

# Multi-Modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs

Nachiket Deo and Mohan M. Trivedi

**Abstract**—To safely and efficiently navigate through complex traffic scenarios, autonomous vehicles need to have the ability to predict the future motion of surrounding vehicles. Multiple interacting agents, the multi-modal nature of driver behavior, and the inherent uncertainty involved in the task make motion prediction of surrounding vehicles a challenging problem. In this paper, we present an LSTM model for interaction aware motion prediction of surrounding vehicles on freeways. Our model assigns confidence values to maneuvers being performed by vehicles and outputs a multi-modal distribution over future motion based on them. We compare our approach with the prior art for vehicle motion prediction on the publicly available NGSIM US-101 and I-80 datasets. Our results show an improvement in terms of RMS values of prediction error. We also present an ablative analysis of the components of our proposed model and analyze the predictions made by the model in complex traffic scenarios.

## I. INTRODUCTION

An autonomous vehicle deployed in complex traffic needs to balance two factors: the safety of humans in and around it, and efficient motion without stalling traffic. The vehicle needs to have the ability to take initiative, such as, deciding when to change lanes, cross unsignalized intersections, or overtake another vehicle. This requires the autonomous vehicle to have some ability to reason about the future motion of surrounding vehicles. This can be seen in existing tactical path planning algorithms [30]–[32], all of which depend upon reliable estimation of future trajectories of surrounding vehicles.

Many approaches use motion models for predicting vehicle trajectories [26]–[29]. However, motion models can be unreliable for longer prediction horizons, since vehicle trajectories tend to be highly non-linear due to the decisions made by the driver. This can be addressed by data-driven approaches to trajectory prediction [5], [10]–[12]. These approaches formulate trajectory prediction as a regression problem by minimizing the error between predicted and true trajectories in a training dataset. A pitfall for regression based approaches is the inherent multi-modality of driver behavior. A human driver can make one of many decisions under the same traffic circumstances. For example, a driver approaching their leading vehicle at a faster speed could either slow down, or change lane and accelerate to overtake. Regression based approaches have a tendency to output the average of these multiple possibilities, since the average prediction minimizes the regression error. However, the average

The authors are with the Laboratory for Intelligent and Safe Automobiles, University of California, San Diego, CA 92092, USA. [ndeo@ucsd.edu](mailto:ndeo@ucsd.edu), [mtrivedi@ucsd.edu](mailto:mtrivedi@ucsd.edu)

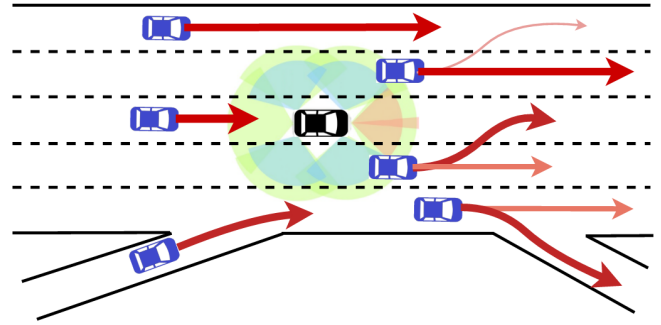


Fig. 1. An autonomous vehicle deployed in complex traffic (shown in the middle), needs to have the ability to predict the future motion of surrounding vehicles. Our proposed LSTM model allows for non-linear and multi-modal predictions of surrounding vehicle trajectories based on maneuver classes. It also assigns a probability to each mode and outputs uncertainty of prediction around each mode.

prediction may not be a good prediction. For instance, in the example scenario described above, the average prediction would be to stay in lane without deceleration. Thus, we need trajectory prediction models that address the multi-modal nature of predictions.

In this paper, we use *maneuvers* for multi-modal trajectory prediction, by learning a model that assigns probabilities for different maneuver classes, and outputs maneuver specific predictions for each maneuver class. Following the success of Long-Short Term Memory (LSTM) networks in modeling non-linear temporal dependencies in sequence learning and generation tasks [5], [23], [24], we propose an LSTM model for vehicle maneuver and trajectory prediction for the case of freeway traffic. It uses as input the track histories of the vehicle and its surrounding vehicles, and the lane structure of the freeway. It assigns confidence values to six maneuver classes and predicts a multi-modal distribution over future motion based on them. We train and evaluate our model using the NGSIM US-101 [2] and I-80 [3] datasets of real vehicle trajectories collected on Californian multi-lane freeways.

## II. RELATED RESEARCH

**Maneuver based models:** Classification of vehicle motion into maneuver classes has been extensively addressed in both advanced driver assistance systems as well as naturalistic drive studies [7]–[9], [19], [20]. A comprehensive survey of maneuver-based models can be found in [1], [6]. Of particular interest are works that use the recognized maneuvers to make better predictions of future trajectories [10], [11], [13]–[16]. These approaches usually involve a

maneuver recognition module for classifying maneuvers and maneuver specific trajectory prediction modules. Maneuver recognition modules are typically classifiers that use past positions and motion states of the vehicles and context cues as features. Heuristic based classifiers [13], Bayesian networks [14], hidden Markov models [10], [11], random forest classifiers [16] and recurrent neural networks have been used for maneuver recognition. Trajectory prediction modules output the future locations of the vehicle given its maneuver class. Polynomial fitting [13], maneuver specific motion models [14], Gaussian processes [11], [15], Gaussian mixture models [10] have been used for trajectory prediction. Many approaches [10], [16]–[18] also take into consideration the interaction between vehicles for assigning maneuver classes and predicting trajectories. Hand crafted cost functions based on relative configurations of vehicles are used in [10], [18] to make optimal maneuver assignments for all surrounding vehicles. However, these approaches can be limited by how well the cost function is designed. Other works [16], [17] implicitly learn vehicle interaction from trajectory data of real traffic. Here we adopt the second approach due to the availability of large datasets of real freeway traffic [2], [3].

**Recurrent networks for motion prediction:** Since motion prediction can be viewed as a sequence classification or sequence generation task, a number of LSTM based approaches have been proposed in recent times for maneuver classification and trajectory prediction. Khosroshahi *et al.* [19] and Phillips *et al.* [20] use LSTMs to classify vehicle maneuvers at intersections. Kim *et al.* [21] propose an LSTM that predicts the location of vehicles in an occupancy grid at intervals of 0.5s, 1s and 2s into the future. Contrary to this approach, our model outputs a continuous, multi-modal probability distribution of future locations of the vehicles up to a prediction horizon of 5s. Alahi *et al.* [5] propose *social LSTMs*, which jointly model and predict the motion of pedestrians in dense crowds through the use of a social pooling layer. However, vehicle motion on freeways has a lot more structure than pedestrians in crowds, which can be exploited to make better predictions. In particular, relative positions of vehicles can be succinctly described in terms of lane structure and direction of travel, and vehicle motion can be binned into maneuver classes, the knowledge of which can improve motion prediction. Lee *et al.* [22] use an RNN encoder-decoder based conditional variational auto-encoder (CVAE) for trajectory prediction. Sampling the CVAE allows for multi-modal predictions. Contrarily, our model outputs the multi-modal distribution itself. Finally, Kuefler *et al.* [4] use a gated recurrent unit (GRU) based policy using the behavior cloning and generative adversarial imitation learning paradigms to generate the acceleration and yaw-rate values of a bicycle model of vehicle motion. We compare our trajectory prediction results with those reported in [4].

### III. PROBLEM FORMULATION

We formulate motion prediction as estimating the probability distribution of the future positions of a vehicle condi-

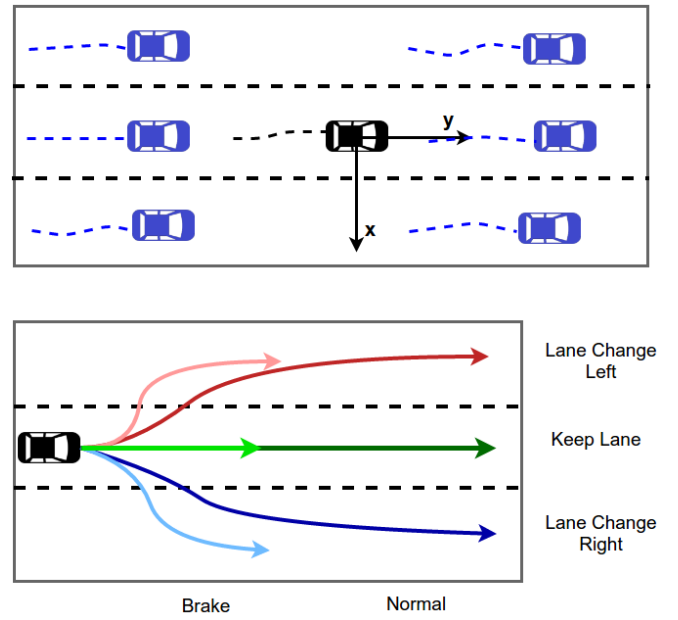


Fig. 2. **Top:** The co-ordinate system used for trajectory prediction. The vehicle being predicted is shown in black, neighboring vehicles considered are shown in blue. **Bottom:** Lateral and longitudinal maneuver classes

tioned on its track history and the track histories of vehicles around it, at each time instant  $t$ .

#### A. Frame of reference

We use a stationary frame of reference, with the origin fixed at the vehicle being predicted at time  $t$  as shown in Fig. 2. The  $y$ -axis points in the direction of motion of the freeway, and the  $x$ -axis is the direction perpendicular to it. This makes our model independent of how the vehicle tracks were obtained, and in particular, can be applied to the case of on-board sensors on an autonomous vehicle. This also makes the model independent of the curvature of the road, and can be applied anywhere on a freeway as long as an on-board lane estimation algorithm is available.

#### B. Inputs and outputs

The input to our model is the tensor of track histories

$$\mathbf{X} = [\mathbf{x}^{(t-t_h)}, \dots, \mathbf{x}^{(t-1)}, \mathbf{x}^{(t)}]$$

where,

$$\mathbf{x}^{(t)} = [x_0^{(t)}, y_0^{(t)}, x_1^{(t)}, y_1^{(t)}, \dots, x_6^{(t)}, y_6^{(t)}]$$

are the  $x$  and  $y$  co-ordinates at time  $t$  of the vehicle being predicted and six vehicles surrounding it as shown in Fig. 2. We choose these six vehicles since they seem to have the most effect on a vehicle's motion.

The output of the model is a probability distribution over

$$\mathbf{Y} = [\mathbf{y}^{(t+1)}, \dots, \mathbf{y}^{(t+t_f)}]$$

where,

$$\mathbf{y}^{(t)} = [x_0^{(t)}, y_0^{(t)}]$$

are the future co-ordinates of the vehicle being predicted

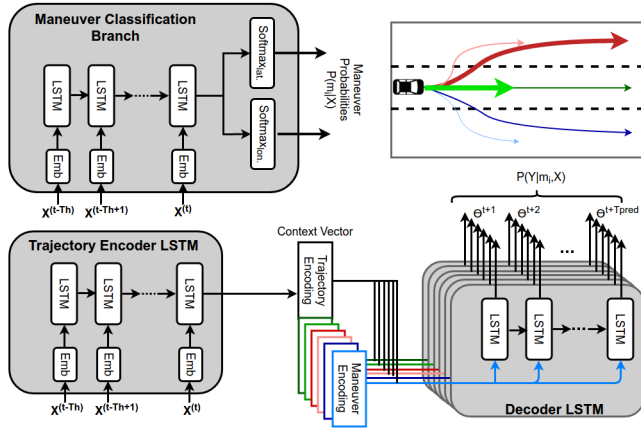


Fig. 3. **Proposed Model:** The trajectory encoder LSTM encodes the track histories and relative positions of the vehicle being predicted and its adjacent vehicles in a context vector. The context vector is appended with maneuver encodings of the lateral and longitudinal maneuver classes. The decoder LSTM generates maneuver specific future distributions of vehicle positions at each time step, and the maneuver classification branch assigns maneuver probabilities

### C. Probabilistic motion prediction

Our model estimates the conditional distribution  $P(\mathbf{Y}|\mathbf{X})$ . In order to have the model produce multi-modal distributions, we expand it in terms of maneuvers  $m_i$ , giving:

$$P(\mathbf{Y}|\mathbf{X}) = \sum_i P_{\Theta}(\mathbf{Y}|m_i, \mathbf{X})P(m_i|\mathbf{X}) \quad (1)$$

where,

$$\Theta = [\Theta^{(t+1)}, \dots, \Theta^{(t+t_f)}]$$

are the parameters of a bivariate Gaussian distribution at each time step in the future, corresponding to the means and variances of future locations.

### D. Maneuver classes

We consider three lateral and two longitudinal maneuver classes as shown in Fig. 2. The lateral maneuvers consist of left and right lane changes and a lane keeping maneuver. Since lane changes involve preparation and stabilization, we define a vehicle to be in a lane changing state for  $\pm 4s$  w.r.t. the actual cross-over. The longitudinal maneuvers are split into normal driving and braking. We define a vehicle to be performing a braking maneuver if its average speed over the prediction horizon is less than 0.8 times its speed at the time of prediction. We define our maneuvers in this manner since these maneuver classes are communicated by vehicles to each other through turn signals and brake lights, which will be included as a cue in future work.

## IV. MODEL

### A. LSTM encoder-decoder

Our proposed model is shown in Fig. 3. We use an encoder-decoder framework [23]. The trajectory encoder LSTM takes as input the frame by frame past locations of the predicted vehicle and its six adjacent vehicles for the past  $t_h$  frames. The hidden state vector of the encoder LSTM is updated at each time step based on the hidden state at the

previous time step and the input frame of vehicle locations at the current time step. The final state of the trajectory encoder LSTM can be expected to encode information about the track histories and relative positions of the 7 vehicles. This context vector is then used by the decoder LSTM as input. At each time step, for  $t_f$  frames into the future, the decoder LSTM state is updated based on the encoded context vector and LSTM state at the previous instant. The decoder outputs at each time step, a 5-D vector  $\Theta^{(t)}$  corresponding to the parameters of a bivariate Gaussian distribution, giving the distribution of the future locations of the predicted vehicle at that time instant, conditioned on the track histories.

### B. Maneuver dependent predictions

The encoder-decoder model described in the previous section outputs a uni-modal maneuver-independent trajectory distribution. In order to have the decoder generate a multi-modal trajectory distribution based on the six maneuver classes defined, we append the encoder context vector with a one-hot vector corresponding to the lateral maneuver class and a one-hot vector corresponding to the longitudinal maneuver class. The added maneuver context allows the decoder LSTM to generate maneuver specific probability distributions  $P_{\Theta}(\mathbf{Y}|m_i, \mathbf{X})$  as given in Eq. 1. To obtain the conditional probabilities  $P(m_i|\mathbf{X})$  for each maneuver class given track histories, we train the maneuver classification branch of the model shown in Fig. 3. The maneuver classification LSTM has the same inputs as the trajectory encoder LSTM. It has two output softmax layers for predicting the probabilities of lateral and longitudinal maneuver classes. Assuming the lateral and longitudinal maneuver classes to be conditionally independent given the track history, we obtain  $P(m_i|\mathbf{X})$  by taking the product of the corresponding lateral and longitudinal maneuver probabilities.

### C. Implementation details

We use LSTMs with 128 units for the encoder, decoder and maneuver classification branch. The input vectors  $X^{(t)}$  are embedded using a 64 unit fully-connected layer with leaky ReLU activation with  $\alpha=0.1$ , prior to being input to the LSTM layer. Although the trajectory encoder-decoder and maneuver classification models are used in tandem during test time, we train the models separately. The trajectory encoder-decoder is trained to minimize the negative log likelihood loss for the ground truth future locations of vehicles under the predicted trajectory distribution. The context vector is appended with the ground truth values of the maneuver classes for each training sample. The maneuver classification model is trained to minimize the the sum of cross-entropy losses of the predicted and ground truth lateral and longitudinal maneuver classes. Both models are trained using Adam [25] with a learning rate of 0.001. The models are implemented using Keras [33].

## V. EXPERIMENTAL EVALUATION

### A. Dataset

We use the publicly available NGSIM US-101 [2] and I-80 [3] datasets for our experiments. Each dataset consists





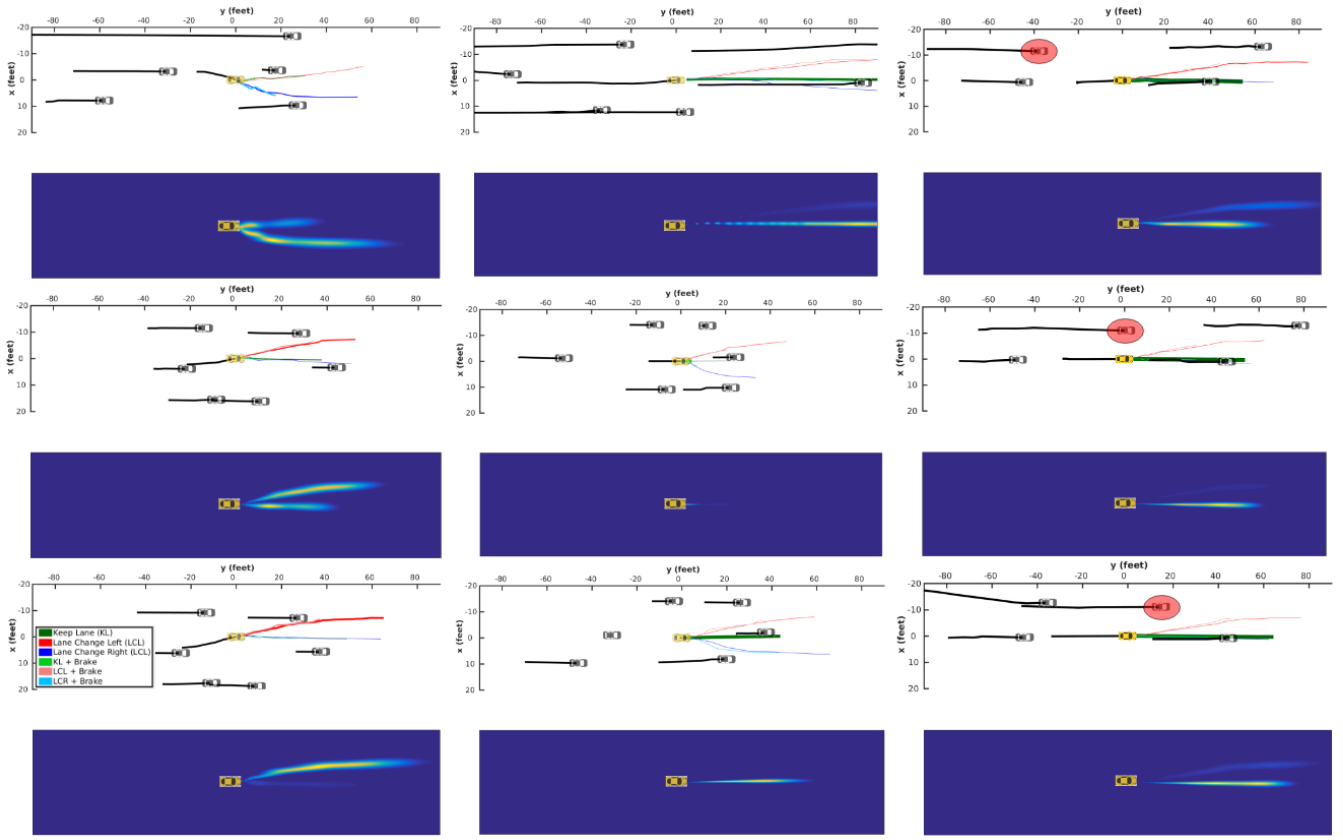


Fig. 6. **Analysis of predictions:** (a) Multimodal predictions, (b) Effect of leading vehicle, (c) Effect of adjacent vehicle

- *Surround LSTM with ground truth maneuvers (M-LSTM (GT))*: Finally, we also consider the M-LSTM with ground truth values of maneuver classes, to gauge the potential improvement in trajectory prediction with improved maneuver recognition

Fig 5 shows the RMS values of prediction error for the 4 system settings considered. We observe that the S-LSTM outperforms the Vanilla LSTM model, suggesting that motion of adjacent vehicles is a significant cue for predicting the future motion of vehicles. The M-LSTM leads to further improvement in prediction accuracy, suggesting the usefulness of maneuver classification prior to motion prediction. Both effects seem to become more pronounced for longer prediction intervals. Additionally, we note from the RMSE values of M-LSTM(GT) that considerable further improvement could have been achieved if maneuver classification was more accurate.

#### E. Qualitative analysis of predictions

In this section we qualitatively analyze the predictions made by our model to gain insights into its behavior in various traffic configurations. Figure 6 shows six different scenarios of traffic. Each figure shows a plot of track histories over the past 3 seconds and the mean predicted trajectories over the next 5 seconds for each maneuver class. The thickness of the plots of the predicted trajectories is proportional to the probabilities assigned to each maneuver

class. Additionally, each figure shows a heat map of the complete predicted distribution.

Fig. 5(a) illustrates the multi-modal nature of the prediction made by the model for vehicles about to change lanes. The predicted distribution has a mode corresponding to the respective lane change, as well as the *keep lane* maneuver. The model becomes more and more confident in the lane change further into the maneuver. We note that the model predicts the vehicle to merge into the target lane for the lane change maneuvers illustrating the ability of the LSTM to model the non-linear nature of vehicle motion.

Fig 5(b) shows the effect of the leading vehicle on the predictions made by the model. The first example shows an example of free flowing traffic, where the predicted vehicle and the leading vehicle are moving at approximately the same speed. In the second example, we note from the track histories that the leading vehicle is slowing down compared to the predicted vehicle. We see that the model predicts the vehicle to brake, although it's current motion suggests otherwise. Conversely, in the third example, we see that the vehicle being predicted is almost stationary, while the leading vehicle is beginning to move. The model predicts the vehicle to accelerate, as is expected in stop-and-go traffic.

Fig 5(c) shows the effect of vehicles in the adjacent lane on the model's predictions. The three examples show the same scenario separated by 0.5 sec, with the vehicle being predicted is in the rightmost lane. We note that in all three cases shown, the model assigns a high probability

to the vehicle keeping lane. However it also assigns a small probability for the vehicle to change to the left lane. We note that the probability to change to the left lane is affected by the circled vehicle shown in the plots. When the circled vehicle is far behind, the model assigns a high probability to the lane change. When the vehicle is right next to the vehicle being predicted, the lane change probability drops. When the vehicle passes and the lane opens up again, the probability for lane change increases again.

## VI. CONCLUSIONS

A novel LSTM based interaction aware model for vehicle motion prediction was presented in this paper, capable of making multi-modal trajectory predictions based on maneuver classes. The model was shown to achieve lower prediction error on two large datasets of real freeway vehicle trajectories, compared to two existing state of the art approaches from literature, demonstrating the viability of the approach. Additionally, an ablative analysis of the system showed the significance of modeling the motion of adjacent vehicles for predicting the future motion of a given vehicle, and detecting and exploiting common maneuvers of vehicles for future motion prediction.

## VII. ACKNOWLEDGEMENT

We gratefully acknowledge the support of our colleagues and sponsors and the comments of the anonymous reviewers.

## REFERENCES

- [1] S. Lefvre, D. Vasquez, and C. Laugier. "A survey on motion prediction and risk assessment for intelligent vehicles." *Robomech Journal* 1, no. 1 (2014): 1.
- [2] J. Colyar, and J. Halkias. US highway 101 dataset. *Federal Highway Administration (FHWA), Tech. Rep.* FHWA-HRT-07-030 (2007).
- [3] J. Colyar, and J. Halkias. US highway 80 dataset. *Federal Highway Administration (FHWA), Tech. Rep.* FHWA-HRT-07-030 (2007).
- [4] A. Kuefler, J. Morton, T. Wheeler, and M. Kochenderfer. Imitating driver behavior with generative adversarial networks. In *Intelligent Vehicles Symposium (IV), 2017 IEEE*, pp. 204-211. IEEE, 2017.
- [5] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 961-971. 2016.
- [6] S. Sivaraman, and M. M. Trivedi. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. *IEEE Transactions on Intelligent Transportation Systems* 14, no. 4 (2013): 1773-1795.
- [7] J. V. Dueholm, M. S. Kristoffersen, R. K. Satzoda, T. B. Moeslund, and M. M. Trivedi. Trajectories and maneuvers of surrounding vehicles with panoramic camera arrays. *IEEE Transactions on Intelligent Vehicles* 1, no. 2 (2016): 203-214.
- [8] S. Sivaraman, B. Morris, and M. M. Trivedi. Learning multi-lane trajectories using vehicle-based vision. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pp. 2070-2076. IEEE, 2011.
- [9] B. T. Morris, and M. M. Trivedi. Learning, modeling, and classification of vehicle track patterns from live video. *IEEE Transactions on Intelligent Transportation Systems* 9, no. 3 (2008): 425-437.
- [10] N. Deo, A. Rangesh, and M. M. Trivedi. How would surround vehicles move? A Unified Framework for Maneuver Classification and Motion Prediction. *arXiv preprint arXiv:1801.06523* (2018).
- [11] C. Laugier, I. E. Paromtchik, M. Perrollaz, M. Yong, J-D Yoder, C. Tay, K. Mekhnacha, and A. Ngre. Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety. *IEEE Intelligent Transportation Systems Magazine* 3, no. 4 (2011): 4-19.
- [12] J. Wiest, M. Hffken, U. Kreel, and K. Dietmayer. Probabilistic trajectory prediction with gaussian mixture models. In *Intelligent Vehicles Symposium (IV)*, pp. 141-146. IEEE, 2012.
- [13] A. Houenou, P. Bonnifait, V. Cherfaoui, and W. Yao. Vehicle trajectory prediction based on motion model and maneuver recognition. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pp. 4363-4369. IEEE, 2013.
- [14] M. Schreier, V. Willert, and J. Adamy. Bayesian, maneuver-based, long-term trajectory prediction and criticality assessment for driver assistance systems. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pp. 334-341. IEEE, 2014.
- [15] Q. Tran, and J. Firl. Online maneuver recognition and multimodal trajectory prediction for intersection assistance using non-parametric regression. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pp. 918-923. IEEE, 2014.
- [16] J. Schlechtriemen, F. Wirthmueller, A. Wedel, G. Breuel, and K-D Kuhnert. "When will it change the lane? A probabilistic regression approach for rarely occurring events." In *Intelligent Vehicles Symposium (IV), 2015 IEEE*, pp. 1373-1379. IEEE, 2015.
- [17] E. Kfer, C. Hermes, C. Whler, H. Ritter, and F. Kummert. Recognition of situation classes at road intersections. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pp. 3960-3965. IEEE, 2010.
- [18] M. Bahram, C. Hubmann, A. Lawitzky, M. Aeberhard, and D. Wollherr. A combined model-and learning-based framework for interaction-aware maneuver prediction. *IEEE Transactions on Intelligent Transportation Systems* 17, no. 6 (2016): 1538-1550.
- [19] A. Khosroshahi, E. Ohn-Bar, and M. M. Trivedi. Surround vehicles trajectory analysis with recurrent neural networks. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pp. 2267-2272. IEEE, 2016.
- [20] D. J. Phillips, T. A. Wheeler, and M. J. Kochenderfer. Generalizable intention prediction of human drivers at intersections. In *Intelligent Vehicles Symposium (IV), 2017 IEEE*, pp. 1665-1670. IEEE, 2017.
- [21] B. Kim, C. M. Kang, S. H. Lee, H. Chae, J. Kim, C. C. Chung, and J. W. Choi. Probabilistic Vehicle Trajectory Prediction over Occupancy Grid Map via Recurrent Neural Network. *arXiv preprint arXiv:1704.07049* (2017).
- [22] N. Lee, W. Choi, P. Vernaza, C. B. Choy, PHS Torr, and M. Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 336-345. 2017.
- [23] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).
- [24] A. Graves. Generating sequences with recurrent neural networks. *arXiv preprint: 1308.0850* (2013).
- [25] D. P. Kingma, and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [26] A. Polychronopoulos, M. Tsogas, A. J. Amditis, and L. Andreone. Sensor fusion for predicting vehicles' path for collision avoidance systems. *IEEE Transactions on Intelligent Transportation Systems* 8, no. 3 (2007): 549-562.
- [27] A. Barth, and U. Franke. Where will the oncoming vehicle be the next second?. In *Intelligent Vehicles Symposium, 2008 IEEE*, pp. 1068-1073. IEEE, 2008.
- [28] R. Schubert, E. Richter, and G. Wanielik. Comparison and evaluation of advanced motion models for vehicle tracking. In *Information Fusion, 2008 11th International Conference on*, pp. 1-6. IEEE, 2008.
- [29] R. Toledo-Moreo, and M. A. Zamora-Izquierdo. IMM-based lane-change prediction in highways with low-cost GPS/INS. *IEEE Transactions on Intelligent Transportation Systems* 10, no. 1 (2009): 180-185.
- [30] S. Ulbrich, and M. Maurer. Towards tactical lane change behavior planning for automated vehicles. In *International Conference on Intelligent Transportation Systems (ITSC), 2015*, pp. 989-995. IEEE, 2015.
- [31] J. Nilsson, J. Silvlin, M. Brannstrom, E. Coelingh, and J. Fredriksson. If, when, and how to perform lane change maneuvers on highways. *IEEE Intelligent Transportation Systems Magazine* 8, no. 4 (2016): 68-78.
- [32] S. Sivaraman, and M. M. Trivedi. Dynamic probabilistic drivability maps for lane change and merge driver assistance. *IEEE Transactions on Intelligent Transportation Systems* 15, no. 5 (2014): 2063-2073.
- [33] F. Chollet. Keras, <https://github.com/fchollet/keras>, 2015.