**Title:One Thousand and One Hours: Self-driving Motion Prediction Dataset**
Source: arXiv preprint arXiv:2006.14480 (2020)
Link: https://arxiv.org/pdf/2006.14480
Year:2020
This paper outlines the features and key aspects of the dataset and explains in-depth the methods used to compile the dataset. The dataset consists of scenes, where each scene is 25 seconds long and captures the perception output of the self-driving system, which maps the precise positions and motions of nearby vehicles, cyclists, and pedestrians over time. The dataset is encoded in the form of n-dimensional compressed Zarr arrays, which enables fast random access throughout the dataset. The data is collected using a fleet of self-driving vehicles driving along a fixed route. The sensors utilized for the perception include seven cameras, three LiDARs, and five radars. Apart from the scenes, the dataset also comprises of Aerial maps and semantic maps. The aerial maps capture the area that runs along the route at a resolution of 6 cm per pixel^3. The semantic maps capture information about the roads and traffic elements. The entire dataset is structured in a manner apt for motion prediction tasks. Additionally, a python toolkit named L5kit is released, alongside for accessing the dataset. It provides useful features such as Multi-threaded data loading and sampling, Customisable scene visualization and rasterization, and Baseline motion prediction solution.

**Title: Rules of the Road : Predicting Driving Behaviour with a Convolutional Model of Semantic Interactions**

Source:IEEE Conference on Computer Vision and Pattern Recognition. 2019.
Link:
https://openaccess.thecvf.com/content_CVPR_2019/papers/Hong_Rules_of_the_Road_Predicting_Driving_Behavior_With_a_Convolutional_CVPR_2019_paper.pdf
Summary:
This paper discusses the problem of predicting future states of entities in complex, real-world driving scenarios. Many of the previous research papers have addressed mainly on how they predict small future time intervals. These papers also take input as raw sensory information(camera, lidar, or radar) which requires a heavy emphasis on extracting high-level representation of entities.
Moreover, the publicly available datasets used are very small and unrealistic. The dataset provided by this paper includes the following- 9,659 unique vehicles in 83,880 prediction scenarios (173 hours), in 88 physically-distinct locations and includes semantic map information. Using the given road data which includes connectivity of roads, lanes, junctions, stop and yield lines, etc., an RGB image that contains elements with unique colors is mapped to its geometric primitives. A neural network is used to map the low-level sensor information to 3D tracked entities. The output model is represented by a probability distribution over the entity state space at each time step, the actions the entity might take at a particular time, and efficiently predicting the full trajectories of the entity. The industry and linear baselines performed worse compared to these methods for predicting a large time interval( 5 seconds into the future), but better in smaller intervals. A useful insight from the paper is a dataset with a wide variety of real-world locations and unique 3D tracks are necessary for future predictions.

Also, the representation of multimodality is crucial in determining real-world planning for driving to avoid collisions and hence accidents.

**Title: “Uncertainty-aware Short-term Motion Prediction of Traffic Actors for Autonomous Driving”**
Source: The IEEE Winter Conference on Applications of Computer Vision. 2020.
Link:
https://openaccess.thecvf.com/content_WACV_2020/papers/Djuric_Uncertainty-aware_Short-term_Motion_Prediction_of_Traffic_Actors_for_Autonomous_Driving_WACV_2020_paper.pdf

This paper provides an approach to solve one of the integral aspects of the deployment of self-driving vehicles, which is the prediction of the actions taken by the various actors (vehicles, pedestrians, etc) on the scene. The approach used consists of two major phases, the rasterization of the maps and surroundings in the vehicles' vicinity followed by the training of a deep convolutional neural network for the prediction of the short-term trajectory for the actors while accounting for the inherent uncertainty in the environment. The rasterization process allows the complex 3D scene to be modeled in a more understandable manner, this is accomplished by utilizing a vector layer to represent the different major parts of the scene such as the roads or vehicles, and then assigning a distinct RGB colour to each of these layers. This rasterization process returns a 2D aerial view with distinct colours from the complex 3D scene which allows it to be more easily parsed by the neural network. The network architecture extracts features from the rasterized results and then passes it through two fully connected layers to obtain the trajectory prediction. This paper provides us with some important insights into how to handle driving data, namely with the rasterization procedure as well as the inherent uncertainty in the environment which must be taken into account while making predictions.

**Title: Multimodal Deep Generative Models for Trajectory Prediction: A Conditional Variational Autoencoder Approach**
Source:arXiv preprint arXiv:2008.03880 (2020)
Link: https://arxiv.org/pdf/2008.03880.pdf
Summary:
The goal of conditional generative modeling is to fit a model of the conditional probability distribution p(y | x), which may be used for downstream applications such as inference, or to generate new samples y given x. The encoder neural network, parameterized by i, takes the input x and produces a distribution pi(z | x) where
z is a latent variable that can be continuous or discrete. The decoder neural network, using the same parameter, uses the input x and samples from the encoder to produce to produce pi(y | x, z) conditional probability. p(y | x) is then obtained by marginalization of the latent variable z. RNNs are leveraged to process time series data without increasing problem size.

Future work on this paper includes developing ways to make the latent space more interpretable using better temporal logic, make the model fit against upstream sensor noise.