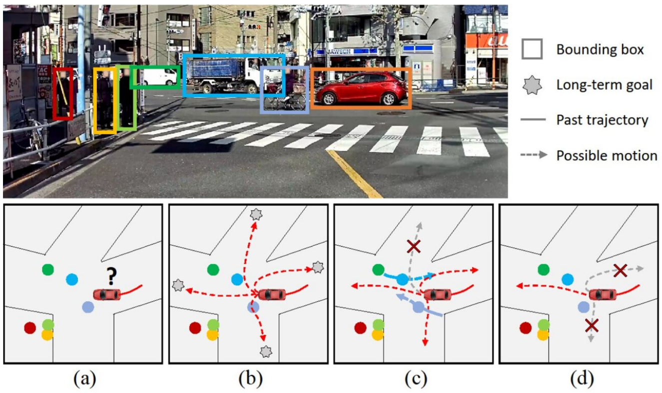
LOKI: An intention dataset to train models for pedestrian and vehicle trajectory prediction

9 September 2021, by Ingrid Fadelli



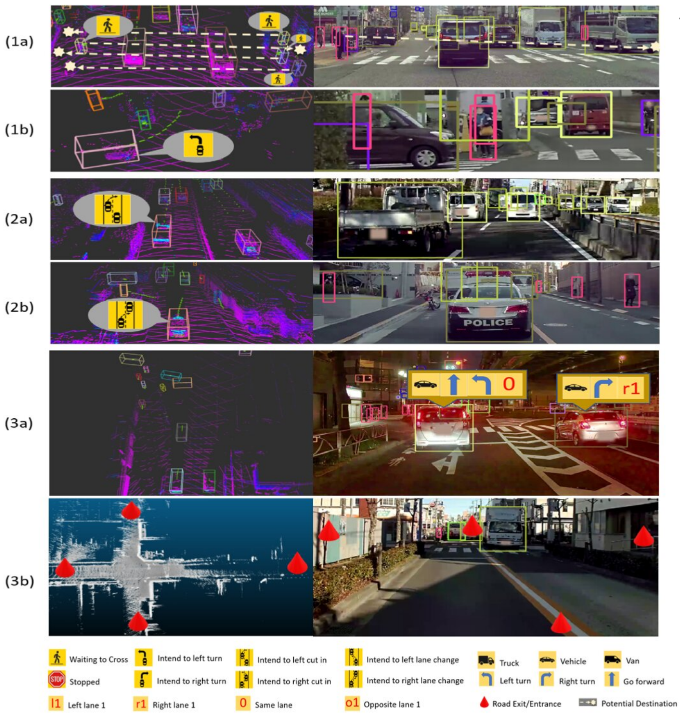
The researchers showed that reasoning about long-term goals and short-term intents plays a significant role in trajectory prediction. With a lack of comprehensive benchmarks for this purpose, they introduced a new dataset for intention and trajectory prediction. An example use case is illustrated in (a) where the team predict the trajectory of the target vehicle. In (b), long-term goals are estimated from agent's own motion. Interactions in (c) and environmental constraints such as road topology and lane restrictions in (d) influence the agent's short-term intent and thus future trajectories. Credit: Girase et al.

Human decision-making processes are inherently hierarchical. This means that they involve several levels of reasoning and different planning strategies that operate simultaneously to achieve both short-term and long-term goals.

Over the past decade or so, an increasing number of computer scientists have been trying to develop computational tools and techniques that could replicate human decision-making processes, allowing robots, autonomous vehicles or other devices to make decisions faster and more efficiently. This is particularly important for robotic systems performing actions that directly impact the safety of humans, such as self-driving cars.

Researchers at Honda Research Institute U.S., Honda R&D, and UC Berkeley have recently compiled LOKI, a dataset that could be used to train models that predict the trajectories of pedestrians and vehicles on the road. This dataset, presented in a paper pre-published on arXiv and set to be presented at the ICCV conference 2021, contains carefully labeled images of different agents (e.g., pedestrians, bicycles, cars, etc.) on the street, captured from the perspective of a driver.

"In our recent paper, we propose to explicitly reason about agents' long-term goals as well as their short-term intents for predicting future trajectories of traffic agents in driving scenes," Chiho Choi, one of the researchers who carried out the study, told TechXplore. "We define long-term goals to be a final position an agent wants to reach for a given prediction horizon, while intent refers to how an agent accomplishes their goal."

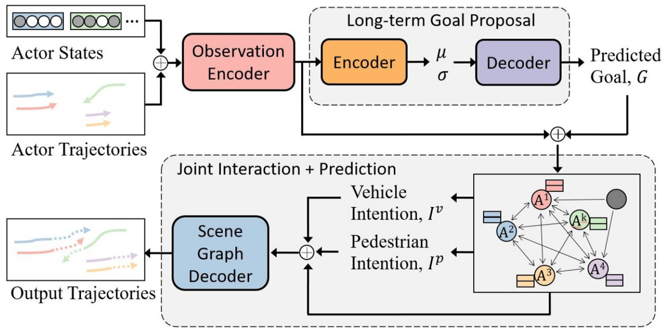


Visualization of three types of labels: (1a-1b) Intention labels for pedestrian; (2a-2b) Intention labels for vehicle; and (3a-3b) Environmental labels. The left part of each image is from laser scan and the right part is from camera. In (1a), the current status of pedestrian is ”Waiting to cross”, and the potential destination shows the intention of pedestrian. In (3a), the blue arrow indicates the possible action of the current lane where the vehicle is on, and the red words present the lane position related to the ego-vehicle. Credit: Girase et al.

Choi and his colleagues hypothesized that to predict the trajectories of traffic agents most efficiently, it is important for machine learning techniques to consider a complex hierarchy of short-term and long-term goals. Based on the agent motions predicted, the model can then plan the movements of a robot or vehicle most efficiently.

The researchers thus set out to develop an architecture that considers both short- and long-term goals as key components of frame-wise intention estimation. The results of these considerations then influence its trajectory prediction module.

"Consider a vehicle at an intersection where the vehicle wants to reach its ultimate goal of turning left to its final goal point," Choi explained. "When reasoning about the agent's motion intent to turn left, it is important to consider not only agent dynamics but also how intent is subject to change based on many factors including i) the agent's own will, ii) social interactions, iii) environmental constraints, iv) contextual cues."



Our model first encodes past observation history of each agent to propose a long-term goal distribution over potential final destinations for each agent independently. A goal, G is then sampled and passed into the Joint Interaction and Prediction module. A scene graph is constructed to allow agents to share trajectory information, intentions, and long-term goals. Black nodes denote road entrance/exit information which provides agents with map topology information. At each timesteps, current scene information is propagated through the graph. We then predict an intent (the action will the agent take in the near future) for each agent. Finally, the trajectory decoder is conditioned on predicted intentions, goals, past motion, and scene before forecasting the next position. This process is recurrently repeated for the horizon length. Credit: Girase et al.

The LOKI dataset contains hundreds of RGB images portrayed different agents in traffic. Each of these images has corresponding LiDAR point clouds with detailed, frame-wise labels for all traffic agents.

The dataset has three unique classes of labels. The first of these are intention labels, which specify 'how' an actor decides to reach a given goal via a series of actions. The second are environmental labels, providing information about the environment that impacts the intentions of agents (e.g., 'road exit' or 'road entrance' positions, 'traffic light," 'traffic sign," 'lane information," etc.). The third class includes contextual labels that could also affect the future behavior of agents, such as weather-related information, road conditions, gender and age of pedestrians, and so on.

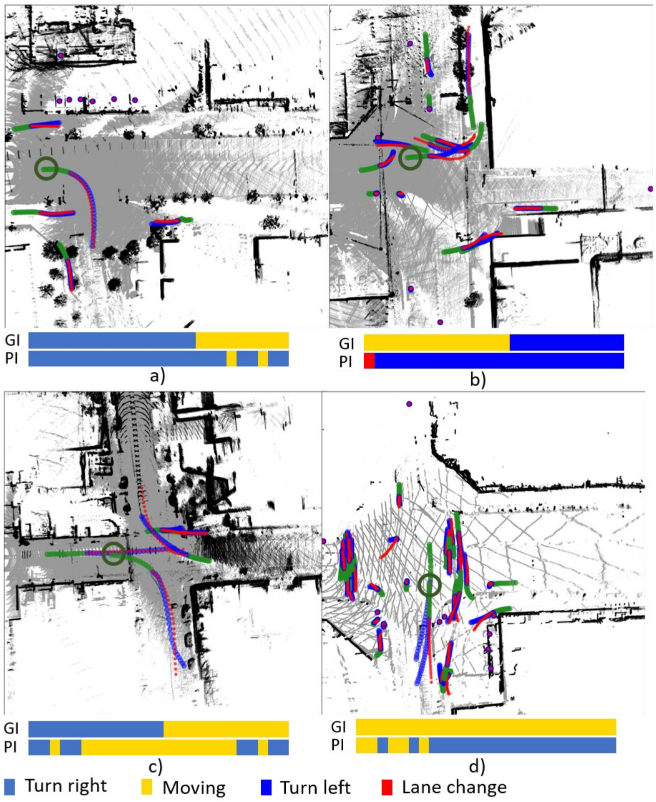
"We provide a comprehensive understanding of how intent changes over a long time horizon," Choi said. "In doing so, the LOKI dataset is the first that can be used as a benchmark for intention understanding for heterogeneous traffic agents (i.e., cars, trucks, bicycles, pedestrians, etc.)."

Category	Set	Instances	Description
Intention Labels (Vehicle)	Stopped	130743	The vehicle is stopped. This can happen in many scenarios such as stopping for a traffic light, waiting to make a turn at an intersection, yielding for a pedestrian, etc.
	Parked	127150	The vehicle is parked along the street or parking lot
	Lane change to the left	2120	The vehicle is merging into the next lane.
	Lane change to the right	2087	
	Cut in to the left	347	The vehicle is cutting into another lane.
	Cut in to the right	736	
	Turn left	15190	The vehicle is turning (ex: at an intersection or towards a highway ramp).
	Turn right	13171	
	Moving / Other	206243	The vehicle is driving forward or some other movement that is not captured in the other labels.
Intention Labels (Pedestrian)	Stopped	32538	The pedestrian is stopped along the street
	Moving	241889	The pedestrian is walking (ex: along the street)
	Waiting to cross	49576	The pedestrian is waiting to cross the intersection.
	Crossing the road	64870	The pedestrian is crossing the road.
	Potential Destination	67862	The potential location where the pedestrian may walk to.
Environmental Labels	Lane information	440338	The possible actions a vehicle can take based on the current lane it is in. (*e.g.* right turn, left Turn, go forward, u-turn, lane change not possible). Note that multiple choices can be selected depending on the situation. For example, a vehicle can be in a lane that goes forward or turns left. In our dataset, if a lane type is possible we select 1 and if it is not possible we select 0. Sometimes, if the vehicle is out of frame and lane information cannot be deduced, we label it as -1.
	Traffic light	42476	The current state of the traffic light (*e.g.* Red straight, Green round, Yellow round, etc.)
	Traffic sign	39066	The type of the traffic sign (*e.g.* Stop, Left turn only, Do not enter for all)
	Road Exit and Entrance	126889	The positions of the road entrances/exits for a given scene. There can be a variable number of road entrances/exits depending on map topology. Refer to figure 9 for more details.
Contextual Labels	Age	166874	The estimated age category (child, adult, senior) of the pedestrian.
	Gender	166874	The gender of the pedestrian (male/female)
	Weather	644	The weather condition of the scenario (Sunny/Dusk/Cloudy/Night).
	Road condition	644	The road surface condition (dry / wet).

Details of the LOKI dataset. We report the various types of labels, number of instances of each label, and descriptions for all label types. Credit: Girase et al.

In addition to compiling the LOKI dataset, Choi and his colleagues developed a model that explores how the factors considered by LOKI can affect the future behavior of agents. This model can predict the intentions and trajectories of different agents on the road with high levels of accuracy, specifically considering the impact of i) an agent's own will, ii) social interactions, iii) environmental constraints, and iv) contextual information on its short-term actions and decision-making process.

The researchers evaluated their model in a series of tests and found that it outperformed other state-of-the-art trajectory-prediction methods by up to 27%. In the future, the model could be used to enhance the safety and performance of autonomous vehicles. In addition, other research teams could use the LOKI dataset to train their own models for predicting the trajectories of pedestrians and vehicles on the road.



Visualization of top-1 trajectory prediction result (green: past observation, blue: ground truth, red: prediction) and frame-wise intention of a particular agent in dark green circle at the start of the observation time step (GI: Ground truth Intention, PI: Predicted Intention) is shown at the bottom of each scenario. Credit: Girase et al.

"We already started exploring other research directions aimed at jointly reasoning about intentions and trajectories while considering different internal/external factors such as agents' will, social interactions and environmental factors," Choi said. "Our immediate plan is to further explore the intention-based prediction space not only for trajectories but also for general human motions and behaviors. We are currently working on expanding the LOKI dataset in this direction and believe our highly flexible dataset will encourage the prediction community to further advance these domains."

More information: Harshayu Girase et al, LOKI: Long term and key intentions for trajectory prediction, arXiv:2108.08236 [cs.CV] arxiv.org/abs/2108.08236