

INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps

Wei Zhan¹, Liting Sun¹, Di Wang^{2,*}, Haojie Shi^{3,*}, Aubrey Clausse⁴, Maximilian Naumann^{5,*}, Julius Kümmerle⁵, Hendrik Königshof⁵, Christoph Stiller⁵, Arnaud de La Fortelle⁴ and Masayoshi Tomizuka¹

Abstract—Interactive motion datasets of road participants are vital to the development of autonomous vehicles in both industry and academia. Research areas such as motion prediction, motion planning, representation learning, imitation learning, behavior modeling, behavior generation, and algorithm testing, require support from high-quality motion datasets containing interactive driving scenarios with different driving cultures. In this paper, we present an INTERnational, Adversarial and Cooperative moTION dataset (INTERACTION dataset) in interactive driving scenarios with semantic maps.

Five features of the dataset are highlighted. 1) The interactive driving scenarios are diverse, including urban/highway/ramp merging and lane changes, roundabouts with yield/stop signs, signalized intersections, intersections with one/two/all-way stops, etc. 2) Motion data from different countries and different continents are collected so that driving preferences and styles in different cultures are naturally included. 3) The driving behavior is highly interactive and complex with adversarial and cooperative motions of various traffic participants. Highly complex behavior such as negotiations, aggressive/irrational decisions and traffic rule violations are densely contained in the dataset, while regular behavior can also be found from cautious car-following, stop, left/right/U-turn to rational lane-change and cycling and pedestrian crossing, etc. 4) The levels of criticality span wide, from regular safe operations to dangerous, near-collision maneuvers. Real collision, although relatively slight, is also included. 5) Maps with complete semantic information are provided with physical layers, reference lines, lanelet connections and traffic rules.

The data is recorded from drones and traffic cameras, and the processing pipelines for both are briefly described. Statistics of the dataset in terms of number of entities and interaction density are also provided, along with some utilization examples in the areas of motion prediction, imitation learning, decision-making and planning, representation learning, interaction extraction and social behavior generation. The dataset can be downloaded via <https://interaction-dataset.com>.

I. INTRODUCTION

In order to enable fully autonomous driving in complex scenarios, comprehensive understanding and accurate prediction

¹W. Zhan, L. Sun, and M. Tomizuka are with the Mechanical Systems Control (MSC) Laboratory, Department of Mechanical Engineering, University of California, Berkeley, CA 94720 USA. (e-mail: wzhan@berkeley.edu).

²D. Wang is with Xi'an Jiaotong University, Xi'an, P.R. China.

³H. Shi is with Harbin Institute of Technology, Harbin, China.

⁴A. Clausse and A. de La Fortelle are with MINES ParisTech, Paris, France.

⁵M. Naumann, J. Kümmerle, H. Königshof and C. Stiller are with FZI Research Center for Information Technology and Karlsruhe Institute of Technology, Karlsruhe, Germany.

* The work was conducted during their visit to the MSC Lab at University of California, Berkeley.

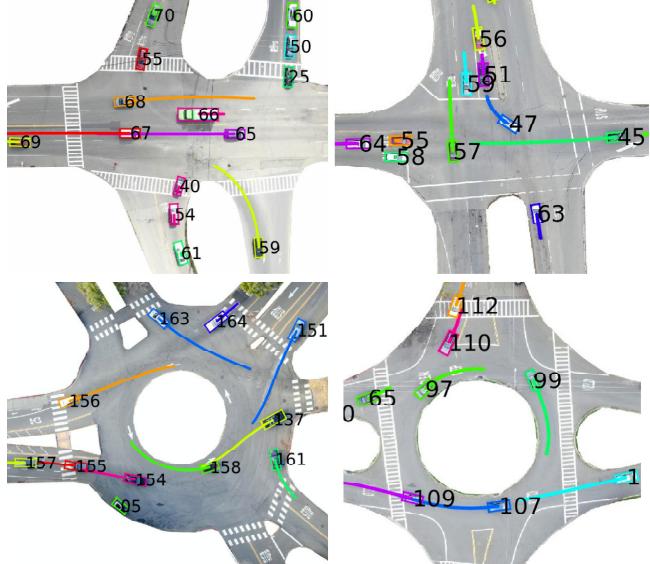


Fig. 1: Examples of the detection and tracking results in highly interactive driving scenarios in the dataset.

of the behavior and motion of other road users are required. Moreover, autonomous vehicles need to behave like vehicles with human drivers to make themselves more predictable to others and thus, facilitate cooperation. These are two of the major challenges in the field of autonomous driving. To overcome these challenges, considerable amount of research efforts have been devoted to: i) predicting the future intention and motion of other road users [1]–[3], ii) modeling and analyzing driving behavior [4], [5], iii) clustering the motion and finding representation of the motion primitives [6], [7], iv) cloning and imitating human and expert behavior [8], [9], and v) generating human-like and social behavior and motion [10]–[12].

All the aforementioned research areas require interactive vehicle motion data from real-world driving scenarios, which is the most fundamental and indispensable asset. NGSIM dataset [13] is the most popular one used in the aforementioned areas, such as prediction [14]–[16], behavior modeling [5], social behavior generation and planning [10], and representation learning [6], since it is publicly available with decent scale and quality. The recently released highD dataset [17]

also greatly assists behavior-related research such as prediction [18]. Public motion datasets such as NGSIM and highD facilitated, but also restricted behavior-related research due to limited diversity, complexity and criticality of the scenarios and behavior. Also, the importance of map information and completeness of interaction entities were under-addressed in most of the existing datasets. However, these missing points are crucial for behavior-related research, which will be discussed in the following.

1) *Diversity of interactive driving scenarios:* Recent behavior-related research using public datasets was mostly restricted to highway scenarios due to the data availability. There are many more highly interactive driving scenarios to explore, such as roundabouts with yield/stop signs, (unsignalized) one/two/all-way stop intersections (shown in Fig. 1), signalized intersections with unprotected left turn, zipper merge in cities, etc.

2) *International driving cultures:* Most of the existing datasets only contain driving data in one specific country. However, driving cultures in different countries and different continents can be distinct for very similar scenarios. Without motion data in similar scenarios from different countries, it is not possible to incorporate the impact of driving cultures in different countries, such as driving styles, preferences, risk tolerance, understanding of traffic rules, etc., for behavior modeling and analysis as well as the design of adaptive prediction and planning algorithms in different countries.

3) *Complexity of the scenarios and behavior:* Most of the scenarios in the existing public datasets are relatively simple and structured with explicit right-of-way. The behavior of the drivers is only occasionally impacted by others. There is very little social pressure (such as several vehicles waiting behind and even honking) on the drivers, so that their behavior is cautious without aggressive and irrational decisions. A motion dataset with much more complex and interactive behavior and scenarios is expected to facilitate the research tackling real and challenging problems.

4) *Criticality of the situations:* Critical situations (such as near-collision cases) are much more challenging and valuable than others for behavior-related research areas. For instance, [15] proposed a fatality-aware prediction benchmark emphasizing prediction inaccuracies in critical situations. However, critical situations are too sparse in existing motion datasets, and can hardly be identified. Therefore, a motion dataset with denser critical situations is necessary to facilitate the research efforts on those difficult problems.

5) *Map information:* Map information with references and semantics such as lanelet connections and traffic rules, are crucial for behavior-related research areas such as motion planning and prediction. It provides key information on input (features), such as route and goal point [9], distance to the merging point [14], [15], lateral position within the lane [10], etc., and makes the algorithms generalizable to other scenarios. Such semantic maps are currently missing for most of the existing public motion datasets.

6) *Completeness of interaction entities:* In order to ac-

curately model, predict and imitate the interactive vehicle behavior, it is crucial to provide motions of all surrounding entities which may impact their behavior in the dataset. This requirement was often overlooked when using motion data collected by onboard sensors due to occlusions and limited field of view of the sensors. Although existing motion datasets collected from onboard sensors contain data collected from a wide range of areas for long time periods, complete and meaningful interaction pairs are relatively sparse.

In this paper, we will emphasize all the aforementioned aspects to construct an international motion dataset collected by drones and traffic cameras.

- *Diverse and international:* It contains a variety of highly interactive driving scenarios from different countries, such as roundabouts, signalized/unsignalized intersections, as well as highway/urban merging and lane change.
- *Complex and critical:* Part of the scenarios are relatively unstructured with inexplicit right-of-way. The driving behavior in the dataset are highly impacted by other drivers, whose behavior can be aggressive or irrational due to the social pressure. Near-collision or slight-collision scenes are contained in the dataset to facilitate the research for critical situations.
- *Semantic map and complete information:* HD maps with semantics are provided to generated key features in the context. Motions of all entities which may influence the driving behavior are included in the dataset.

The proposed dataset can significantly facilitate behavior-related research such as motion prediction, imitation learning, decision-making and planning, representation learning, interaction extraction and social behavior generation. Results from exemplar methods in all these areas are provided utilizing the proposed dataset.

II. RELATED WORK

A. Datasets from Bird's Eye View

As mentioned in Section I, NGSIM dataset [13] is the most popular vehicle motion dataset among the behavior-related research communities. The raw data was collected by cameras mounted on buildings and processed automatically [20]. The accuracy of the dataset is mostly acceptable. However, there may be steady errors, and the image projection can significantly enlarge the size of the vehicles. Researchers proposed methods [21] to rectify the errors, but it can only improve the quality of a small part of the dataset. In view of the problems in NGSIM, highD dataset [17] was constructed by using a drone with more accurate vehicle motions and larger amount of high way driving data than NGSIM. Other datasets [22], [23] from bird's eye view are more focused on pedestrian behavior without strong vehicle interactions.

The driving scenarios presented in NGSIM and highD are quite limited. NGSIM contains highway driving (including ramp merging and double lane change) and signalized intersection scenarios. In fact, signalized intersections are mostly controlled by the traffic lights and interactions are very rare

TABLE I: Comparison with existing motion datasets

	highly interactive scenarios	complexity of scenarios	density of aggressive behavior	near-collision situations and collisions	HD maps with semantics	completeness of interaction entities & viewpoint
NGSIM [13]	ramp merging, (double) lane change	structured roads, explicit right-of-way	low	very few near-collision	no	yes, bird's-eye-view from a building
highD [17]	lane change	structured roads, explicit right-of-way	low	very few near-collision	no	yes, bird's-eye-view from a drone
Argoverse [19]	unsignalized intersections, pedestrian crossing	unstructured roads, inexplicit right-of-way	low	no	yes, but partially	only for the ego data-collection vehicle
INTERACTION	roundabouts, ramp merging, double lane change unsignalized intersections	unstructured roads, inexplicit right-of-way	high	yes	yes	yes, bird's-eye-view from a drone

and slight. A small amount of lane changes are interactive, but most of them are neither interactive nor critical. Ramp merging and double lane change can be highly interactive when the traffic is relatively dense, but the amount of interaction is still relatively limited in NGSIM. HighD only contains highway driving scenarios with car following and lane change. Urban scenarios which contain densely and highly interactive behavior, such as roundabouts and unsignalized intersections are not included in either of the two public datasets of vehicle motions.

B. Datasets from Onboard Sensors

In addition to the bird's-eye-view motion datasets, two types of onboard-sensor-based ones are also publicly available. One includes motion data of surrounding entities from onboard LiDARs and front-view cameras, such as Argoverse [19] and HDD dataset [24]. The other only contains motions of many data-collection vehicles from onboard GPS, such as 100-car study [25].

There are two major advantages for datasets from onboard sensors. One is that a variety of driving scenarios with relatively long data recording time are usually included in those datasets, such as urban driving at signalized/unsignalized intersections and highway driving with ramp merging, etc. The other is that the occlusions of LiDARs and cameras are recorded so that the actual occlusions from perspective of the ego vehicle can be partially recovered.

Completeness of interaction entities is a major problem when using datasets from onboard sensors for behavior-related research. For motion datasets with GPS-based fleets, it is hard to determine whether the vehicles in an "interactive" motion segment was actually interacting with each other since there is no motion recording of other surrounding vehicles (or even pedestrians) without GPS devices installed. For motion datasets constructed from onboard LiDARs and cameras, it is hard to guarantee that all the surrounding objects impacting the behavior of other vehicles are included in the dataset when predicting the motions of others. Therefore, complete interactions are relatively sparse in such kind of datasets. If the sensors cannot cover the full field of view, it will be even impossible to guarantee the completeness of information for the surrounding entities of the ego data collection vehicle.

Also, the data collected in a large area may lead to very few repetitions at the same location. It is hard to learn multimodal driving behavior for prediction or planning since only

one sequence of motions can be found with similar features at the same location.

Map information is also missing in most of the motion datasets. To the best of our knowledge, Argoverse is the only motion dataset providing relatively rich map information. Physical layer (locations of curbs, road markings, etc.) is contained and semantic information (lane bounds and turn directions, etc.) required by prediction and planning is partially included.

Table I provides a comparison of the three most useful public vehicle motion datasets as well as the one presented in this article. The proposed dataset contains much more diverse, complex and critical scenarios and vehicle motions comparing to the other three. In addition, HD maps with full semantic information are provided, and the completeness of interaction entities is superior to datasets from onboard sensors.

III. FEATURES OF THE DATASET

In this section, we will illustrate the features of the proposed dataset by highlighting the diversity, internationality, complexity, criticality, and semantic map.

A. Diversity

Fig. 2 illustrates a variety of highly interactive driving scenarios from traffic cameras and drones in our dataset, including zipper merging in a city (Fig. 2 (a)), ramp merging and lane change on a highway (Fig. 2 (b)), five roundabouts with yield and stop signs (Fig. 2 (c) - (g)), several unsignalized intersections with one/two/all-way stops (Fig. 2 (h) - (j)), and unprotected left turn at a signalized intersection (Fig. 2 (k)). In Fig. 2, the first two letters of the names represent the sources of the data (drone as *DR* and traffic camera as *TC*), while next three letters represent the corresponding country and the last two represent the scenario code in the dataset. The numbers in circles denote the branch ID for each scenario.

Fig. 2 (b) contains several subscenarios. The subscenario with the upper two lanes (that merge into one finally) is a zipper merging which is similar to the urban counterpart in Fig. 2 (a), where vehicles strongly interact with each other. It is also a ramp for the middle two lanes. The subscenario with the lower three lanes (that merge into two finally) is a forced merging and vehicles have to change their lanes.

The roundabout in Fig. 2 (f) is an extremely busy 7-way roundabout with one "yield" branch and six "stop" branches.

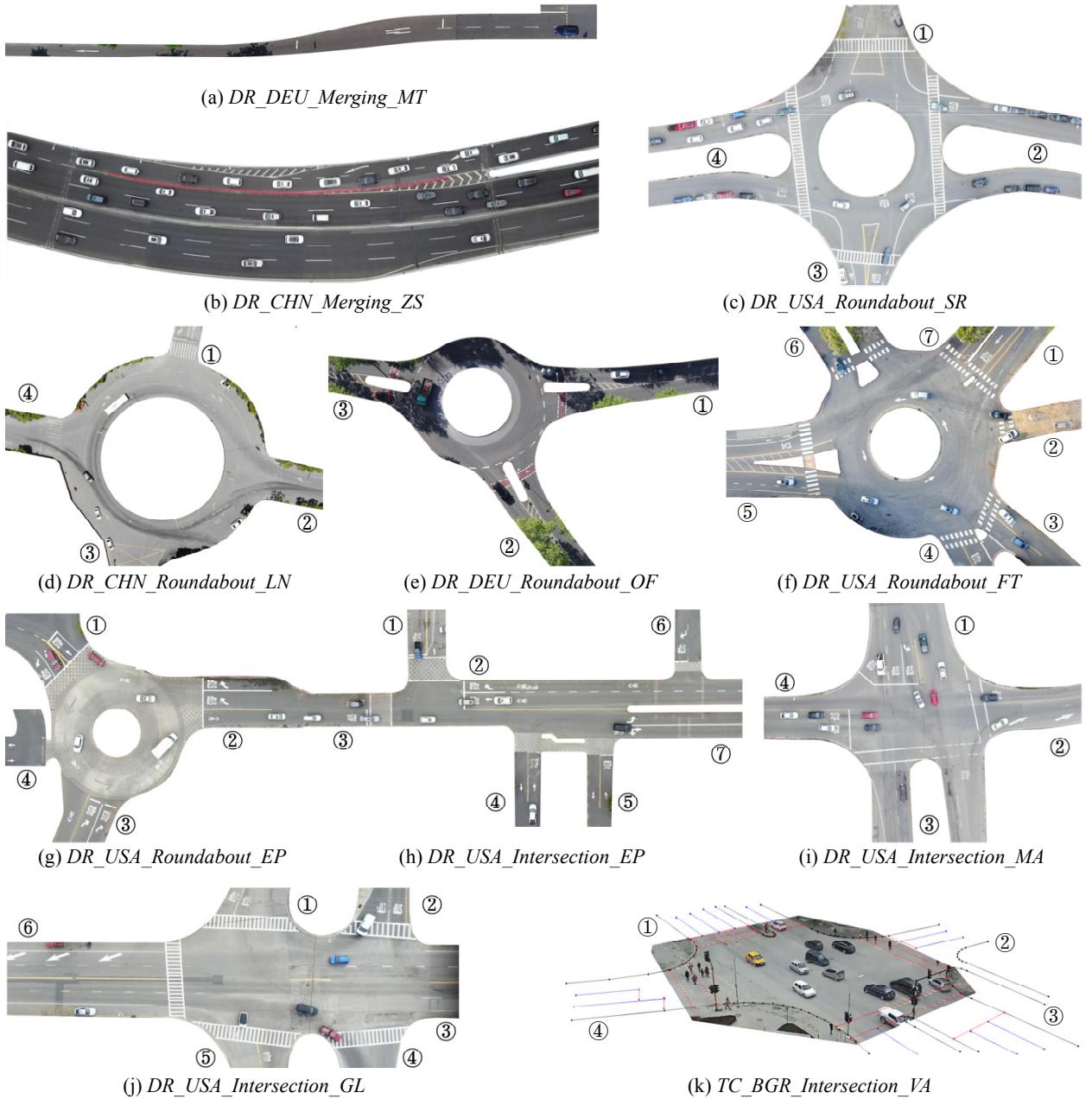


Fig. 2: A variety of highly interactive driving scenarios recorded by drones in the dataset, including: (a) urban merging, (b) highway ramp merging and lane change, (c)-(g) five roundabouts, and (h)-(j) unsignalized intersections, and (k) unprotected left turn at a signalized intersection.

Lots of vehicles enter the roundabout at the same time with intensive interactions and relatively high speeds. The branches of the roundabouts in Fig. 2 (c)-(e) are controlled by yield signs, while all branches of the roundabout in Fig. 2 (g) are controlled by stop signs.

Figure 2 (i) shows an extremely busy all-way-stop intersection with 9 lanes controlled by stop signs. Multiple vehicles are interactively inching to compete. The scenario shown in Fig. 2 (j) contains three branches (Branch 1, 2, 5) controlled by stop signs, while vehicles from Branch 3 and 6 have the right-of-

way (RoW). Lots of vehicles are entering the intersections from all branches (except Branch 4), and vehicles holding RoW on the straight road are with relatively high speed. A busy all-way-stop T-intersection is shown in Fig. 2 (h), while three other branches (Branch 4-6) are also controlled by stop signs.

B. Internationality

The motion data was collected from three continents (North America, Asia and Europe). Motion data collected by drones

are from four countries, namely, the US, China, Germany and Bulgaria, as indicated in the names of the scenarios (*USA/CHN/DEU/BGR*). Vehicles in all these countries are driven on the right-hand side of the road. However, driving culture in these countries is with remarkable distinctions.

We provide motion data from three roundabouts with similar traffic rules, namely, *SR* from the US, *OF* from Germany and *LN* from China. All the three roundabouts do not have stop signs, and the nominal traffic rule is that the vehicles entering the roundabout should yield the ones which is already in the roundabout.

We also provide motion data from two zipper merging scenarios, those are, *MT* from Germany and *ZS* from China (the upper two lanes in Fig. 2 (b)). Although *MT* is urban road and *ZS* is the entrance of highway, the “zipper” rule remains the same, and the speeds are similar when the traffic is heavy.

C. Complexity

In addition to regular driving behavior such as car-following, lane change, stop and left/right/U-turn, our dataset emphasizes highly interactive and complex driving behavior with cooperative and adversarial motions of the vehicles. By carefully choosing the locations and corresponding rush hours for the data collection, we were able to gather large amounts of strong interactions within relative short period of time. Strongly interactive pairs of vehicles can even appear every few seconds from time to time for scenarios such as the ramp in *ZS*, the entrance branches in *FT*, the all-way-stop intersections in *EP* and *MA* as well as the two-way-stop intersection in *GL*.

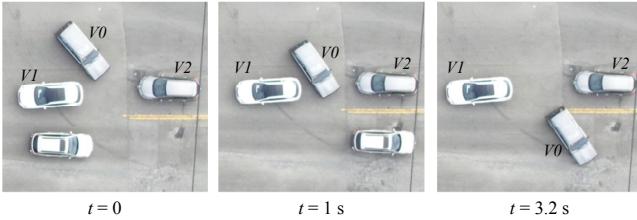


Fig. 3: A sequence of images of a dangerous insertion in *GL* in the proposed dataset.

Also, scenarios in *FT* and *GL* are relatively unstructured since there is no explicit lane restrictions in the roundabout or intersections. Vehicles can exploit the space to achieve their goals, sometimes showing irrational and highly dangerous behavior. For instance, Fig. 3 shows a dangerous insertion of *V0* between two vehicles (*V1* and *V2*) stopping and making left turns from Branch 3 to Branch 5 in *GL* (refer to Fig. 2). The driver of *V0* intended to drive from Branch 1 to Branch 4 but there was no explicit road structure for the driver.

Moreover, aggressive or irrational behavior can often be found due to inexplicit nominal or practical RoW. Vehicles may arrive at the stop bars almost at the same time and drivers may negotiate with each other by inching or even accelerating in *MA* and *EP*. The traffic in *FT* and *GL* can be very busy and

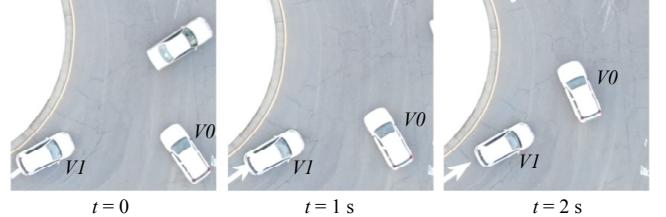


Fig. 4: A sequence of images of a violation for the right-of-way in a roundabout in the proposed dataset.

it may take even minutes for the vehicle without nominal RoW to enter and pass, making the driver impatient. Also, there may be a queue of vehicles waiting behind and even honking to put social pressures to the one in the front of queue. Although there are explicit traffic rules on who goes first for roundabouts or 2-way-stop intersections, vehicles without nominal RoW may be aggressive, and vehicles with nominal RoW are mostly aware of such potential violations and are ready to react. For example, *V0* in Fig. 4 was entering the roundabout in *FT* from Branch 3, while *V1* was in the roundabout holding the RoW. However, *V0* violated the rule and forced *V1* to stop and yield.

Those factors significantly increase the complexity of the motions in the dataset and bring forward lots of challenging but valuable research topics for the community.

D. Criticality

As discussed in Section III-C, vehicles holding the nominal RoW (in the roundabout of *FT* or on the straight road of *GL*) may often encounter slight violations from vehicles without nominal RoW (entering the roundabout or intersection from branches controlled by stop signs). Moreover, the vehicles holding the RoW may have relatively high speed (40 km/h or even higher). Therefore, critical situations can be observed in the dataset where time-to-collision-point (TTCP) can be extremely low. A slight collision can even be found in the dataset.

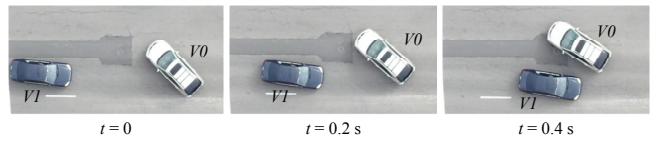


Fig. 5: A sequence of images of a near-collision case in the proposed dataset.

Fig. 5 shows a near-collision case in *GL*. *V0* was making a left turn from Branch 5 (with a stop sign) to Branch 6, while *V1* (with the RoW) was going straight forward from Branch 6 to Branch 3 with a relatively high speed. *V1* had to execute emergency swerve to avoid the collision with *V0*, which was very dangerous.

Besides the critical, near-collision cases, a slight collision shown in Fig. 6 can also be found in the dataset in *GL*. *V0* was making a right turn from Branch 5 (with a stop sign) to



Fig. 6: A sequence of images of a slight collision in the proposed dataset.

Branch 3, while $V1$ (with the RoW) was making a right turn from Branch 6 to Branch 4. In this situation, the driver of $V0$ might have predicted that $V1$ was going straight to Branch 3, so that $V0$ could accelerate in advance.

E. Semantic Map

Map information is crucial for behavior-related research areas. The information required is twofold. The basic requirement is the physical layer containing a set of points or curves representing curbs, road markings (lane markings, stop bars, etc.) and other key features. In addition to the physical layer, semantic information is also necessary, which includes but is not limited to, 1) reference paths, 2) lanelets as well as their connections and turn directions, 3) traffic rules and RoW associated, etc. Moreover, such information needs to be organized with consistent format and toolkit to facilitate the users when utilizing the map. All the aforementioned requirements are met in our dataset, and more detailed information on map construction can be found in Section IV-C.

IV. CONSTRUCTION OF MOTION DATA AND MAPS

In this section, we will discuss the pipeline for constructing the motion data from both drones and traffic cameras, as well as the corresponding semantic maps.

A. Motions from Drone Data

We used drones such as DJI Mavic 2 and DJI Phantom 4 to collect the raw video data. The raw videos were 4K (3840x2160) by 30Hz. We downsampled the video to 10 Hz and process the data. The processed results are partially illustrated in Fig. 1. The bounding boxes are very accurate and the paths are smooth after going through our processing pipeline with the following three steps.

- *Video stabilization and alignment:* Due to gradual or sudden drift and rotation of drones, the collected videos need to be stabilized via video stabilization algorithms with transformation estimator. Also, similarity transformation is applied to project all the frames to the first one and aligned with the map.
 - *Detection:* In order to obtain accurate bounding boxes of the moving obstacles, Faster R-CNN [26] is applied. The boxes are highly accurate, and very few inaccurate detections are rectified manually.
 - *Data association, tracking and smoothing:* Kalman filter is applied for data association and tracking. To obtain smooth motions of the vehicles, a Rauch-Tung-Striebel (RTS) smoother [27] is also incorporated.

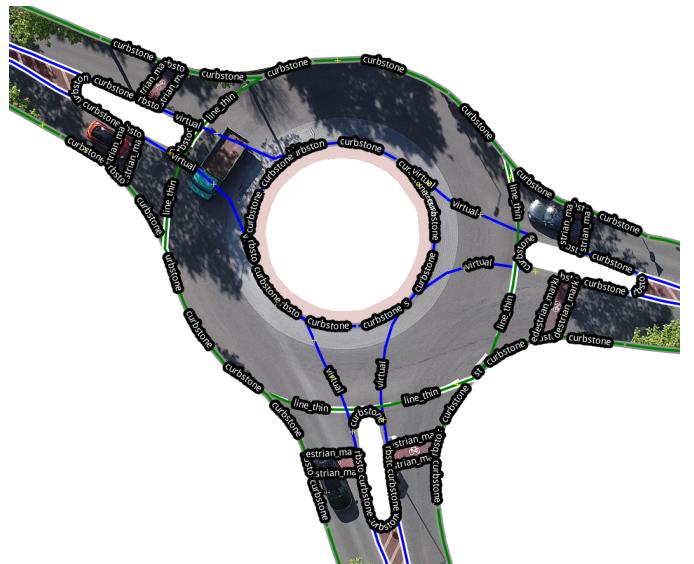


Fig. 7: An exemplary physical layer of a lanelet2 map [34].

B. Motions from Traffic Camera Data

The data processing pipeline for motions from traffic camera data mainly contains the following steps, and more details, including the camera parameter estimation, can be found in [28].

- *Detection:* To detect vehicles and pedestrians in each frame, we use a state-of-the-art object detector [29], which provides detections with 2D bounding box, instance mask and instance type.
 - *Data association:* Detections are grouped into tracks using a combination of an Intersection-over-Union [30] tracker which associates detections with high mask overlap in successive frames, and a visual tracker [31] to compensate for miss detections.
 - *Tracking and smoothing:* Once detections are grouped into tracks, trajectories on the ground plane are estimated using a RTS smoother. For the observation model, we use a pin-hole camera model [32]. This allows to incorporate measurements and uncertainty directly in pixels, capturing the uncertainty due to the resolution, position and orientation of the camera. For vehicles, the RTS smoother uses a bicycle model [33] as process model, allowing to capture the kinematics constraints of vehicles.

C. Construction of the High Definition Maps

As public roads are structured environments, the particular road layout of a certain area strongly affects the motion of all traffic participants. The structure for vehicles mostly starts by subdividing the road into lanes, and later combining them to create junctions, roundabouts, on ramps and so on. Further, movement within this structured area is guided by traffic rules, such as speed limits or prioritizing one road over another. In order to model such coherence, simply mapping center-lines of all lanes is not sufficient anymore.

TABLE II: Summary of the dataset.

Scenarios	Locations	Video length (min)	number of vehicles	Total video length (min)	Total number of vehicles
roundabout	USA_Roundabout_SR	40.90	965	365.1	10479
	CHN_Roundabout_LN	24.24	227		
	DEU_Roundabout_OF	55.04	1083		
	USA_Roundabout_FT	207.62	7496		
	USA_Roundabout_EP	37.30	708		
unsignalized intersection	USA_Intersection_EP	66.53	1367	433.33	14867
	USA_Intersection_MA	107.37	2982		
	USA_Intersection_GL	259.43	10518		
merging and lane change	DEU_Merging_MT	37.93	574	132.55	10933
	CHN_Merging_ZS	94.62	10359		
signalized intersection	TC_Intersection_VA	60	3775	60	3775

Thus, in order to allow for a thorough analysis of the recorded trajectories, we provide centimeter-accurate high definition maps in the lanelet2 format [34]. Within lanelet2, the physical layer of the road network, such as road borders, lane markings and traffic signs is stored. An exemplary physical layer is visualized in Figure 7. From this layer, atomic lane elements, called *lanelets*, are created. They describe the course of the lane and form the basis for so called regulatory elements, which determine traffic regulations such as the right of way or the speed limit.

When used alongside the recorded trajectories, these lanelet2 maps facilitate the reasoning about why some vehicles decelerate while approaching a junction, or why others do not, depending on the right of way but also on the presence of other traffic participants that potentially interact.

V. STATISTICS OF THE DATASET

A. Scenarios and Vehicle Density

The dataset contains motion data collected in four categories of scenarios: roundabout, unsignalized intersection, signalized intersection, merging and lane change, as shown in Fig. 2. A detailed summary of the dataset is listed in Table II. In the roundabout scenarios, 10479 trajectories of vehicles from five different locations were recorded for around 365 minutes. Similarly, in the unsignalized intersection scenarios, three locations were included and 14867 trajectories were collected for around 433 minutes. In the merging and lane change scenarios, 10933 trajectories were recorded at two locations for around 133 minutes. Finally, one location was selected for the signalized intersection, which provided 3775 trajectories for around 60 minutes.

B. Metrics for Interactive Behavior Identification

To represent the density of the interactive behavior of the proposed dataset, we use the metric - number of interaction pairs per vehicle (IPV) as in proposed in [35]. To calculate the IPV, a set of rules were proposed in [35] to extract the interactive behavior under different spatial representations of vehicle paths. The set of rules and metric are briefly reviewed below.

- 1) Minimum time-to-conflict-point difference ($\Delta TTCP_{\min}$): $\Delta TTCP_{\min}$ is a metric to describe the relative states of two moving vehicles in a scenario

where the paths of the two vehicles share a conflict point but without any forced stop. As shown in Fig. 8, such vehicle paths include two categories: (1) paths with static crossing or merging points such as intersections (Fig. 8 (a)-(b)), and (2) paths with dynamic crossing or merging points such as ramping and lane-changing, as shown in Fig. 8 (c)-(d). In such scenarios, merging can happen anywhere in the shaded area. We define $\Delta TTCP_{\min}$ as

$$\begin{aligned} \Delta TTCP_{\min} &= \min_{t \in [T_{\text{start}}, T_{\text{end}}]} \Delta TTCP^t \\ &= \min_{t \in [T_{\text{start}}, T_{\text{end}}]} (TTCP_1^t - TTCP_2^t) \quad (1) \end{aligned}$$

where $TTCP_i^t = \Delta d_i^t / v_i^t, i = 1, 2$ is the traveling time to the conflict point of each vehicle in the interactive pairs. v_i^t and Δd_i^t are, respectively, the speed of the i -th vehicle and its distance to the conflict point along the path at time t . For the scenarios with dynamic merging points, we use the actual merging points of the vehicle trajectories as the conflict points. In (1), T_{start} and T_{end} are set to be long enough to cover the interaction period between vehicles. If $TTCP_{\min} \leq 3$ s, then it is defined that interaction exists.

- 2) Waiting Period (WP): WP is a metric for vehicles with forced stops along their paths. In [35], the default waiting period at stops was set as 3 s, and the behavior deviation from the default one was used as an indicator of the interactivity, i.e., interaction exists when $WP > 3$ s.

C. Distribution of Interactivity

Based on the set of rules, there are 13375 interactive pairs of vehicles in the proposed dataset. We compare the interactivity among three datasets: the proposed INTERACTION dataset, the highD dataset, and the NGSIM dataset. Results are shown in Fig. 9, where the x-axis represents the length of $\Delta TTCP_{\min}$ in seconds, and the y-axis are the number of vehicles (Fig. 9 (a)) and the density of vehicles¹ (Fig. 9 (b)), respectively. We can see that the INTERACTION dataset contains more intensive interactions with $\Delta TTCP_{\min} \leq 1$ s.

¹The density is given by:

$$\text{density} = \frac{\text{number of vehicles with particular } \Delta TTCP_{\min}}{\text{total number of vehicles in the dataset}}$$

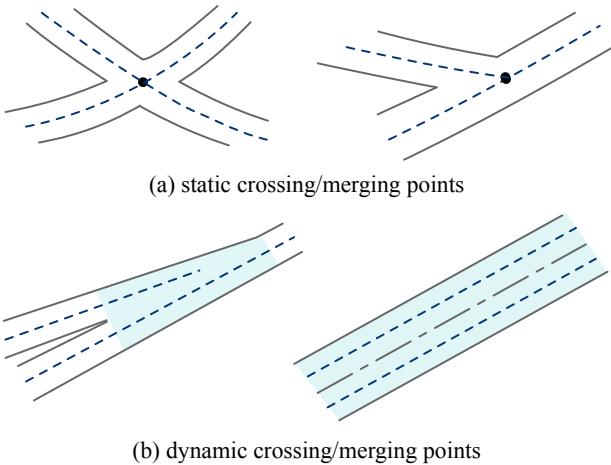


Fig. 8: Geometry of different interactive paths. In (a), the crossing/merging points between two paths are static and fixed, while in (b), the crossing/merging points are dynamic.

We also summarized the distributions of $\Delta TTCP_{\min}$ and WP of all vehicles in the dataset over different driving scenarios. The results are shown in Fig. 10. Similarly, the x-axis represents the length of $\Delta TTCP_{\min}$ and WP in seconds, and the y-axis is the density of vehicles in each scenario. We can see that the dataset contains highly interactive trajectories with a high density of $\Delta TTCP_{\min} \leq 1$ s, and WP greater than 3 s.

VI. UTILIZATION EXAMPLES

The proposed dataset is intended to facilitate researches related to driving behavior, as mentioned in Section I. In this section, we provide several utilization examples of the proposed dataset, including motion/trajetory prediction, imitation learning, motion planning and validation, motion clustering and representation, interaction extraction and human-like behavior generation.

A. Motion Prediction and Behavior Analysis

Motion/trajetory prediction is of vital importance for autonomous vehicles, particularly in situations where intensive interaction happens. To obtain an accurate probabilistic prediction model of vehicle motion, both learning- and planning-based approaches have been extensively explored. By providing high-density interactive trajectories along with HD semantic maps, the proposed dataset can be used for both approaches.

For instance, [36] proposed a deep latent variable model based on Wasserstein auto-encoder (WAE) to improve the interpretability. It incorporated the structure of recurrent neural network with vehicle kinematic model such that the output can be constrained. The motion data in *FT* was utilized to train and test the model in comparison with other state-of-the-art models such as variational auto-encoder (VAE), auto-encoder, and generative adversarial network (GAN). Quantitative results shown in Section VI-A demonstrated that the proposed WAE-based method can outperform other state-of-the-art models,

TABLE III: Comparisons of prediction accuracy from [36].

Methods	features	RMSE	MAE
WAE-based approach	x	0.013/0.011	0.046/0.035
	y	0.006/0.014	0.019/0.041
	ψ	0.006/ 0.008	0.018/0.042
VAE	x	0.018/0.016	0.25/0.22
	y	0.006/0.003	0.14/0.22
	ψ	0.006/ 0.008	0.13/0.21
Auto-encoder	x	0.315/0.044	1.026/0.315
	y	0.057/0.141	0.182/0.479
	ψ	0.011/0.066	0.037/0.078
GAN	x	0.024/0.020	0.324/0.273
	y	0.007/0.017	0.188/0.241
	ψ	0.005/0.048	0.107/0.286

when comparing the root mean square error (RMSE) and mean absolute error (MAE) of the prediction for position and yaw angle.

On the other hand, [37] took advantage of the HD semantic maps and combined the learning-based and the planning-based prediction methods. A deep learning model based on conditional variational auto-encoder (CVAE) and an optimal planning framework based on inverse reinforcement learning are dynamically combined to predict both irrational and rational behavior of the vehicles. Benefiting from the HD semantic information, features for the deep learning model were defined in Frenet frame, which generated much better prediction performance in terms of generalization. Some exemplar results are given in Fig. 11.

B. Imitation Learning

The driving behavior in the proposed dataset can also be used for imitation learning which directly imitates how human drive in complicated scenarios. We extended the fast integrated learning and control framework proposed in [38] in the *FT* roundabout scenario. As shown in Fig. 12, both the semantic HD map information and the states of surrounding vehicles (the red boxes) were included as the features. The grey box represents the current position of the ego vehicle. The green boxes and blue boxes, respectively, are the ground truth future positions and generated future positions of the ego vehicle via the imitation network.

C. Validation of Decision and Planning

Besides motion prediction and imitation, the motion data and maps in the dataset can also be used for testing different decision making and motion planning algorithms. The data-replay motions in the dataset are more suitable to test the performances of the decision-maker and planner when the motions of surrounding entities are independent of the ego motions. For example, the motion of the ego vehicle may not effect others when it does not have the RoW, or it has the RoW but others violate the rules or ignore the ego motion.

The environmental representation and motion planning methods proposed in [39] were tested in the *FT* roundabout

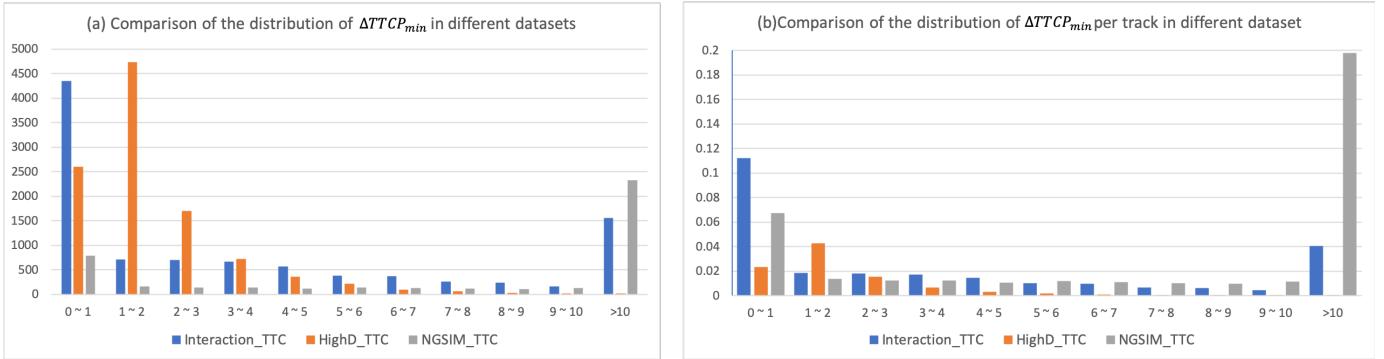


Fig. 9: Distribution of the $\Delta TTCP_{min}$ in three vehicle motion datasets: the proposed INTERACTION dataset, the HighD dataset and the NGSIM dataset.

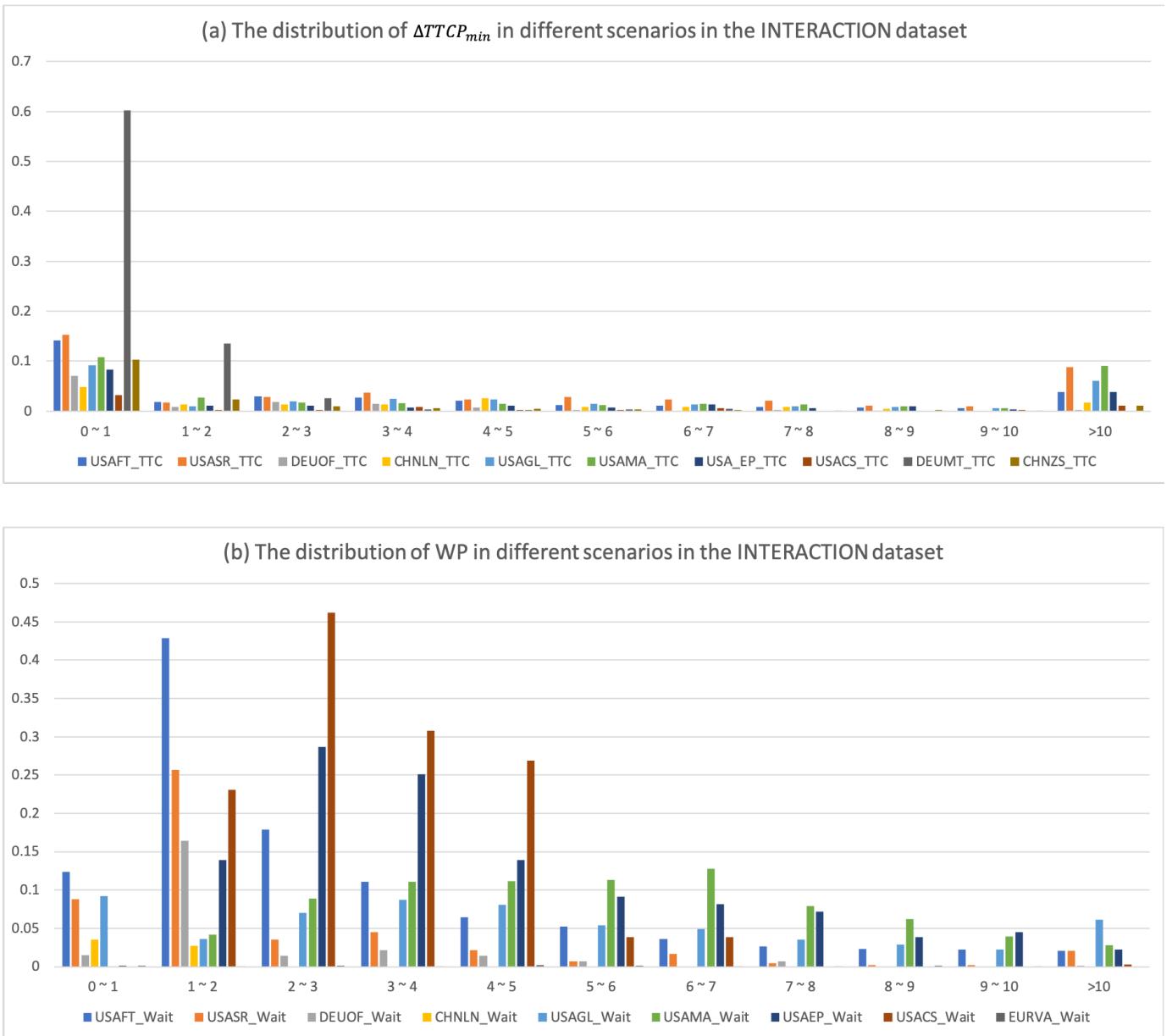


Fig. 10: Distribution of the $\Delta TTCP_{min}$, and WP across different locations and scenarios in the dataset.

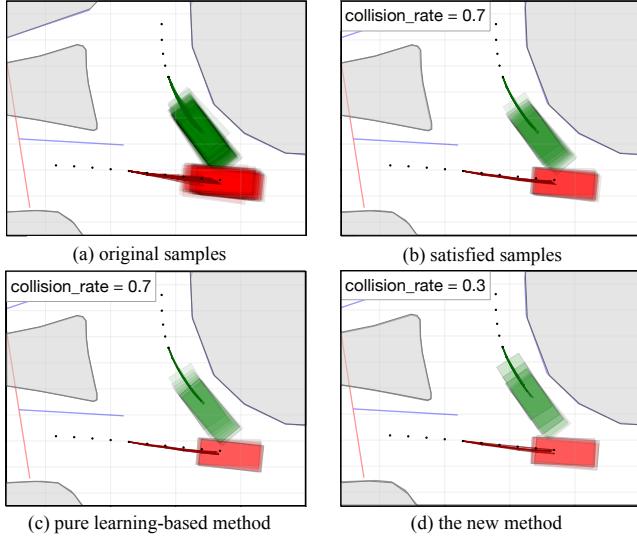


Fig. 11: Some exemplar prediction results from [37].

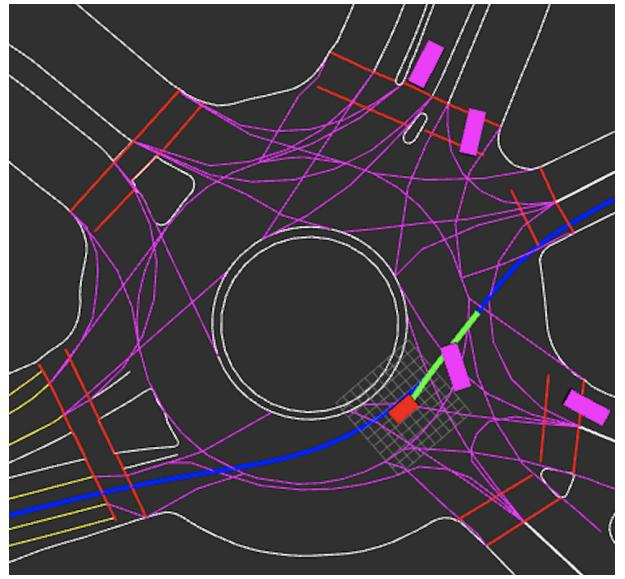


Fig. 13: A screen-shot of simulation when testing the motion planner in [39] with the proposed dataset.

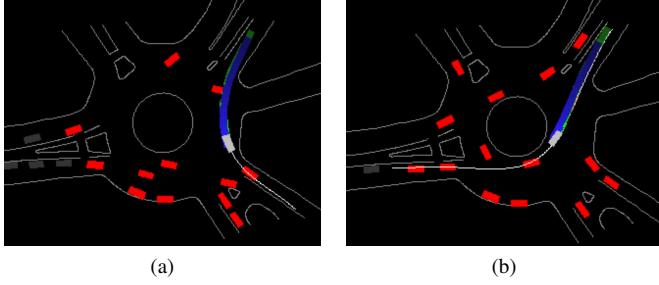


Fig. 12: Two examples of the imitation learning results by employing the method in [38].

scenario. Fig. 13 is a bird's-eye-view screen-shot of the simulation. The red rectangle represents the autonomous vehicle with the planner in [39]. It was decelerating to avoid the collision with a vehicle entering the roundabout although it has the RoW.

We also combined the integrated decision and planning framework proposed in [40] and the sample-based motion planner proposed in [41] to design the decision-maker and planner under uncertainty. The predictor was designed according to [42] based on dynamic Bayesian network (DBN) to provide the probabilities of the intentions of others.

Figure 14 shows the results of the planned speed profile with corresponding bird's-eye-view screenshots of the situations at specific time steps. The host autonomous vehicle was entering the FT roundabout, and the vehicle in the roundabout, retrieved from the proposed dataset, was exiting. When it was not clear whether the target vehicle was going to exit or not, such as the time step in Fig. 14 (a), the predictor returned $P(\text{exit})$ as 0.626. With the non-conservatively defensive strategy proposed in [40], the ego vehicle was able to keep accelerating to enter the roundabout as planned for the next 0.5 s, so that the potential threat with low probability (the target vehicle stays

in the roundabout) did not affect the efficiency and comfort of the ego vehicle. The long-term planning corresponding to yielding case (red curve in Fig. 14 (b)) guaranteed that the ego vehicle was able to fully stop for the worst case.

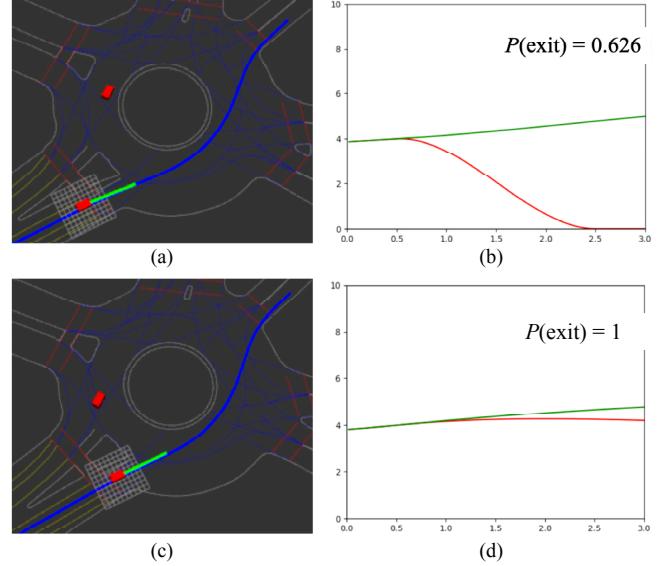


Fig. 14: Screenshots of the situations and corresponding planned speed profiles by implementing decision and planning methods in [40], [41] with predictor in [42] by utilizing the proposed dataset.

D. Motion Clustering and Representation Learning

The X-means algorithm [43] was employed to cluster the trajectories and obtain motion patterns with results shown in Fig. 15. We constructed a feature space with vehicle motions in Frenét Frame based on map information. Fig. 15 (a) shows

the clustered trajectory segments in different colors with the map. Fig. 15 (b) and (d) demonstrate the cluster results with longitudinal positions and speeds of the two interacting vehicles as the coordinates. The clustering results with the first and second components of principle component analysis (PCA) for the feature space are shown in Fig. 15 (c). In the figures we can see that different interactive motions are separated and similar ones are clustered, which are desirable results to obtain motion patterns.

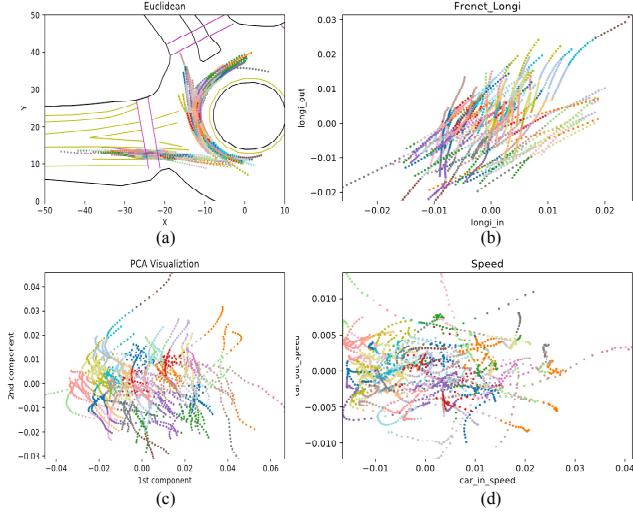


Fig. 15: Results of X-means [43] motion clustering using the proposed dataset.

E. Extraction of Interactive Agents and Trajectories.

The proposed dataset can also be used to learn the interaction relationships between agents. We implemented the learning method and network structure proposed in [44] to extract the interaction frames of two agents. Some example results are given in Fig. 16, where Fig. 16 (a) and (b) provide one exemplar pair of interacting cars in the *FT* scenario, while Fig. 16 (c) and (d) represent another pair. In Fig. 16 (a) and (c), the paths of both of the interacting cars are provided, and in Fig. 16 (b) and (d), the trajectories along longitudinal directions are shown. We can see that the extracted interaction frames (purple circle) align quite well with the ground truth frames (blue star).

F. Human-like Decision and Behavior Generation

We can also learn decision-making models that generate human-like decisions and behaviors with the proposed dataset. In [45], an interpretable human behavior model was proposed based on the cumulative prospect theory (CPT). As a non-expected utility theory, CPT can well explain some systematically biased or “irrational” behavior/decisions of human that cannot be explained by the expected utility theory. Parameters of three different models were learned and tested using the data in the *FT* roundabout scenario: a predefined model based on time-to-collision-point (TTCP), a learning-based model based

on neural networks, and the proposed CPT-based model. The results (Fig. 17) showed that the CPT-based model outperformed the TTCP model and achieved similar performance as the learning-based model with much less training data and better interpretability.

VII. CONCLUSION

In this paper, we presented a motion dataset in a variety of highly interactive driving scenarios from the US, Germany, China and other countries, including signalized/unsignalized intersections, roundabouts, ramp merging and lane change from cities and highway. Complex interactive motions were captured, featuring inexplicit right-of-way, relatively unstructured roads, as well as aggressive and irrational behavior caused by impatience and social pressure. Critical (near-collision and slight-collision) situations can be found in the dataset. We also included high-definition (HD) maps with semantic information for all scenarios in our dataset. The data was recorded from drones and traffic cameras and the data processing pipeline was briefly described. Our map-aided dataset with diversity, internationality, complexity and criticality of scenarios and behavior can significantly facilitate driving-behavior-related research such as motion prediction, imitation learning, decision-making and planning, representation learning, interaction extraction, and human-like behavior generation, etc. Results from various kinds of methods of these research areas were demonstrated utilizing the proposed dataset.

VIII. ACKNOWLEDGEMENT

The authors also would like to thank the Karlsruhe House of Young Scientists (KHYS) for their support of Maximilian’s research visit at MSC Lab.

REFERENCES

- [1] S. Lefèvre, D. Vasquez, and C. Laugier, “A survey on motion prediction and risk assessment for intelligent vehicles,” *ROBOMECH Journal*, vol. 1, no. 1, pp. 1–14, Jul. 2014.
- [2] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, “Human motion trajectory prediction: A survey,” *arXiv preprint arXiv:1905.06113*, 2019.
- [3] W. Zhan, A. de La Fortelle, Y.-T. Chen, C.-Y. Chan, and M. Tomizuka, “Probabilistic prediction from planning perspective: Problem formulation, representation simplification and evaluation metric,” in *Intelligent Vehicles Symposium (IV), 2018 IEEE*, 2018, pp. 1150–1156.
- [4] H. Okuda, N. Ikami, T. Suzuki, Y. Tazaki, and K. Takeda, “Modeling and Analysis of Driving Behavior Based on a Probability-Weighted ARX Model,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 98–112, Mar. 2013.
- [5] K. Driggs-Campbell, V. Govindarajan, and R. Bajcsy, “Integrating Intuitive Driver Models in Autonomous Planning for Interactive Maneuvers,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3461–3472, Dec. 2017.
- [6] Q. Lin, Y. Zhang, S. Verwer, and J. Wang, “MOHA: A Multi-Mode Hybrid Automaton Model for Learning Car-Following Behaviors,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–8, 2018.
- [7] W. Wang, W. Zhang, and D. Zhao, “Understanding V2V Driving Scenarios through Traffic Primitives,” *arXiv:1807.10422 [cs, stat]*, Jul. 2018, arXiv: 1807.10422.
- [8] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer, “HG-Dagger: Interactive Imitation Learning with Human Experts,” to appear in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

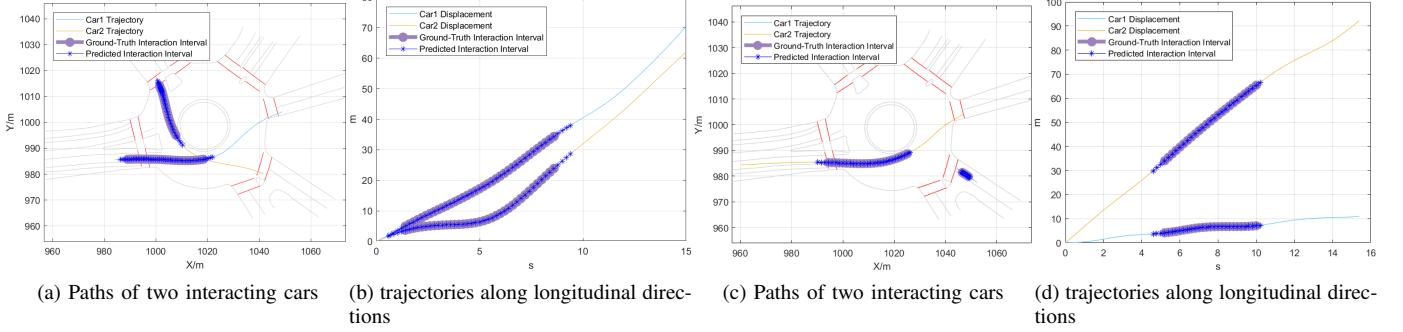


Fig. 16: Two examples of the extracted interaction pairs by implementing the learning method and network structure in [44].

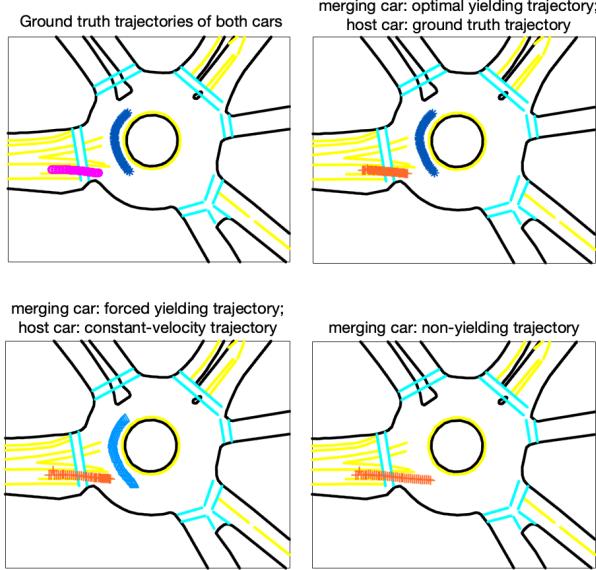


Fig. 17: Results of interpretable human behavior model based on the cumulative prospect theory (CPT) [45] using the proposed dataset.

- [9] N. Rhinehart, R. McAllister, and S. Levine, “Deep Imitative Models for Flexible Inference, Planning, and Control,” Oct. 2018.
- [10] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan, “Courteous autonomous cars,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 663–670.
- [11] C. Guo, K. Kidono, R. Terashima, and Y. Kojima, “Toward Human-like Behavior Generation in Urban Environment Based on Markov Decision Process With Hybrid Potential Maps,” in *2018 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2018, pp. 2209–2215.
- [12] M. Naumann, M. Lauer, and C. Stiller, “Generating Comfortable, Safe and Comprehensible Trajectories for Automated Vehicles in Mixed Traffic,” in *Proc. IEEE Intl. Conf. Intelligent Transportation Systems*, Hawaii, USA, Nov 2018, pp. 575–582.
- [13] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale, “The Next Generation Simulation Program,” *Institute of Transportation Engineers. ITE Journal; Washington*, vol. 74, no. 8, pp. 22–26, Aug. 2004.
- [14] L. Sun, W. Zhan, and M. Tomizuka, “Probabilistic Prediction of Interactive Driving Behavior via Hierarchical Inverse Reinforcement Learning,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 2111–2117.
- [15] W. Zhan, L. Sun, Y. Hu, J. Li, and M. Tomizuka, “Towards a Fatality-Aware Benchmark of Probabilistic Reaction Prediction in Highly Interactive Driving Scenarios,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 3274–3280.
- [16] F. Alché and A. de La Fortelle, “An LSTM network for highway trajectory prediction,” in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Oct. 2017, pp. 353–359.
- [17] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, “The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 2118–2125.
- [18] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, “Relational recurrent neural networks for vehicle trajectory prediction,” in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019.
- [19] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, “Argoverse: 3d Tracking and Forecasting With Rich Maps,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8748–8757.
- [20] Z. Kim, G. Gomes, R. Hranac, and A. Skabardonis, “A machine vision system for generating vehicle trajectories over extended freeway segments,” in *12th World Congress on Intelligent Transportation Systems*, 2005.
- [21] B. Coifman and L. Li, “A critical evaluation of the Next Generation Simulation (NGSIM) vehicle trajectory dataset,” *Transportation Research Part B: Methodological*, vol. 105, pp. 362–377, Nov. 2017.
- [22] D. Yang, L. Li, K. Redmill, and U. Özgüner, “Top-view Trajectories: A Pedestrian Dataset of Vehicle-Crowd Interaction from Controlled Experiments and Crowded Campus,” *arXiv:1902.00487 [cs]*, Feb. 2019, arXiv: 1902.00487.
- [23] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, “Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes,” in *ECCV 2016*. Springer International Publishing, 2016, pp. 549–565.
- [24] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko, “Toward Driving Scene Understanding: A Dataset for Learning Driver Behavior and Causal Reasoning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7699–7707.
- [25] V. L. Neale, T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman, “An overview of the 100-car naturalistic study and findings,” *National Highway Traffic Safety Administration, Paper*, vol. 5, p. 0400, 2005.
- [26] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.
- [27] D. Simon, *Optimal State Estimation: Kalman, H_{infinity}, and Nonlinear Approaches*. New York, NY, USA: Wiley-Interscience, 2006.
- [28] A. Clausse, S. Benslimane, and A. De La Fortelle, “Large-scale extraction of accurate vehicle trajectories for driving behavior learning,” *30th IEEE Intelligent Vehicles Symposium (IV)*, 2019.
- [29] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, “Mask R-CNN,” *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.
- [30] E. Bochinski, V. Eiselein, and T. Sikora, “High-speed tracking-by-detection without using image information,” in *International Workshop on Traffic and Street Surveillance for Safety and Security at IEEE AVSS 2017*, Lecce, Italy, Aug. 2017. [Online]. Available: <http://elvera.nue.tu-berlin.de/files/1517Bochinski2017.pdf>

- [31] A. Lukežič, T. Vojjíř, L. Čehovin Zajc, J. Matas, and M. Kristan, “Discriminative correlation filter tracker with channel and spatial reliability,” *International Journal of Computer Vision*, 2018.
- [32] D. C. Brown, “Close-range camera calibration,” *PHOTOGRAMMETRIC ENGINEERING*, vol. 37, no. 8, pp. 855–866, 1971.
- [33] P. Polack, F. Altché, B. d’Andréa-Novel, and A. de La Fortelle, “The kinematic bicycle model: A consistent model for planning feasible trajectories for autonomous vehicles?” in *2017 IEEE Intelligent Vehicles Symposium (IV)*, June 2017, pp. 812–818.
- [34] F. Poggenhans, J. Pauls, J. Janosovits, S. Orf, M. Naumann, F. Kuhnt, and M. Mayr, “Lanelet2: A high-definition map framework for the future of automated driving,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 1672–1679.
- [35] W. Zhan, L. Sun, D. Wang, Y. Jin, and M. Tomizuka, “Constructing a Highly Interactive Vehicle Motion Dataset,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.
- [36] H. Ma, J. Li, W. Zhan, and M. Tomizuka, “Wasserstein Generative Learning with Kinematic Constraints for Probabilistic Prediction of Interactive Driving Behavior,” in *2019 IEEE Intelligent Vehicles Symposium*, 2019.
- [37] Y. Hu, L. Sun, and M. Tomizuka, “Generic prediction architecture considering both rational and irrational driving behaviors,” in *2019 22st International Conference on Intelligent Transportation Systems (ITSC)*, to appear, 2019.
- [38] L. Sun, C. Peng, W. Zhan, and M. Tomizuka, “A Fast Integrated Planning and Control Framework for Autonomous Driving via Imitation Learning,” in *ASME 2018 Dynamic Systems and Control Conference*. American Society of Mechanical Engineers, Sep. 2018, pp. 1–11.
- [39] W. Zhan, J. Chen, C. Y. Chan, C. Liu, and M. Tomizuka, “Spatially partitioned environmental representation and planning architecture for on-road autonomous driving,” in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2017, pp. 632–639.
- [40] W. Zhan, C. Liu, C. Y. Chan, and M. Tomizuka, “A non-conservatively defensive strategy for urban autonomous driving,” in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 459–464.
- [41] T. Gu, J. Atwood, C. Dong, J. M. Dolan, and J.-W. Lee, “Tunable and stable real-time trajectory planning for urban autonomous driving,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 250–256.
- [42] J. Schulz, C. Hubmann, J. Lehner, and D. Burschka, “Interaction-Aware Probabilistic Behavior Prediction in Urban Environments,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 3999–4006.
- [43] D. Pelleg, A. W. Moore *et al.*, “X-means: Extending k-means with efficient estimation of the number of clusters.” in *ICML*, vol. 1, 2000, pp. 727–734.
- [44] T. Shu, Y. Peng, L. Fan, H. Lu, and S.-C. Zhu, “Perception of human interaction based on motion trajectories: From aerial videos to decontextualized animations,” *Topics in cognitive science*, vol. 10, no. 1, pp. 225–241, 2018.
- [45] L. Sun, W. Zhan, Y. Hu, and M. Tomizuka, “Interpretable modelling of driving behaviors in interactive driving scenarios based on cumulative prospect theory,” in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019.