function. One of the strengths of these approaches is that modeling other agents' behavior is not required, and the possible future interactions are embedded in the value functions or Q-functions given the state space. However, they always suffer from lacking explainability of the output decision and transferability to real-world scenarios. Moreover, designing a suitable reward function is not straightforward and can be affected a lot by the preferences of the developer.

In order to better imitate human driving behavior, Inverse Reinforcement Learning (IRL) approaches [4]–[6] try to infer the reward function from naturalistic human driving data. However, similar to RL, large neural networks are usually still required in order to coop with large state space. How the input data is processed and why a certain state-action pair outputs a higher Q-value is still hardly interpretable. Another challenge for these approaches is driving safety. Shuojie et al. [7] propose a RL-based method and Monte Carlo tree search algorithm to minimize unsafe behaviors but still without formal guarantees. Some other approaches [8], [9] utilize a safety layer that is easily verifiable to prevent executing unsafe commands from RL, but breaks the continuity of the actions and reduces the performance of the behavior as the output is constrained inside one rule-based envelop.

In urban driving, sensor occlusion builds another barrier to balancing safe and progressive decisions, which the previously mentioned learning-based methods hardly discussed. Kamran et al. [2] claim to provide actions that are less overcautious than the rule-based policy and safer than the usual RL-based approach, but still not collision-free. Hubmann et al. [10] propose a generic Partially Observable Markov Decision Process (POMDP) formulation that can be applied to various scenarios for urban driving. Its performance is demonstrated at unsignalized intersections with multiple vehicles and occlusions caused by static and dynamic objects. However, the run-time performance of this approach can still largely limit the choice of the state and action spaces.

In our previous publication [11], we propose one behavior cloning (BC) approach that produces provably safe and human-like behavior. The output decision is transparent to humans and can easily be tracked. Unlike most of other BC approaches that map the full state space or even images to low-level control commands with neural networks [12],

This work is accomplished within the project "UNICARagil" (FKZ 16EMO0287) and the financial support from the Federal Ministry of Education and Research of Germany (BMBF) is acknowledged.

Lingguang Wang, Carlos Fernandez and Christoph Stiller are with the Institute of Measurement and Control Systems, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany, lingguang.wang, carlos.fernandez, stiller@kit.edu

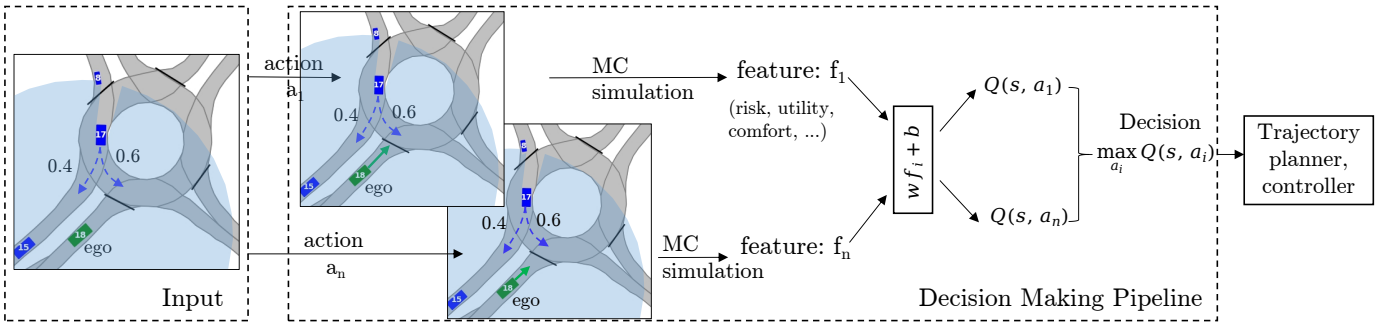


Fig. 1: Proposed decision making pipeline taking a roundabout scenario as an example.

[13], which can have the same issue as RL approaches, our approach only serves as a high-level decision-making module. The approach is proven to provide human-like decisions that balance efficiency, comfort, perceived safety, and politeness in lane changing, merging, and exiting scenarios on multi-lane roads.

In this paper, we generalize the BC approach to urban driving scenarios, especially the highly challenging unsignalized intersections, e.g. unprotected left turn, yielding, pedestrian crossing, roundabouts, etc. The pipeline is illustrated in Fig. 1. The minimum inputs of the decision making are the preception results of the environment and the self-localization, where noisy states (location, velocity, etc.) of other traffic participants within the sensing field of view (FoV) and of the ego vehicle are provided. In order to obtain traffic rules and information about the road topology, a High-Definition (HD) Map, e.g. lanelet2 [14], is provided as well. The output is one of the predefined high-level actions, which is further processed and refined by the trajectory planner and controller to be executed on the car. One optional input that could improve the planning quality is extra information about the intentions of other traffic participants, e.g. turning intention at intersections, exiting intention at roundabouts, etc.

The overall pipeline of the extended approach is summarized as follows: while approaching the intersection, several high-level action candidates are generated. Assuming the ego vehicle following a specific action, we use Monte-Carlo (MC) sampling (or MC Simulation) to generate possible future episodes of the current scene, where the uncertain behaviors and uncertain intentions of the known traffic participants and the existence probability of the phantom vehicles in occluded areas are considered. After performing a large number of MC simulations, the consequence of executing the action is estimated, where several indices (features) are used to characterize the action. We categorize the features into risk, utility, comfort, and politeness, and they will be explained in detail in later chapters. One linear function is utilized to compute the Q-value for each action with unknown parameters (weights w and bias b), and the final decision is the one that maximizes the Q-value.

Our method is essentially one BC approach because the goal is to update the weights w and bias b of the linear function, such that the Q-value of the more human-like actions is maximized, and the Q-values of other actions are minimized, similar to the classification problem.

This paper serves as a generalization and extension of the previous approach, and several novelties and contributions are made additionally:

- We extend the RSS concept [15] to unsignalized intersections for different type of conflict zones, different traffic participants, close consecutive intersections, and occluded intersections, etc.
- We rethink and extend the definition of risk that was proposed in [11] to cover emergency situations, not only the fall-back possibility.
- We propose a realistic environment modeling for unsignalized intersections and extend the MC simulations to consider possible phantom vehicles from occlusions.
- We further prove the generalizability of our approach by extending and applying it to unsignalized intersections.

II. Preliminary Formulations

A. RSS safety for Unsignalized Intersections

We first discuss the RSS safety concept that is crucial to be considered while making decisions at unsignalized intersections. The idea of this concept is to guarantee to never cause a collision, instead of never being involved in a collision. At unsignalized intersections, besides making sure to "not hit someone from behind", another two RSS "common sense" rules "right-of-way is given, not taken" and "be careful of areas with limited visibility" are essential as well.

1) Safety for Merging and Crossing Conflict Zones: Naumann [16] proposes to distinguish between crossing and merging conflict zones and define different safety rules. Fig. 2 presents one example at an unsignalized intersection, where the ego vehicle should give way to the oncoming prioritized vehicles. From all their routing options of the prioritized vehicles, two of them intersect with the route of the ego vehicle (green dashed line), resulting in a crossing and a merging conflict zone.

The formulation in [16] can not be generalized for our scenarios. Therefore, we reformulate his proposal with additional notions to enable a better understanding but omit the mathematical details.

For crossing conflict zones, the safety of the non-prioritized vehicle (ego) can be ensured when at least one of the conditions C_1 and C_2 is held:

- C_1: Ensure to be able to stop before the conflict zone.

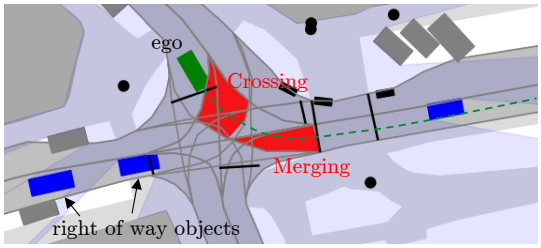


Fig. 2: Crossing and Merging conflict zones between the ego vehicle and the prioritized vehicles.

- C_2: Ensure to safely pass the conflict zone.

To satisfy C_1, ego vehicle needs to be able to stop before the conflict zone with less than its maximum deceleration $a_{\text{max,decel,ego}}$. There are two ways to satisfy C_2, which are:

- $C_2(a)$: At the time ego vehicle enters the conflict zone with its maximum ability, i.e. maximum acceleration $a_{\text{max,accel,ego}}$ until the allowed speed limit v_{limit}, the prioritized vehicle is at sufficient distance, i.e. its required deceleration to stop in front of the conflict zone is acceptable, e.g. less than its $a_{\text{soft,decel,obj}}$.
- $C_2(b)$: Ego vehicle can guarantee to have left the conflict zone (with its maximum ability) for a predefined time of zone clearance (TZC) $t_{\text{TZC,min}}$, before the prioritized vehicle can enter it with its maximum acceleration $a_{\text{max,accel,obj}}$ until its assumed maximum velocity $v_{\text{max,obj}}$.

For merging conflict zones, C_1 still holds. However, the conflict zone can never be actually "passed", as its length is potentially unlimited. Therefore, C_3 is proposed instead of C_2:

- C_3: From the time the ego vehicle enters the conflict zone with its maximum ability, the distance of the prioritized vehicle to the ego vehicle always remains larger than the RSS safe following distance. In addition, the prioritized vehicle shall not brake more than $a_{\text{soft,decel,obj}}$.

In other words, the ego vehicle becomes the leading vehicle of the prioritized vehicle after merging. If the prioritized vehicle only needs to brake slightly in order to maintain RSS safe distance to the ego vehicle, the merging is considered as not impeding and rude. RSS safety for merging conflict zone is similar to the on-ramp merging that is comprehensively discussed in [11], where the mathematical details of one general approach and an extended version are presented.

We use the recommended RSS parameters from [11]. As for the additional parameter $t_{\text{TZC,min}}$, $t_{\text{TZC,min}} = 0$s is sufficient for safety. However, drivers of the prioritized vehicles might still feel endangered. Therefore, we analyzed 4057 crossing scenarios and the associated TZCs from the inD dataset [17]. The TZC is computed for each vehicle with the assumptions in $C_2(b)$ when C_1 is not possible anymore. As a result, 81.4% of the vehicles cross with more than 0.5s TZC and 61.6% with more than 1s. We select $t_{\text{TZC,min}} = 0.5$s as 81.4% of the crossing drivers regard it as not making prioritized drivers feel endangered.

2) Safety for Cyclists and Pedestrian Crossing: Cyclists driving on the vehicle lanes are treated as vehicles when considering RSS safety, but with different RSS parameters (e.g. maximum deceleration), as their dynamics differ from vehicles. Those located on the walkway will be treated as normal pedestrians. Therefore, there is no need to introduce specific RSS safety rules for cyclists.

Pedestrian Crossing belongs to crossing conflict zones, where C_1 and C_2 apply, prioritizing pedestrians instead of vehicles. As pedestrians have different dynamics, we reformulate C_2:

- $C_{2,\text{p}}$: Ego vehicle can guarantee to have left the conflict zone (with its maximum ability) for a predefined time of zone clearance (TZC) $t_{\text{TZC,min}}$, before the prioritized pedestrian can enter it with its maximum velocity $v_{\text{max,p}}$.

Note that pedestrians are assumed to be able to accelerate to $v_{\text{max,p}}$ in an infinitely short time. The maximum recorded velocity of pedestrians in inD dataset is $15.07\frac{\text{m}}{\text{s}}$, which is obviously an outlier. Therefore, we select $v_{\text{max,p}} = 5.02\frac{\text{m}}{\text{s}}$ at 0.99 percentile.

3) Stop Line and Relevant Conflict Zones: We use lanelet2 maps where traffic rules are well-defined and encoded in regulatory elements. For unsignalized intersections, right-of-way regulatory elements control the traffic rules and decide who has priority and who needs to yield. One right-of-way regulatory element has the following components:

- *Stop line*: before which the yielding vehicle is recommended to stop, such that no conflict zone is impeded.
- *Traffic sign* (optional): related traffic sign, e.g. yielding sign, stop sign, etc.
- *Yielding lanelet*: vehicles tending to pass through this lanelet should yield to prioritized traffic participants.
- *Right-of-way lanelets*: traffic participants that are possible to pass one of the right-of-way lanelets have priority over the yielding vehicles.

In the example of Fig. 2, the left turning lanelet of the ego vehicle should be the yielding lanelet, and the oncoming lanelets of the west arm and the south arm are right-of-way lanelets. If there is an oncoming vehicle from the south arm, it will create two more conflict zones with the ego vehicle. The conflict zones can be inferred by checking intersecting areas between possible succeeding lanelets of the yielding lanelet and of the right-of-way lanelets.

In the example of Fig. 2, if the ego vehicle tends to go straight or turn right, the traffic rules will be completely different. In the former case, only the oncoming lanelets of the west arm will be the right-of-way lanelets for the going-straight yielding lanelet. In the latter case, the ego vehicle does not need to yield to any vehicles.

4) Supplemental Constraints for Maximum Ability: It was not discussed in [16] how the maximum ability of the ego vehicle in C_2 and C_3 can be affected. In the author's opinion, there are two important affecting sources.

The first one is when the ego vehicle has one leading vehicle. If the leading vehicle is possible to come to a full stop at or right after the conflict zone, the time for the ego vehicle to reach or leave the conflict zone could be potentially infinite. In this case, C_2 and C_3 can not be satisfied at all.

The second one is in the case of two close consecutive stop lines or regulatory elements. In the example of Fig. 2, there is one zebra crossing right after the conflict zones. If

safely passing the pedestrian crossing is not guaranteed, the maximum ability of the ego vehicle is additionally limited by "being able to stop before the zebra crossing", when examining safety for the two red conflict z ones. T his g uarantees that when the ego vehicle tries its best to reach or pass the red conflict z ones i n order t o s atisfy C_2 a nd C_3, t he s peed i s still sufficiently l ow, s uch t hat s topping b efore t he n ext pedestrian crossing still remains possible. Therefore, we introduce $C^*_{2,3}$ supplemental to $\{C_2, C_{2,p}, C_3\}$ (notation for one of the C_2, $C_{2,p}$ and C_3 depending on the type of the conflict zone):

- $C^*_{2,3}(a)$: If ego vehicle has a leading vehicle, the maximum ability of the ego vehicle in C_2, $C_{2,p}$ and C_3 is limited by maintaining a minimum RSS safe distance to the leading vehicle, while the leading vehicle decelerates with its maximum deceleration $a_{\text{max,decel,lead}}$.
- $C^*_{2,3}(b)$: If the next stop line is close to the current one, and safely passing all the conflict zones of the next stop line is not guaranteed, the maximum ability of the ego vehicle is additionally limited by being able to safely stop before the first conflict zone of the next stop line with $a_{\text{max,decel,ego}}$.

5) Safety Considering Occlusions: Under occlusion, the information about whether prioritized traffic participants exist in occluded areas is lost. It is proposed in [18] to over-approximate the possible states in occluded road sections, which provides provable safety even in the worst case. However, the resulting driving behavior is overly conservative. In [19], the occluded road sections are tracked through time, and the possible state intervals of the hidden traffic participants are significantly reduced, which guarantees the same safety but allows a more efficient driving behavior.

In this paper, we utilize the approach in [19] to check the RSS safety against $\{C_2, C_{2,p}, C_3\}$. It is assumed that the worst-case traffic participants from the possible state intervals in occlusions are at the sensing edges. Thus, when passing a strongly occluded intersection, $\{C_2, C_{2,p}, C_3\}$ are only able to be satisfied when the FoV of the ego vehicle covers far enough of the prioritized roads.

B. Action Space and Basic Policy

In this paper, we only make longitudinal high-level decisions. Obstacle avoidance, fine speed control and comfort maximizing, etc. are supposed to be accomplished within the subsequent trajectory planner.

1) Basic Actions and Provably Safe Basic Policy: We introduce three basic high-level actions and propose one rule-based driving policy that is provably safe regarding RSS safety.

- *Stop*: Stop before the first conflict zone (or the stop line).
- *Pass*: Try to pass the conflict zones with maximum ability.
- *Squeeze*: Carefully advancing with minimum velocity.

In order to fulfill RSS safety, the simplest policy can be formulated as follows: The ego vehicle can *pass* only when $\{C_2, C_{2,p}, C_3\}$ is fulfilled for all the conflict zones related to the current stop line, otherwise *stop*. This policy guarantees that at least one of C_1 and $\{C_2, C_{2,p}, C_3\}$ is satisfied. However, in some scenarios, no intersection between C_1 and $\{C_2, C_{2,p}, C_3\}$ can be found, where a traversal from one to another only with *stop* and *pass* is no possible, e.g. in scenarios with extremely strong occlusions, leading to a deadlock situation [20]. In this case, we introduce the third action *squeeze* which allows the ego vehicle to slowly approach, even enter the conflict zone to gain more visibility, e.g. with $1\frac{\text{m}}{\text{s}}$. The authors regard this action as RSS-safe as well. We name this simple policy the first basic policy (B1).

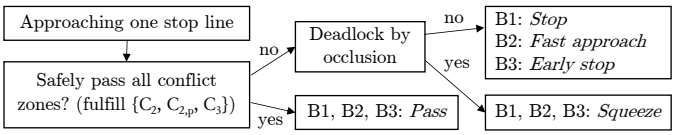


Fig. 3: Visualization of the rule-based policies B1, B2 and B3.

2) Extended Actions and Advanced Policy: After checking the behavior of human drivers in the datasets, we observe that in some scenarios, human drivers are able to find a smoother transition between C_1 and $\{C_2, C_{2,p}, C_3\}$, e.g. at intersections with slight occlusion. Instead of *stopping* before the conflict zones with constant deceleration, until $\{C_2, C_{2,p}, C_3\}$ is satisfied to switch to *pass*, they try to decelerate less at the beginning. In this way, before they must execute a harsh brake to not break C_1, $\{C_2, C_{2,p}, C_3\}$ is satisfied and they switch to *pass*. In order to mimic this behavior, we can introduce another action similar to *stop* but with less deceleration at the beginning, and more deceleration when getting closer to conflict zones. On the contrary, there are also scenarios where the ego vehicle better slows down more at the beginning, and less when getting closer, e.g. to show its cooperative stopping intention for other traffic participants.

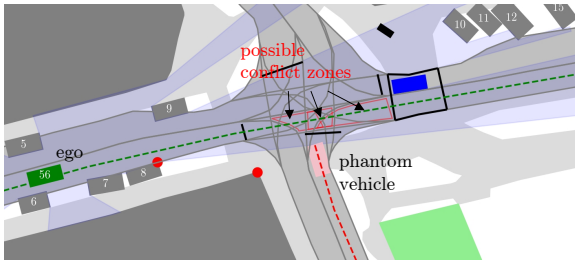
Theoretically, there are an infinite number of approaching styles, and *stop* is one of them as well. In order to not blow up our action space, we introduce only additionally one *fast approach* and one *early stop* actions. The low-level realization of these actions is presented in the next section. With these additional actions, more basic policies are created, i.e. B2 (substituting *stop* in B1 with *fast approach*) and B3 (substituting *stop* in B1 with *early stop*). In addition, more advanced and human-like policies are possible, i.e. instead of sticking always to one of the approaching actions, they can be freely selected at each decision step.

The policies B1, B2, and B3 are illustrated in Fig. 3.

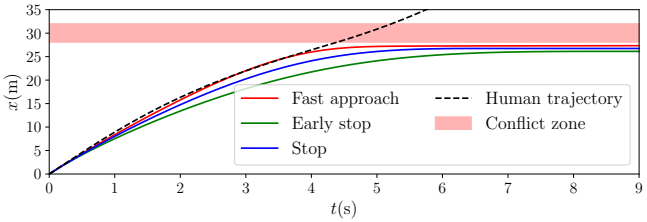
3) Proof-of-Concept Low-level Execution of Actions: The Intelligent Driver Model (IDM) [21] is utilized to generate longitudinal acceleration $\dot{v}_{\text{IDM}}$ for all approaching actions, where a virtual obstacle with 0 velocity is assumed to locate just before the conflict zone. In order to generate different approaching styles *fast approach*, *stop* and *early stop*, we add another parameter α into the original formulation and consider the virtual obstacle as an additional leading vehicle,

$$\dot{v}_{\text{IDM}} = a_{\max}\left(1 - \left(\frac{v}{v_{\text{d}}}\right)^4 - \max_{i=1}^{2} \alpha \left(\frac{d^*\left(v, \Delta v_{\text{f}_i}\right)}{d_{\text{f}_i}}\right)^2\right) \quad (1)$$

where $d^*\left(v, \Delta v\right) = d_0 + vT_{\text{d}} + \frac{v\Delta v}{2\sqrt{ab}}$ is the desired distance to the vehicle ahead. The parameters to set are: maximum acceleration $a_{\max}$, desired velocity v_{d}, minimum accepted distance d_0, desired time gap T_{d} and desired deceleration b.



(a) Ego vehicle approaching an occluded intersection



(b) Human approaching profile and generated action profiles

Fig. 4: The ego vehicle is approaching an occluded intersection with potential phantom vehicles. The y coordinate of (b) is the longitudinal distance along the green dashed route of the ego vehicle in (a). One recorded human driver tries to approach with the black trajectory. Our modified IDM generates three different profiles for the three approaching actions.

α aims to control the impact of the leading obstacle to the overall acceleration and is 1 in the original IDM formulation. In our paper, α is set to 0.5 for *fast approach*, 1 for *stop* and 2 for *early stop*.

Fig. 4 presents one typical occluded intersection. The example human driver is apparently not approaching with the typical IDM deceleration to stop before the conflict zones. However, a smoother transition between *stop* and *pass* is found, which satisfies the safety and is more comfortable. Among the three predefined actions, the *fast approach* trajectory matches the human driver's trajectory the best and should be selected in this case by an advanced human-like policy.

For *pass* action, the longitudinal acceleration should be the maximum ability of the ego vehicle before passing all conflict zones, because they need to be cleared as soon as possible. Therefore, the output acceleration is the maximum acceleration that fulfills $\mathrm{C}^*_{2,3}$ until reaching v_{limit}. For *squeeze* action, the output acceleration is just the output of a speed P-controller that tries to maintain $1\frac{\text{m}}{\text{s}}$.

C. Relevant Features for Approaching Actions

The features that affect decision making are categorized into four groups: *utility*, *ride comfort*, *perceived safety*, and *politeness*. In this paper, we only give a brief review of the previously proposed features and will reformulate the *perceived safety* with extended definitions. Note that all the values of the features will be normalized between 0 and 1.

1) Utility: The following three features are relevant:

- U_1: How fast the overall progress can be made.
- U_2: How soon the desired maneuver can be achieved.
- U_3: How possible the desired maneuver can be finished.

U_1 will mostly be affected by the velocity and can be formulated as $1 - |\frac{v}{v_{\text{des}}} - 1|$ where v denotes the average velocity achieved by an action or a trajectory and v_{des} represents the desired velocity of the driver. In scenarios where clear goals are defined (lane change, merging, passing intersection, etc.), U_2 and U_3 are additionally interesting. For example, at intersections, U_2 describes how soon and U_3 represents how probable the intersection can be passed. We will introduce how to estimate U_2 and U_3 by the MC simulations in Section III.

2) Ride comfort: In the context of longitudinal decision-making, our primary focus lies on ensuring longitudinal comfort. In the domain of trajectory planning, it is common to penalize jerk and acceleration within the cost function. However, as we employ acceleration outputs derived from the IDM for vehicular control inputs in MC simulation, it is not feasible to obtain jerk values. Moreover, devising high-level plans optimized with respect to jerk, rather than acceleration, offers limited advantages while incurring a substantial computational overhead during the MC simulation. Consequently, we define longitudinal comfort exclusively as a function of acceleration: $C = 1 - |\frac{a_{\text{l}}}{a_{\text{max}}}|$, where a_{l} represents the average absolute longitudinal acceleration of a maneuver or trajectory.

3) Perceived safety: This feature is also treated as risk. Similar to the Q-value in a Markov-Decision-Process (MDP), the perceived safety is not only depending on the current state, but also on the action that is about to be executed. Previous publications [22] usually focus on collision risk for perceived safety, where the probability of a collision is computed. However, as a collision is rare in real traffic, the collision probability is difficult to be validated. Instead, we introduce two risk definitions: *emergency risk* R_1 and the *fall-back risk* R_2.

The *emergency risk* represents the probability of emergency situations. RSS treats events to be emergencies when the RSS safety is broken, either passively (e.g. by intruding traffic participants that disrespect RSS) or actively (e.g. by violating C_1 and $\{\mathrm{C}_2, \mathrm{C}_{2,\text{p}}, \mathrm{C}_3\}$ at the same time), where a "proper response" (emergency reaction) needs to be performed, e.g. braking with $a_{\text{max,decel,ego}}$.

The *fall-back risk* represents the probability of switching to the most uncomfortable (fall-back) plan, where the RSS safety is on the verge of being undermined. The fall-back plan is still risky because the driver or the passengers might feel endangered, but does not harm RSS safety. In highway on-ramp merging, a fall-back plan can be a failed merge followed by a harsh stop at the end of the merging lane, due to the incorporation of the vehicle on the target lane. At an occluded intersection, when the vehicle approaches not cautiously (e.g. with *fast approach*), hoping that no prioritized vehicle is behind the occlusion such that switching to *pass* is soon possible, but one suddenly appears. In this case, the most uncomfortable part of the trajectory that nearly has a_{max} deceleration has to be executed. A *fallback* is defined as a deceleration over a threshold when approaching the intersection, e.g. $0.8a_{\text{max}}$.

We list some situations where R_1 and R_2 may be greater than 0. $R_1 > 0$ when

- Traffic participants behave beyond RSS assumptions (e.g. decelerate more than $a_{\text{max,decel,obj}}$).
- Traffic participants break traffic rules (e.g. violates speed limit, take way, cut in, or cross disrespect RSS).
- Highly uncertain perception results (e.g. ghost objects, extremely large estimation error).

$R_2 > 0$ when the ego vehicle has

- Wrong estimation of turning/routing/cooperation intention of other traffic participants.
- Too optimistic estimation of uncertainty in occlusions.

R_1 can hardly be eliminated and can only be minimized by always estimating the worst. However, by doing this, the utility will be largely damaged. We assume a reasonably correct perception such that the third point is ignored in this paper.

R_2 is possible to be 0, e.g. by always following *stop* or *early stop* actions, and can also be reduced, e.g. by having a better prediction module that generates a more accurate estimation of the environments. However, as the consequence of switching to fall-back is not as bad as an emergency situation, human drivers always risk the fall-back to try to be more efficient. In other words, they tolerate 1 harsh brake in 100 uncertain crossings, rather than 100 soft brakes among which 99 are unnecessary. The goal is to find a good balance between *utility* and *perceived safety* with the recorded human driving trajectories.

4) Politeness: Experienced drivers focus not only on their own benefit but behave in a way such that others' convenience is affected as less as possible. A good action allows a smoother overall traffic flow as well. We measure politeness $P_1 = \frac{1}{n}\sum_{i=1}^{n} U_{1,i}$ and $P_2 = \frac{1}{n}\sum_{i=1}^{n} C_i$ by looking at the average utility U_1 and average comfort C of the n surrounding traffic participants where $U_{1,i}$ and C_i are for the i-th object.

III. Monte-Carlo Simulation

As mentioned before, an advanced policy has more freedom than the basic policies B1, B2 and B3. It is able to decide between the approaching styles at each decision step based on the features (risk, utility, comfort and politeness) that characterize each approaching style.

A. Feature Estimation via Monte-Carlo Simulation

The features are approximated via MC simulations. The estimated feature vector $[U_1^*, U_2^*, U_3^*, C_1^*, R_1^*, R_2^*, P_1^*, P_2^*]$ are noted with $*$. MC simulations of N episodes start from the current scene and end after a certain simulation horizon t_{max}, where the ego vehicle follows one of B1, B2 and B3, and the environment reacts with our environment model. A different episode of MC simulations can result in a totally different future because randomness is introduced to the environment model, which will be explained in the next section. After many episodes of MC simulations for one approaching style, the feature values can be computed based on the simulation histories, e.g. how often the ego vehicle is expected to fall back, or how soon the ego vehicle is expected to pass the intersection on average.

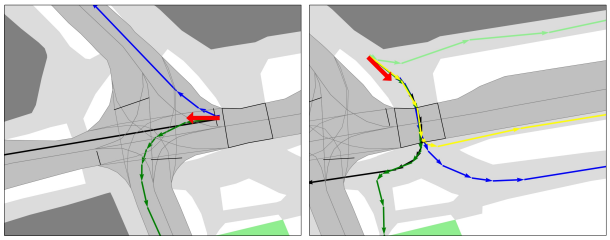


Fig. 5: Different route options from lanelet2 map for vehicles (left) and pedestrians (right) given the pose (red arrow).

The computation of U_1^*, U_2^*, U_3^*, C^*, P_1^* and P_2^* do not differ with our previous work [11], where the values U_1^*, C^*, P_1^* and P_2^* are just averaged over all the episodes.

From one MC simulation, not only numerical values (e.g. average velocity) but also other semantic information, e.g. whether a maneuver succeeds or a fall-back deceleration has been executed, can be obtained. If the maneuver succeeds in i-th episode, the simulation time $t_{\text{finish},i}$ is recorded. Otherwise, t_{max} is utilized for $t_{\text{finish},i}$. As a result, $U_2^* = \frac{1}{N}\sum_{i=1}^{N}(\frac{t_{\text{finish},i}}{t_{\text{max}}})$ and $U_3^* = \frac{n_{\text{finish}}}{N}$ can be formulated mathematically, where n_{finish} represents the number of episodes where the desired maneuvers are completed. Similarly, R_1^* and R_2^* represent the ratios of episodes where the ego vehicle has executed the fall-back deceleration and where the RSS safety is violated either passively or actively.

B. Environment Modeling for MC Simulation

In order to have well-estimated features by MC simulations, the environment should behave as close to reality as possible. We explain the uncertainties that are considered and how the behaviors and intentions of other traffic participants are estimated in this section.

1) State Uncertainty: We assume that uncertainty exists in the state (position, velocity, acceleration, size, etc.) of the ego vehicle and the states of surrounding traffic participants, and the distributions are provided by the localization and perception module. When initiating MC simulations, the states of all the traffic participants are randomly sampled from the distributions.

2) Behavior Modeling and Intention Estimation of Vehicles: The most common traffic participants are surrounding human-driving vehicles. In MC simulations, they follow the IDM model during car-following and the basic yielding policy for crossing intersections. They are assumed to have perfect knowledge of the traffic rules as well.

In reality, the street is populated with different types of drivers. In order to simulate different car-following styles and yielding styles, we predefine three driving styles (aggressive, normal, defensive) by different levels of aggressiveness, each of which is associated with a different IDM parameter and RSS parameter. Their basic yielding policy will be one of B1, B2 and B3 depending on the aggressive level as well. At initialization of the MC simulations, each vehicle will be randomly assigned with one aggressive level and thus with the corresponding parameters and basic yielding policy. A more realistic approach will be deciding the aggressive level of each

vehicle by observing its historical driving behavior but is not included in the scope of this work.

Each vehicle has different goals as well, which is not known to the ego vehicle without additional information (indication, etc.). The goal is represented by a global route, which is composed of a sequence of lanelets on the map. An example is shown in Fig. 5. As input for the MC simulation, the probabilities $\{P(r_i), \dots, P(r_I)\}_{v_n}$ of each vehicle v_n following all its possible routes $\{r_i, \dots, r_I\}_{v_n}$ for $I \in \mathbb{N}$ are required. Without external prediction modules, the probabilities will be assigned to be equal. We adopt one basic routing prediction method [23] that generates the routing probability by matching the state distribution of the vehicle to the centerline of each route. First, the squared Mahalanobis distance $d(v_n, r_i)$ between the vehicle v_n and the route r_i is computed, then the probability is referred by assuming a Boltzmann distribution

$$P(r_i)_{v_n} = \frac{e^{-d(v_n, r_i)}}{\sum_{j=1}^{I} e^{-d(v_n, r_j)}}, \text{for } i \in \{1, \dots, I\} \tag{2}$$

In general, any prediction module (e.g. [24]) that is able to provide the same information can be adopted, which increases the modularity of our method and makes it prediction agnostic.

3) Behavior Modeling and Intention Estimation of Pedestrians: In MC simulation, pedestrians may have several potential routes as well and the estimation is done with the same method as for vehicles. When no zebra is in front, they are assumed to move with constant velocity following one of the routes. Otherwise, when they are closer to zebra than a threshold $d < d_{\text{min}}$, but not jet on it, they will start making crossing decisions. After locating on the zebra, they are simulated to cross straight-forward with a predefined maximum velocity.

We assume an interactive behavior model of pedestrians when they try to cross zebra. The basic idea is that pedestrians tend to start crossing with a higher probability when the traffic is clear, but will hesitate when the street is busy or vehicles are driving fast and do not show decelerating intention explicitly. We learn a logistic regression model for predicting the crossing probability of pedestrians at zebra from inD dataset

$$P_{\text{cross}} = \frac{1}{1 + e^{-(\theta_{\text{p}}^T f_{\text{p}} + b_{\text{p}})}} \tag{3}$$

with θ_{p} to be the weight vector, b_{p} the bias and $f_{\text{p}} = [a_{\text{v}}, a_{\text{v,need}}, v_r]$ the feature vector, where a_{v} denotes the current acceleration of the closest on-coming vehicle to zebra, $a_{\text{v,need}}$ is its needed acceleration to stop before zebra, and $v_r = 1 - \frac{v}{v_{\text{limit}}}$ is its normalized speed to speed limit. The results are $\theta_{\text{p}} = [0.75, -0.5, -1.5]$ and $b_{\text{p}} = -2$. Crossing decision is made when $P_{\text{cross}} > 0.5$. Before stepping onto the zebra, they will renew their decision every 1s of simulation time by computing P_{cross} again.

4) Behavior Modeling and Intention Estimation of Cyclists: Cyclists are recorded in the datasets as well. We do not introduce specific behavior models or intention estimation methods for cyclists but assign either pedestrian-like behavior or vehicle-like behavior in MC simulations. The perceived cyclists that locate on the walkway will be modeled with pedestrian behaviors. Those who drive on the vehicle lanes will be treated similarly to vehicles, with cyclist-specific IDM and RSS parameters though.

5) Modeling of Abnormal Behaviors: Sec. II-C3 lists several situations where the traffic participants behave with abnormal behaviors, leading to emergency reactions of the ego vehicles ($R_1 > 0$). In order to have a realistic estimation of R_1, these behaviors should also be modeled in MC simulation. We analyzed several traffic rules non-compliant behaviors and counted their incidence in the datasets. The results are presented in Table I.

Therefore, besides assigning vehicles with the three aggressive levels, we assign abnormal behaviors to the vehicles that will be initialized in MC simulation according to the real occurrence rate in the datasets, e.g. by initiating with an abnormal RSS parameter or a high desired speed, etc. Note that there is no unexpected pedestrian crossing recorded that is not on the zebra crossing and causes the other vehicle to brake more than $a_{\text{soft,decel,obj}}$. However, we still assign 0.1% pedestrians to not follow the optional routes but may cross the street unexpectedly in MC simulations. Optimally, this probability can be computed by tracking the movement of the pedestrian via external modules.

6) Simulation of FoV and Sampling Phantom Vehicles from Occlusion: In the MC simulation, vehicles are supposed to cross intersections according to the basic policy. Simulating the FoV for every agent in the scene is computationally intractable. Therefore, we assume that other vehicles have a perfect perception of the environment and make reasonable decisions. However, for the ego vehicle, the FoV polygon is simulated forward as it moves. The obstacles that are used for computing the simulated future FoV are currently perceived static obstacles and simulated dynamic obstacles (without pedestrians). With the limited FoV, the ego vehicle can only *pass* when the safety condition in Sec. II-A5 is fulfilled.

During MC simulation, if the ego vehicle is following a *fast approach* action, it may always be able to switch to *pass* smoothly as only perceived vehicles are added in MC simulation and no vehicle will come out from occlusion. However, in reality, if the ego vehicle does *fast approach* as well, thinking that a smooth transition is certain according to its MC simulations, it may fall back when suddenly actual vehicles appear from occlusion. In order to catch this potential fall-back probability, phantom vehicles should be sampled from occluded road sections as well. The idea is to sample based on the perceived traffic density. In order to reduce the burden of MC simulations, we only sample phantom vehicles on the occluded sections of prioritized road sections.

Fig. 6 illustrates an example of possible occluded prioritized routes under the FoV, where phantom vehicles need to be sampled in MC simulations. For doing this, the traffic density $g = \frac{N}{L}$ is first computed, where N represents the number of the perceived dynamic vehicles in the scene (including the ego vehicle), and L is the total length of roads in all directions covered by the FoV. The expected number of phantom vehicles on each occluded section will be $n_{i,\text{exp}} = g l_i \in \mathbb{R}$ where l_i represents the length of the i-th section. The actual number of sampled vehicles is $n_{i,\text{sample}} = \lceil \mathcal{N}(n_{i,\text{exp}}, 1) \rceil$ and $\lceil \rceil$ is a ceiling function. After $n_{i,\text{sample}}$ is decided, the longitudinal

TABLE I: Traffic rules non-compliant behaviors and their occurrence rate.

Type	Datasets	Number of cases analyzed	Occurrence rate (%)
Exceeding 20% speed limit	InD, rounD	19671 vehicles	4.8
Exceeding RSS acceleration/deceleration	InD, rounD	19671 vehicles	0.01
Taking way disrespect RSS safety	InD, rounD	11081 intersection crossings	17.1
Unexpected pedestrian crossing	InD	3093 pedestrians	0

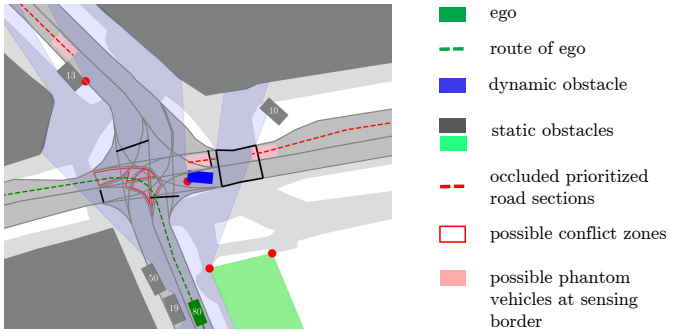


Fig. 6: Possible phantom vehicles on occluded prioritized road sections.

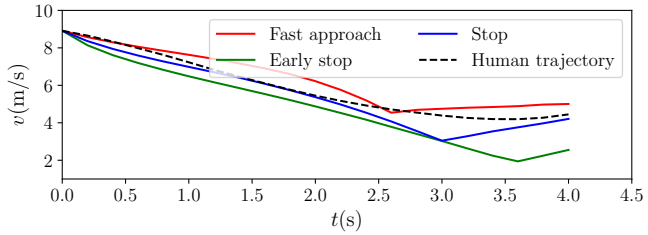


Fig. 7: Comparision of the average velocity profiles from MC simulations and the human-driven trajectory.

position $s_{i,j}$ of the j-th vehicle on the i-th section follows a uniform distribution $s_{i,j} \sim \mathcal{U}(S_{i,\text{free}})$, with

$$S_{i,\text{free}} = \{s \in [0, l_i] \wedge s \notin \bigcup_{k=1}^{n_{i,\text{sample}}} S_{k,i,\text{occupied}} \text{ for } k \neq j\} \quad (4)$$

where $S_{k,i,\text{occupied}} = \{s \in [s_{k,i} - d_{\text{safe}}, s_{k,i} + d_{\text{safe}}]\}$ is the occupied distance range of k-th already sampled phantom vehicle on i-th section. $s_{k,i}$ is the longitudinal position of k-th phantom vehicle on i-th section. $d_{\text{safe}} = 0.5v_{k,i}$ is a minimum longitudinal distance between vehicles. This guarantees that the sampled vehicles do not overlap with each other and even have at least 0.5s time headway between each other.

The sampled phantom vehicles are initialized with a speed following a uniform distribution $\mathcal{U}(0.8v_{\text{limit}}, 1.2v_{\text{limit}})$. Their behavior models and other parameters are initialized the same way as the visible vehicles.

7) Performance Evaluation: Unlike POMDP which usually builds search trees and can not be well-parallelized, each of the single MC simulations is independent of others and thus can be parallelized in a multi-core system. The feature values converge as the number of MC simulations increases. In a selected set of test scenarios, the variance of one feature value (e.g. R_2) related to the MC simulations with 100, 500, and 1000 repetitions is decreased from 0.06, 0.025, to 0.015. The run-time for evaluating the three actions (each with 500 MC simulations) on a laptop with a CORE-i7 8th-Gen Intel CPU with 8 threads is 20ms, 80ms, and 140ms respectively. We take 500 repetitions as a good balance of run-time and accuracy.

IV. Learning Driving Policy via Behavior Cloning

As depicted in Fig. 1, after the features for each action are approximated by the results of MC simulations, one linear function with parameters w and b is utilized to generate Q-value for the action. As the goal is to learn a policy that balances the feature values similar to human drivers, e.g. not overly egoistic (weight utility and comfort too much) or overly cautious (weight risk too much), we first need to determine the decision preferences of humans in certain scenarios. For this purpose, we utilize the inD and rounD [25] datasets that contain a huge amount of trajectories of human drivers at several intersections and roundabouts.

The reason for not using a neural network to represent the Q-function is that we want to explicitly watch the weighting for each feature to better understand the decision and prevent overfitting to the limited data.

A. Generation of Training Data

By performing MC simulations, not only the estimated feature $f^*_{a_i} = [U^*_{1,a_i}, U^*_{2,a_i}, U^*_{3,a_i}, C^*_{a_i}, R^*_{1,a_i}, R^*_{2,a_i}, P^*_{1,a_i}, P^*_{2,a_i}]$ for each action $a_i \in \mathcal{A}, \mathcal{A} = \{a_1, a_2, a_3\} = \{\textit{fast approach}, \textit{stop}, \textit{early stop}\}$ can be computed, but also the average velocity profiles $\overline{V}_{a_i,t_{\max}} = [\overline{v}_{a_i,t_0}, \overline{v}_{a_i,t_1}, \ldots, \overline{v}_{a_i,t_{\max}}]$, where $\overline{v}_{a_i,t_j} = \frac{1}{M}\sum_{k=1}^{M} v_{a_i,k,t_j}$ and v_{a_i,k,t_j} is the velocity of the ego vehicle at t_j time step of k-th MC simulation by following the action a_i. M is the total number of MC simulations for one action.

For each valid frame[1] of each recorded vehicle in the dataset, M episodes of MC simulations are performed for each of the three actions, resulting in three different velocity profiles. Afterward, an error $\sigma_{a_i,t_{\max}}$ between the ground-truth velocity profile $V_{\text{gt},t_{\max}} = [v_{\text{gt},t_0}, \ldots, v_{\text{gt},t_{\max}}]$ of the vehicle and the three generated velocity profiles $\overline{V}_{a_i,t_{\max}}$ is computed with

$$\sigma_{a_i,t_{\max}} = \sum_{t_j} (v_{\text{gt},t_j} - \overline{v}_{a_i,t_j})^2 \quad (5)$$

The probability P_{a_i} of how humans would select each of the actions a_i is estimated assuming a Boltzmann distribution

$$P_{a_i} = \frac{e^{-\sigma_{a_i,t_{\max}}}}{\sum_{a_j} e^{-\sigma_{a_j,t_{\max}}}}, a_i, a_j \in \mathcal{A} \quad (6)$$

[1]Frames are valid when the vehicle in the ground-truth trajectory has not passed all intersections

TABLE II: Learned weights w.

	Utility			Comfort	Risk		Politeness	
	U_1	U_2	U_3	C	R_1	R_2	P_1	P_2
w	**6.3**	**-6.8**	**5.8**	0.5	**-3.2**	-1.5	0.8	1.2

As an example, we execute the MC simulations with $t_{\max} = 4$s under the scene of Fig. 6, where the average velocity profiles for the three actions and the ground-truth human trajectory are recorded and depicted in Fig. 7. The turning points of the three velocity profiles are the switching points between slowing down and *pass*, where the simulated FoV can cover enough of the prioritized lanes and safety conditions are satisfied. Using eq. 6, the matched probability of the three actions are $P_{\text{fast approach}} = 0.635$, $P_{\text{stop}} = 0.364$ and $P_{\text{early stop}} = 0.001$.

One training data d_i of the training dataset $\mathcal{D}$ is composed of the data $[f^*_{a_1}, f^*_{a_2}, f^*_{a_3}]$ and the label $[P_{a_1}, P_{a_2}, P_{a_3}]$. There are four intersections and three roundabouts containing in total 56 recordings of ca. 30 minutes in inD and rounD datasets. For each intersection or roundabout, we use half of the recordings to generate training data. In total, we evaluated 20240 valid frames and generated training data of the same size.

B. Learning Driving Policy from Datasets

We use softmax cross-entropy loss L for back-propagation and updating the parameters w and b, which is formulated for the entire training dataset as

$$\begin{aligned} L &= -\sum_{d_i \in \mathcal{D}} \sum_{a_i \in \mathcal{A}} P_{a_i} \log\left(\frac{e^{-Q_{a_i}}}{\sum_{a_j \in \mathcal{A}} e^{-Q_{a_j}}}\right) \\ &= -\sum_{d_i \in \mathcal{D}} \sum_{a_i \in \mathcal{A}} P_{a_i} \log\left(\frac{e^{-(w f^*_{a_i} + b)}}{\sum_{a_j \in \mathcal{A}} e^{-(w f^*_{a_j} + b)}}\right) \end{aligned} \tag{7}$$

As a result, the learned weights is presented in Table II and $b = 0.8$. As the features are normalized between 0 and 1, the weights are representative in reflecting the preferences of human drivers. Obviously, human drivers pay more attention to their overall utility and risk.

V. Evaluation

For showing the strength of our approach, we first evaluate it on an interactive simulation that is built upon the datasets. We further show the generalization of our approach with the evaluation on the new roundabout from the Interaction dataset [26]. We categorize the evaluation scenarios into three types: unsignalized intersections with slight occlusions caused by static obstacles, with severe occlusions, and roundabouts. Quantitative evaluations and interesting case studies comparing four policies are performed in the end.

A. Evaluation Simulation

There are some existing simulators or benchmark toolings. Carla [27] provides realistic environment representation and sensor simulation. However, it is not straightforward to integrate the lanelet2 maps, and designing an interface for low-level controlling requires much effort as well. CommonRoad [28] is another benchmark for evaluating planning algorithms. It provides simulations that are partly recorded from real traffic and partly hand-crafted to create dangerous situations. However, it focuses on evaluating the cost functions of motion planning algorithms but is less interesting for our high-level behavior generation approach. BARK [29] targets to provide a realistic and interactive simulation environment that is initiated from datasets, but with reactive surrounding agents following several pre-designed behavior models. It meets our requirements best but does not support lanelet2 maps as well, and the benchmarking function is not yet fully released.

Evaluating behavior models in totally offline datasets has the advantage that other agents are moving according to the recorded trajectories and are realistic. However, the drawback is that they do not react to the movement of the automated ego vehicle which diverges from the ground truth since the second simulation step. Following the idea of BARK, we build a similar simulation upon the datasets, but on the data that were not used for training, i.e. the other half of the inD and rounD datasets, and the Interaction dataset.

After running the simulation, one of the vehicles in the scene is regarded as the ego vehicle which will follow our driving policy and replace its original trajectory. Other agents will behave following their recorded trajectories. However, they will be overridden by automated driving (AD) agents that diverge from the original trajectories once one of the following conditions is met:

- The distance to its front agent is less than the RSS safe distance computed from a relaxed RSS parameter.
- Starting to cross the intersection if the crossing is not RSS-safe according to a relaxed RSS parameter.

The overridden AD agents will be randomly assigned one aggressive level and their behavior models and parameters are initialized the same as in MC simulation.

As the FoV of the recording drone is limited, new agents might spawn from the edge of the scene that may collide with the overridden agents. We do not spawn these new agents once one of the following conditions is met:

- The spawned position is already occupied by other agents.
- The distance of the spawning position to its front agent or following agent is less than the RSS safe distance computed from a relaxed RSS parameter.

A relaxed RSS parameter allows more aggressive driving as it assumes e.g. a larger $a_{\text{max,decel,ego}}$, and a smaller $t_{\text{TZC,min}}$, etc. With these modifications, the simulation is as close to reality as possible, and on the other hand, is populated with reactive agents that try to avoid collisions and follow traffic rules.

For validating our simulation, we compared it with the one that only replays the offline datasets for other vehicles. Each vehicle in the datasets will be treated as the ego vehicle once and follows our learned policy, while others behave reactively or only follow their recorded trajectories. After running through the whole test dataset, the number of resulting collisions between all agents is recorded. On average, the number of collisions for simulating one AD agent is reduced from 2.3 to 0.05, by introducing our modifications. The remaining few collisions are mostly from edge cases, e.g. the

TABLE III: Statistics for simulation evaluation for intersections with neglectable static occlusion.

Policies	MDE (m)	Average velocity ($\frac{m}{s}$)	Fall-back ratio (%)	Velocity gain of the traffic ($\frac{m}{s}$)
B1	**8.46**	6.19	**5.02**	0
B2	10.51	**6.91**	17.14	-0.02
Learning (L1)	**9.94**	6.74	9.72	**0.003**
Learning + better prediction (L2)	10.25	**6.81**	**9.06**	**0.002**

TABLE IV: Statistics for simulation evaluation for intersections with severe static occlusion.

Policies	MDE (m)	Average velocity ($\frac{m}{s}$)	Fall-back ratio (%)	Velocity gain of the traffic ($\frac{m}{s}$)
B1	9.98	5.33	**2.5**	0
B2	**9.38**	**5.86**	30.3	**0.09**
Learning (L1)	**9.51**	**5.66**	**7.3**	**0.06**
Learning + better prediction (L2)	**9.55**	**5.63**	**6.6**	**0.07**

front vehicle is not clearly identified as they drive close to the border, or the bounding boxes of the vehicles are not accurate enough and have slight overlap, etc. Our simulation is still regarded as realistic enough, as on average only 9.5% agents of the scene have been overridden, and most of the other agents still follow their trajectories.

B. Compared Policies and Metrics

We compare four policies and three of them are already explained, i.e., B1, B2, and our learned policy (L1). In order to show how the prediction module affects the quality of our learned policy, we utilize a better proof-of-concept routing prediction module instead of eq. 2 and apply the learned policy on that, which makes up the fourth policy (L2). As all the ground-truth routing of the surrounding agents are known (no matter whether they are overridden), their future 3s of points on the ground-truth routing with 0.5s interval is used for matching with their possible routes, instead of only the current pose. We omit the mathematical details here because they are similar to eq. 2. This prediction module has significantly better accuracy as it "cheats" to use the ground-truth data. However, this is acceptable as our goal is not to provide a good prediction module, but only to show how a good one can help improve the planning quality.

We evaluate the following four metrics:

- *Mean distance error (MDE)* to the ground-truth trajectory: to show the human-likeness of each policy.
- *Average velocity*: proportional to the inverse of average crossing time, but includes the parts after the intersection and is more representative of the overall utility.
- *Fall-back ratio*: how often does the ego AD agent need to switch to *fallback*.
- *Velocity gain of the traffic*: how the average velocity of traffic flow is improved compared to policy B1.

C. Massive Evaluation on Test Scenarios

1) Intersections with Neglectable Static Occlusion: There are three intersections in inD dataset where the occlusions caused by the static obstacles are not severe to hinder driving. In total, 458 vehicles are evaluated that have encountered at least one yielding intersection.

The quantitative results are presented in Table III. It can be seen that B1 has the most human-like performance, but achieves a relatively low average velocity. As it applies *stop* as the approaching action, it has the least fall-back ratio. The reason why the fall-back ratio is not 0 for B1 is that some vehicles start to be recorded really close to an intersection and intend to cross at high speed. However, the RSS safety is not fulfilled and the B1 policy decides to stop, which results in a high deceleration and triggers the fallback. B2 achieves the overall highest velocity but results in the largest fall-back ratio and is the least human-like policy. In addition, the traffic flow is compromised as well.

The two learned policies have similar performances. They are more human-like and achieve way less fall-back ratio than B2, but with only slight velocity loss. The overall traffic flow is improved as well, as politeness is considered in the features. By applying a better prediction, the fall-back ratio is decreased and the average velocity is improved as well.

2) Intersections with Severe Static Occlusion: There is one intersection in inD dataset where the buildings and the parked vehicles occlude all the arms of the intersection severely, as shown in Fig. 6. There is additionally one pedestrian crossing on the west arm of the intersection marked with a black polygon, where the ego vehicle should yield to pedestrians. In total, 824 vehicles are evaluated that have encountered at least one yielding intersection or one pedestrian crossing.

The quantitative results are presented in Table IV. With severe occlusions, B1 is the least human-like policy and achieves the lowest velocity. B2 is the fastest policy which is the most human-like one as well. However, it leads to a 30.3% fall-back ratio which is extremely unpleasant. The learned policies reduce the fall-back ratio a lot, with slightly increased MDE and decreased velocity. All the policies that achieve a faster average velocity allow a smoother traffic flow as well. The better prediction module again helps a little in reducing the fall-back ratio as the turning intentions of prioritized vehicles are better estimated.

3) Roundabouts: There are three different roundabouts in rounD dataset where the streets are mostly clear in all directions. In total, 4282 vehicles are evaluated.

The quantitative results are presented in Table V. Simulation results are similar to the intersections with severe static occlusions. The learned policies achieve the highest average velocity and balance the risk to an acceptable level. With a better prediction module, the performance is again increased in almost all metrics.

4) Unseen Roundabout in Interaction Dataset: The learned policies do show better performance in different scenarios of

TABLE V: Statistics for simulation evaluation for roundabouts.

Policies	MDE (m)	Average velocity ($\frac{m}{s}$)	Fall-back ratio (%)	Velocity gain of the traffic ($\frac{m}{s}$)
B1	6.40	5.41	**0.8**	0
B2	**6.31**	**5.65**	24.2	**0.025**
Learning (L1)	**6.36**	5.53	**4.0**	0.004
Learning + better prediction (L2)	**6.31**	**5.59**	**4.8**	**0.013**

TABLE VI: Statistics for simulation evaluation for an unseen roundabout in Interaction dataset.

Policies	MDE (m)	Average velocity ($\frac{m}{s}$)	Fall-back ratio (%)	Velocity gain of the traffic ($\frac{m}{s}$)
B1	6.58	6.83	**0.4**	0
B2	**6.13**	**6.94**	13.4	**0.03**
Learning (L1)	**6.02**	**6.92**	**0.9**	0.01
Learning + better prediction (L2)	**6.09**	**6.93**	**1.3**	0.02

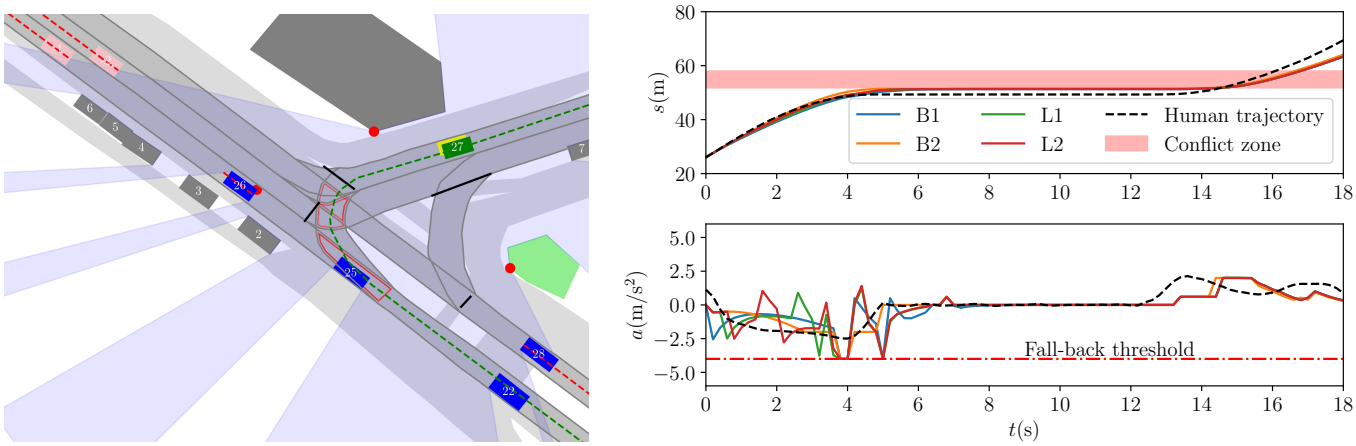


Fig. 8: The ego vehicle (green) tries to finish an unprotected left turn. The yellow polygon in the background visualizes the ground-truth position of the ego vehicle.

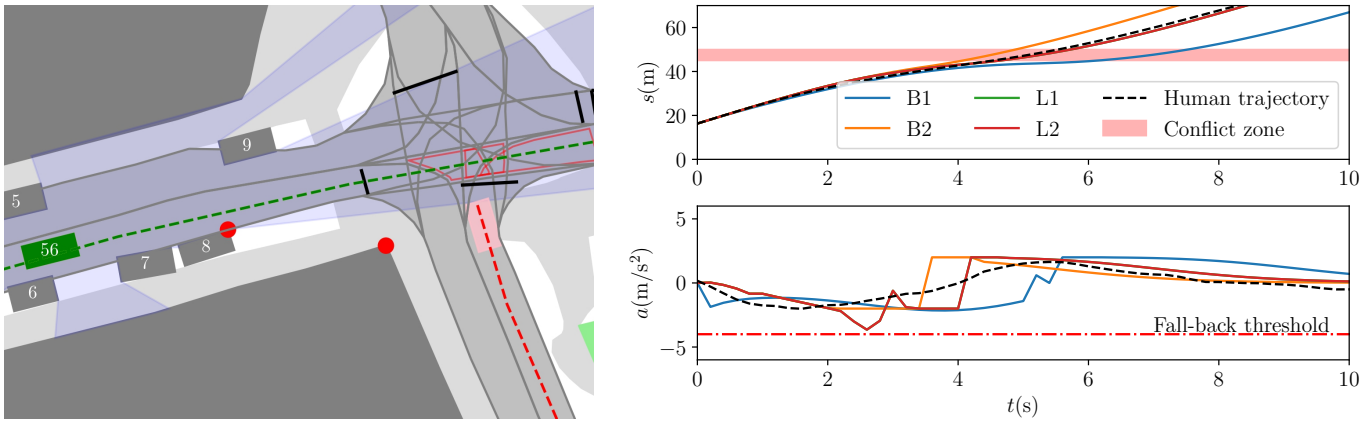
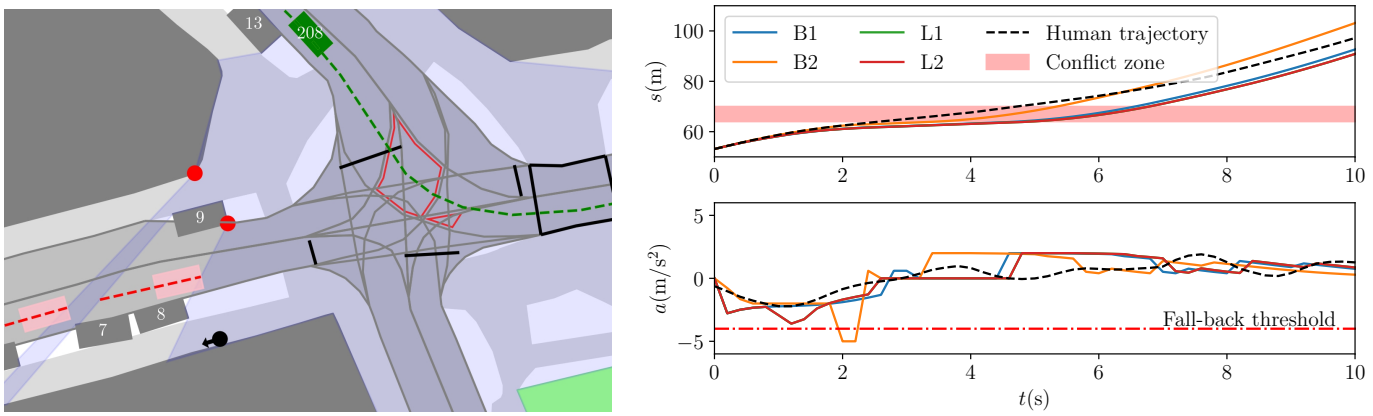


Fig. 9: The ego vehicle (green) is approaching an intersection with ordinary occlusion. The profiles of L1 and L2 overlap.

the datasets, but still on the same intersections as the training data. In order to present the generalization of our approach, we select the roundabout *DR_DEU_Roundabout_OF* from Interaction Dataset and apply the same quantitative evaluation. Other intersections are either highly unstructured or are not unsignalized intersections with the yielding traffic rule. In total, 552 vehicles are evaluated. The results are presented in Table VI. The statistics show a similar pattern as for the three roundabouts in inD dataset, where the learned policies are able to maximize the utility of the ego vehicle while keeping the fall-back ratio at a reasonably low level.

Evaluation results on the unseen roundabout prove that our approach is map agnostic. As the MC simulations perform directly on the map and take the traffic participants as input without abstracting any information, the estimated features have similar accuracy for any unseen intersection. Exceptions are e.g. when performing MC simulations on new scenarios where the traffic participants have significantly different behavior (e.g. in different countries), and the behavior modeling in Sec. III-B may produce a big error.



n. The profiles of L1 and L2 overlap.

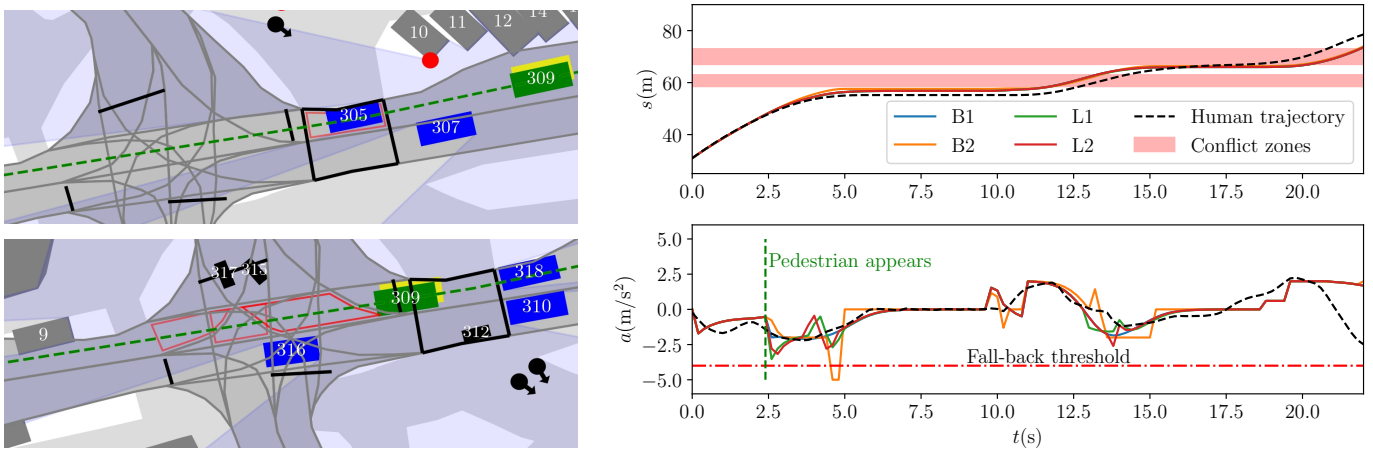


Fig. 11: The ego vehicle (green) is approaching consecutive stop lines. The first one is a pedestrian crossing, and the second one is a yielding.

D. Case Study

We take several representative scenarios and analyze the driving history of each driving policy, i.e. s-profiles (longitudinal distance along the route) and a-profiles (acceleration).

Fig. 8 presents one complex unprotected left turn with several potential conflicting zones, where potential occluded vehicles and some visible vehicles have priority. The approaching accelerations of the policies are different, but all the policies are able to pass the intersection after the intersection is clear with similar s-profiles.

Fig. 9 presents one scenario where the intersection is not severely occluded but cautious driving is still needed. Approaching relatively fastly (with B2, L1 and L2) allows a smooth transition between slowing down and *pass*. With B1, the ego vehicle is reduced to a low velocity. B2 achieves a faster speed even than the human trajectory, which is efficient but might let the passengers feel endangered. L1 and L2 produce the most human-like trajectories.

Fig. 10 illustrates a scenario on the same intersection, but the ego vehicle is coming from the north arm. The prioritized lane (west arm) is additionally occluded by the parked cars. B2 again achieves the highest utility, but has to execute a fallback as stopping in front of the conflict zones is no longer guaranteed, and *pass* is not safe as well. B1 slows down more but allows passing the conflict zones without a fallback. L1 and L2 produce the same behavior, they decelerate more than B1 at the beginning but then could accelerate earlier than B1. Thus, B1, L1 and L2 achieve almost the same utility, which is slightly lower than the human, but all without a fallback. The human driver does not slow down too much at the beginning but intrudes into the conflict zones more aggressively than our *squeeze* action.

Fig. 11 shows the behaviors on consecutive stop lines where the first one is a pedestrian crossing and the second one is yielding to prioritized cyclists. After the ego vehicle perceives the pedestrian, L1 and L2 execute *early stop* action and decelerate more than B1 and B2 to show cooperative intention to the pedestrian. The reason for executing *early stop* is that in MC simulations, decelerating early motivates the pedestrian to make *cross* decision according to our pedestrian behavior models in Sec. III-B3. Thus, the pedestrian can pass the zebra faster and clear the conflict zone earlier, which allows the ego vehicle to pass faster as well. In this way, both the utilities of the ego vehicle and the pedestrian are expected to be increased. Note that B2 leads to a fallback again as the conflict zone is not cleared soon enough. Afterward, all four policies traverse the second stop line and its conflict zones similar to the human trajectory.

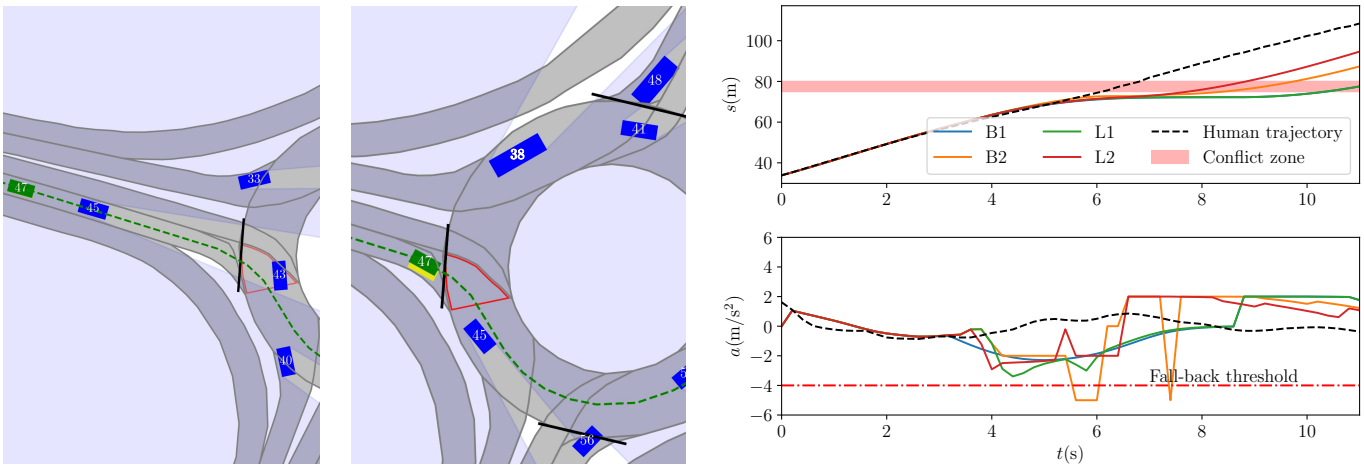


Fig. 12: The ego vehicle (green) is approaching a roundabout. The vehicle with id 38 is recorded to exit the roundabout via the west exit.

Fig. 12 presents a roundabout scenario. The human driver approaches the roundabout with only slight deceleration and enters with almost no hesitation. The reason is that the intention of the vehicle with id 38 is well-estimated by humans, e.g. through indicator signals. The basic policy B1 and the learned policy L1 have similar behavior, where they are not certain about the exiting intention of id 38. Therefore, they need to come to a full stop to ensure safety. B2 tries to approach fastly but safety is not fulfilled until the exiting intention of id 38 is certain, leading to a fallback. However, with a better prediction module and a better intention estimation, L2 adjusts the velocity such that entering the roundabouts becomes safe earlier, but without a fallback.

In summary, B1 shows the most conservative driving behavior but is the least risky one. With B2 which is the opposite of B1, the utility of the ego vehicle is maximized, but with the cost of the most frequent fallback. The learned policy L1 finds a good balance between utility, risk and the overall traffic flow. It achieves similar human-likeness, utility and traffic flow as B2, but leads to significantly less fallback. With a better prediction module (L2), the fall-back ratio is further reduced without affecting other metrics.

VI. Conclusions And Outlook

In this study, we generalize our previously proposed behavior cloning concept for learning high-level decisions in urban driving scenarios, particularly at unsignalized intersections. We extend the RSS safety concept to address various conflict zones, pedestrian safety, and occlusions. Our action space representation encompasses both aggressive and conservative behaviors, enabling the generation of provably safe driving policies. To attain more human-like behavior, actions are selected based on their feature values. We adapt feature definitions to urban scenarios and broaden the concept of risk to encompass uncertainties that may necessitate emergency responses. These features are estimated using MC simulations, projecting the current uncertain environment into the future.

The approach boasts several advantages, including a highly modular design. It accepts uncertain perception results in any format as input and can optionally incorporate advanced prediction modules to enhance performance. The output comprises high-level decisions, which can be converted into low-level control commands by any trajectory planning module. Additionally, tracing the resulting decisions is straightforward, either by examining the MC simulations or the Q-value from the linear function.

An intriguing avenue for future research involves extending the current approach to unstructured environments, wherein HD maps are unavailable. The biggest challenge lies in defining appropriate traffic regulations, which are essential for the RSS safety concept. Additionally, exploring the influence of learned weights on driving styles warrants further investigation. Such research could enable the provision of predefined driving styles or even facilitate real-time tuning of driving preferences to accommodate diverse user requirements.

References

[1] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. A. Sallab, S. Yogamani, and P. Pérez, "Deep reinforcement learning for autonomous driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–18, 2021.

[2] D. Kamran, C. F. Lopez, M. Lauer, and C. Stiller, "Risk-aware high-level decisions for automated driving at occluded intersections with reinforcement learning," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 1205–1212.

[3] B. Mirchevska, C. Pek, M. Werling, M. Althoff, and J. Boedecker, "High-level decision making for safe and reasonable autonomous lane changing using reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2156–2162.

[4] C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning," *Robotics and Autonomous Systems*, vol. 114, pp. 1–18, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889018302021

[5] M. Wulfmeier, D. Rao, D. Z. Wang, P. Ondruska, and I. Posner, "Large-scale cost function learning for path planning using deep inverse reinforcement learning," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1073–1087, 2017. [Online]. Available: https://doi.org/10.1177/0278364917722396

[6] Z. Huang, J. Wu, and C. Lv, "Driving behavior modeling using naturalistic human driving data with inverse reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10 239–10 251, 2022.

[7] S. Mo, X. Pei, and C. Wu, "Safe reinforcement learning for autonomous vehicle using monte carlo tree search," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6766–6773, 2022.

[8] C. Lazarus, J. G. Lopez, and M. J. Kochenderfer, "Runtime safety assurance using reinforcement learning," in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 2020, pp. 1–9.

[9] D. Kamran, Y. Ren, and M. Lauer, "High-level decisions from a safe maneuver catalog with reinforcement learning for safe and cooperative automated merging," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 804–811.

[10] C. Hubmann, N. Quetschlich, J. Schulz, J. Bernhard, D. Althoff, and C. Stiller, "A pomdp maneuver planner for occlusions in urban scenarios," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 2172–2179.

[11] L. Wang, C. Fernandez, and C. Stiller, "High-level decision making for automated highway driving via behavior cloning," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 923–935, 2023.

[12] F. Codevilla, E. Santana, A. M. Lopez, and A. Gaidon, "Exploring the limitations of behavior cloning for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[13] W. Farag and Z. Saleh, "Behavior cloning for autonomous driving using convolutional neural networks," in *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2018, pp. 1–7.

[14] F. Poggenhans, J.-H. Pauls, J. Janosovits, S. Orf, M. Naumann, F. Kuhnt, and M. Mayr, "Lanelet2: A high-definition map framework for the future of automated driving," 11 2018, pp. 1672–1679.

[15] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," *CoRR*, vol. abs/1708.06374, 2017. [Online]. Available: http://arxiv.org/abs/1708.06374

[16] M. Naumann, "Probabilistic motion planning for automated vehicles," Ph.D. dissertation, Karlsruher Institut für Technologie (KIT), 2020.

[17] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 1929–1934.

[18] P. F. Orzechowski, A. Meyer, and M. Lauer, "Tackling occlusions limited sensor range with set-based safety verification," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 1729–1736.

[19] L. Wang, C. Burger, and C. Stiller, "Reasoning about potential hidden traffic participants by tracking occluded areas," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 157–163.

[20] P. Narksri, E. Takeuchi, Y. Ninomiya, and K. Takeda, "Deadlock-free planner for occluded intersections using estimated visibility of hidden vehicles," *Electronics*, vol. 10, no. 4, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/4/411

[21] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2, p. 1805–1824, Aug 2000. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.62.1805

[22] L. Wang, C. F. Lopez, and C. Stiller, "Realistic single-shot and long-term collision risk for a human-style safer driving," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 2073–2080.

[23] D. Petrich, T. Dang, D. Kasper, G. Breuel, and C. Stiller, "Map-based long term motion prediction for vehicles in traffic environments," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, pp. 2166–2172.

[24] J. Quehl, H. Hu, S. Wirges, and M. Lauer, "An approach to vehicle trajectory prediction using automatically generated traffic maps," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 544–549.

[25] R. Krajewski, T. Moers, J. Bock, L. Vater, and L. Eckstein, "The round dataset: A drone dataset of road user trajectories at roundabouts in germany," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.

[26] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kümmerle, H. Königshof, C. Stiller, A. de La Fortelle, and M. Tomizuka, "INTERACTION dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," *CoRR*, vol. abs/1910.03088, 2019. [Online]. Available: http://arxiv.org/abs/1910.03088

[27] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[28] M. Althoff, M. Koschi, and S. Manzinger, "Commonroad: Composable benchmarks for motion planning on roads," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 719–726.

[29] J. Bernhard, K. Esterle, P. Hart, and T. Kessler, "Bark: Open behavior benchmarking in multi-agent environments," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 6201–6208.

Lingguang Wang studied Automotive engineering in Shanghai, China and Karlsruhe, Germany, and received a B. Sc. from Tongji University, Shanghai, China and M. Sc. from Karlsruhe Institute of Tech-nology, Karlsruhe, Germany in 2015 and 2018. He is currently pursuing a Ph.D. degree within the Institute of Measurement and Control Systems at Karlsruhe Institute of Technology, Germany, supervised by Prof. Dr.-Ing. Christoph Stiller. His research interests include risk assessment, motion planning and high-level decision making for autonomous driving.

Carlos Fernandez studied Computer Science Engi-neering at University of Alcala (Spain). He received the B. Sc. degree in 2008 and the M.Sc. in Advanced Electronics Systems and Intelligent Systems in 2010. His PhD was awarded with the highest mark CUM LAUDE in 2016 and it was focused on computer vi-sion applied to intelligent transportation systems and autonomous driving under the supervision of Prof. Dr. Miguel Angel Sotelo. Since 2017 he is group leader at Institute of Measurement and Control Sys-tems at Karlsruhe Institute of Technology, Germany.
He is reviewer of IEEE conferences and journals and his research interests include smart infrastructure, perception methods and trajectory prediction for autonomous driving.

Christoph Stiller received the Electrical Engineer-ing degree from Aachen, Germany, and Trondheim, Norway, and the Diploma and Dr. Ing. degrees from Aachen University of Technology, Aachen, Germany, 1988 and 1994, respectively. He became a Postdoctoral Scientist with the INRS Telecommuni-cations, Montreal, QC, Canada In 1994. In 1995, he joined the Corporate Research and Advanced Development of Robert Bosch GmbH, Hildesheim, Germany. In 2001, he became a chaired Professor at Karlsruhe Institute of Technology, Germany. In
2010, he spent three months by invitation at CSIRO in Brisbane, Australia. In 2015, he was a Guest Scientist for five months with the Bosch RTC and Stanford University, Palo Alto, CA, USA. He served as Editor-in-Chief of the IEEE Intelligent Transportation Systems Magazine (2009-2011) and as Associate Editor for the IEEE Transactions on Image processing (1999-2003), for the IEEE Transactions on Intelligent Transportation Systems (2004-2015), for the IEEE Intelligent Transportation Systems Magazine (2012-ongoing) and as Senior Editor for the IEEE Transactions on Intelligent Vehicles (2015-ongoing). His Autonomous Vehicle AnnieWAY was a finalist in the Urban Challenge 2007 and the winner and second winner of the Grand Cooperative Driving Challenge 2011 and 2016, respectively. In 2013, he collaborated with Daimler on the automated Bertha Benz Memorial Tour.