MOTION PLANNING IN AUTONOMOUS VEHICLES

One of the most critical challenges in autonomous driving systems is on road motion planning which includes finding a path with safest manoeuvre and determining the most feasible trajectory. The basic principle of autonomous vehicles is to use several sensors and cameras to extract informations(e.g. position, velocity, angular velocity) of the external agents(e.g. vehicles, pedestrians, objects, trees, e.t.c.) to take optimal decisions which were governed by predetermined algorithms. However, due to the complex traffic situations and uncertainty leading to unimaginable edge cases, it becomes hard to devise a general motion planning system for these vehicles. But, with the advent of Machine Learning techniques such as deep learning we are heading towards a future where these autonomous vehicles would be an easy deal. Designing an autonomous vehicle planning system needs an approach which would be able to replicate humans' decision making process while driving. When we are driving a car, our actions such as changing steering angles and applying brakes are not just based on instantaneous driving decisions. In fact, the current driving decision is based on what was traffic/road condition in a fast few seconds. Hence here comes the role of time series models such as LSTM and GRU. The augmentation of the agents around an autonomous driving vehicle is the most important factor for motion planning. Since these models will be used for on road vehicles in motion, the processing time becomes a very important factor. Among the 2 types of RNN, using GRU would be more favourable due it's lesser complexity than LSTM to reduce the computational time. It has been inferred from previous experiments that CNN perform better than LSTM in extracting the features of an image but is not powerful enough for time series prediction so an amalgam of these different concepts of LSTM and CNN can produce desired results. The convolutional LSTM has the ability to learn the time-serial features about the traffic environment around an autonomous vehicle. Just like a video can be split into frames, similarly capturing images after regular time intervals and using these frames to produce the optimal driving decision can be used to initiate the process. To deal with time series data LSTM is the most apt neural network but this does not take images/pixels as inputs like CNN but needs a numerical representation of the image.

There have been breakthrough results in feature extraction techniques through state-of-the-art architectures namely Resnet, Inception, Googlenet, Squeezenet, e.t.c. Tuning the most suitable CNN architecture by tweaking the last layer to produce an embedding of fixed length as a feature extractor for images makes the data preparation possible.

The output of the LSTM layer would represent the embedding of the frame for the next time step. Thus, the vehicle can get a prior intuition of how it would be after a time interval as a result of its previous motion which will in turn make it feasible to make further optimal decisions. To take these optimal decisions the vehicle should be trained to learn in various situations. It is nearly impossible to produce accurate labels for a wide range of environments in which the vehicle could be. We can resort to training our model in an online fashion where the learning procedure will commence with human intervention in determining the optimal motion plan for the various situations.

Combining CNN with LSTM can yield state-of-the-art results with the embeddings of images captured after regular time intervals which are extracted through CNN and fed into the LSTM layer sequentially.