A Review of Deep Learning-Based Trajectory Prediction for Autonomous Vehicles

Xinran Li

Chang'an University, Xi'an, China.

2022902980@chd.edu.cn

Abstract. Trajectory prediction is one of the key technologies of the autonomous driving system. The accurate and efficient prediction can assist the autonomous driving system in making correct decisions and ensuring driving safety. This paper summarizes the research on trajectory prediction for autonomous vehicles based on deep learning. First, trajectory prediction models' input representation and output types are analyzed. Then, deep learning-based trajectory prediction methods are reviewed with their advantages and limitations. Finally, future research directions are proposed to address the shortcomings of existing methods in terms of multi-agent interaction modeling and generalization capability.

Keywords: Autonomous Driving, Trajectory Prediction, Deep Learning, Review

1. Introduction

With the increasing degree of informatization and intelligence, autonomous driving technology has been developing rapidly. The automotive industry has witnessed the emergence of autonomous driving systems equipped with intelligent driving features tailored for specific scenarios [1]. However, to realize safe driving, autonomous vehicles must be able to predict the future state of the surrounding environment in real time like a real-person driver. Therefore, trajectory prediction for traffic participants becomes one of the key technologies for autonomous driving systems. Trajectory prediction predicts the future state of a vehicle based on its surroundings information and historical trajectories in the current environment, and transmits the accurate and effective trajectory prediction results to the decision planning module to help the vehicle perform the correct behaviors, which greatly improves the traffic safety .

However, accurately predicting the future movement trajectories of traffic participants has become the key to improving the safety of autonomous driving due to the behavioral uncertainty of traffic participants and the complex interaction environment. Literature [3] reviewed the behavior prediction methods in intersection scenarios, and then analyzed the safety of three types of participants, vehicles, drivers, and pedestrians at intersections, but the prediction scenarios are very limited. Mozaffari et al. [4] in 2019 categorized and reviewed deep learning-based vehicle trajectory prediction methods based on three categories of criteria, input representation, output type, and prediction method, but did not address newer trajectory prediction ways. Literature [5] provides a more comprehensive review of autonomous driving trajectory prediction methods proposed in the last two decades, but does not mention the big model-based prediction methods that have emerged in recent years.

Therefore, this paper summarizes the research results for the task of vehicle trajectory prediction, with the main contents as follows.

1) Input representations and output types of trajectory prediction models are reviewed and analyzed.

2) Trajectory prediction methods based on Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Integration of CNN and RNN(CNN&RNN), Graph Neural Network (GNN), Attention Mechanism (AM), and Large Language Models (LLM) are summarized and analyzed in detail.

3) The future outlook of trajectory prediction for autonomous vehicles is presented.

2. Problem Definition

The trajectory prediction problem can be represented as predicting the future state of a vehicle based on its historical trajectory in the current environment. The historical trajectory coordinates of a traffic participant are denoted as follows.

$$\mathrm{X} = \left\{(x_0^t, y_0^t), (x_1^t, y_1^t), \dots, \left(x_j^t, y_j^t\right), \dots, (x_n^t, y_n^t)\right\} \quad (1)$$

X is the input of the prediction model, which includes the historical trajectory information of the traffic participant. $t \in \{0,1,2,\dots,t_h\}$ denotes the historical time step and t_h denotes the total historical time step. $\left(x_j^t, y_j^t\right)$ represents the value of horizontal and vertical coordinates of the vehicle $j \in \{1,2,\dots,n\}$ at the historical time step of *t*. *n* refers to all traffic vehicles detected by the ego vehicle. The output of the prediction model can be expressed as follows.

$$\mathrm{Y} = \left\{(x_0^{t_f}, y_0^{t_f}), (x_1^{t_f}, y_1^{t_f}), \dots, (x_j^{t_f}, y_j^{t_f}), \dots, (x_n^{t_f}, y_n^{t_f})\right\} \quad (2)$$

Y is the output of the prediction model, which is the predicted trajectory information of the target vehicle in the future time period T. $t_f \in \{t_h + 1, t_h + 2, \dots, t_h + \mathrm{T}\}$ denotes the prediction time step and T denotes the total predicted time step. $(x_j^{t_f}, y_j^{t_f})$ stands for the horizontal and vertical coordinate values of vehicle $j \in \{1,2,\dots,n\}$ at the prediction time step t_f.

3. Input Representations and Output Types for Trajectory Prediction Models

3.1 Input Representation for Trajectory Prediction Models

The input representation of the trajectory prediction model includes agent trajectory information and high-definition map information.

3.1.1 Agent Trajectory Information

Agent trajectory information contains several physical elements, such as position information, velocity, acceleration, and heading angle. Traditional model-driven trajectory prediction based methods require position information, velocity, acceleration, etc., as inputs to build a vehicle motion or dynamics model to realize the prediction of the motion trajectory of a single vehicle in a short period of time. Literature [6] used Kalman linear filters to integrate the model and used position, velocity, and acceleration as inputs for trajectory prediction, which can predict the future position of the vehicle. Literature [7] proposed a recursive neural network-based prediction method to predict the behavior of a vehicle using a trajectory history sequence of its horizontal and vertical position coordinates, velocity, acceleration, and heading angle. In general, the more information input to the trajectory prediction model, the more reliable and accurate the prediction results.

3.1.2 High Definition Map Information

High Definition Map (HD Map) is fused with historical trajectories to effectively make accurate predictions about the future driving status of vehicles. As a key part of the autonomous driving system, the core function of HD Map is to realize centimeter-level spatial positioning and lane-level path planning functions [8], which integrates multi-dimensional static environmental feature parameters, mainly including road geometry topology, lane lines, spatial coordinates of traffic markers, and other key elements. HD Map is usually classified into rasterized maps and vectorized maps.

Rasterized Map representation transforms an HD Map into a Bird's Eye View (BEV), which is then modeled as a semantic segmentation task. Literature [9] introduces a deep learning based approach that considers the current state of the environment and generates a rasterized image near each participant. The rasterized map is then used as an input to a deep convolutional model to predict the future trajectories of traffic participants. The rasterized representation has the advantages of being highly interpretive and fusing more map information, but it also suffers from the problems of its restricted perceptual field, missing information in the rasterization process, and high computational cost.

To address the limitations of rasterized maps, Literature [10] proposes a graph-centric motion prediction model and constructs complex interdependencies of vehicles, topologies, and lanes using vectorization methods. The advantage of vectorized maps is that no matter how the scaling operation is performed on them, it will not change their clarity. Meanwhile, vectorized maps can also reduce computational resources for greater efficiency.

3.2 Output Types for Trajectory Prediction Models

The output types of trajectory prediction models are mainly categorized into three types, driving intention, unimodal trajectory, and multimodal trajectory.

3.2.1 Driving Intention

Driving intention can be used either as a part of the prediction result or as an intermediate result. Driving intention can be categorized into changing lanes to the left, going straight line, changing lanes to the right, and so on.

Liu et al. [11] used a hybrid algorithm of Hidden Markov Models and Support Vector Machines to achieve intention prediction by analyzing the multi-source parameter differences of human-vehicle-road systems in driving scenarios. Literature [12] designed a two-layer Hidden Markov framework for real-time prediction of operational intentions and behaviors under complex driving situations. Ji et al. [13] introduced a long and short-term memory network that not only predicts driving intentions but also generates future trajectories. Compared with traditional prediction methods, this method, based on long and short-term memory networks, performs better in long-term trajectory prediction, whose prediction mechanism, combined with interaction information, significantly improves the accuracy of prediction. However, the existing driving intention prediction methods do not give the exact trajectory information, which leads to possible error in the prediction results, thus causing driving safety accidents.

3.2.2 Unimodal Trajectories

Unimodal trajectory prediction methods output single trajectory with the greatest possibility.

Literature [14] uses recurrent neural networks to predict short-term and long-term trajectories using a single algorithm and through three parameters of a cubic polynomial. In Literature [15], the decoder outputs a sequence of trajectories based on the parameters of a binary Gaussian distribution, and an optimization algorithm is used to make the prediction results converge to typical motion patterns. The unimodal trajectories are generally presented as time series, and the results are characterized by simplicity, intuition, ease of understanding, and low error compared to outputting driving intentions. However, such methods all assume that the future motion state of the car is deterministic and ignore its uncertainty.

3.2.3 Multimodal Trajectories

In contrast, multimodal trajectory prediction methods take into account the uncertainty of traffic behavior and can predict the set of possible trajectories under different driving behavior assumptions.

Literature [16] proposes interactive multi-modeling, which combines multiple parallel models into a single weighted estimate that outputs multimodal trajectories for different driving modes. Literature [17] proposed an improved interactive multi-model Kalman filtering algorithm that incorporates three prediction mechanisms, namely, dynamics model, behavioral features, and interaction relationships, to generate multiple sets of possible motion trajectory predictions in a short period of time. The multimodal trajectory prediction takes into account the uncertainty of vehicles in complex and changing environments, which ensures the prediction accuracy.

4. Deep Learning-Based Trajectory Prediction Methods

In recent years, deep learning-based motion trajectory prediction techniques have received a lot of attention, because they can be applied to more complex scenarios by taking into account not only kinematic and roadway factors but also interaction factors.

Deep learning based trajectory prediction methods are mainly categorized into RNN, CNN, CNN&RNN, GNN, Attention Mechanism, and LLM.

4.1 RNN

RNN has strong temporal feature extraction capabilities. This network architecture saves historical trajectory information through memory units and judges the output of the system based on current inputs and hidden states. However, traditional RNN suffers from the problem of long-term dependency. The network is prone to gradient vanishing or gradient explosion as the time-series span increases. For this reason, researchers have proposed to utilize Long Short-Term Memory (LSTM) networks to solve the above problems.

Altché et al. [18] repeated inputs of current target vehicle history states to a single-layer LSTM neural network to predict the future longitudinal and lateral trajectories of a single target vehicle on a highway. The single RNN model is only suitable for unimodal trajectory prediction. With the continuous development of neural network technology, multiple RNN architectures have been widely used for more complex trajectory prediction tasks. For example, in the dual LSTM framework proposed in Literature [19], the first LSTM network generates intermediate semantic features by analyzing historical trajectory data to predict driving intentions. The second LSTM performs multimodal trajectory prediction based on this high-level semantic information. This hierarchical processing mechanism effectively solves the intention-trajectory association modeling problem in long-distance prediction.

4.2 CNN

The structure of CNN adopts a “sequence-sequence” relationship, which takes historical trajectories as input, superimposes a convolutional layer, and then utilizes a fully connected layer to output future trajectories. Therefore, some scholars have proved that the predicted vehicle trajectories using CNN have stronger spatial-temporal continuity compared to RNN [20].

CNN is generally used to extract image features for trajectory prediction using Bird Eye View (BEV). Literature [21] first used a six-layer CNN architecture to extract deep features from the binary environment representation of the BEV to realize lane change intention prediction. Then the prediction results were input into the speed control module to complete the trajectory prediction. Casas et al. [22] designed a dual-channel CNN architecture to process the lidar point cloud and raster map data of the BEV separately, and then perform obstacle detection, behavioral intention prediction, and motion trajectory prediction.

4.3 Integration of CNN and RNN

RNN is able to extract temporal features, which is very suitable for processing time series information. While CNN is able to extract spatial features, which is suitable for processing spatial structure information. Based on the complementary advantages of these two networks, scholars have proposed a hybrid CNN-RNN architecture to collaboratively mine spatio-temporal correlation features in vehicle motion states.

Deo et al. [23] replaced the traditional fully connected layer with a convolutional pooling layer to enhance spatial feature extraction, whose core structure contains three key components. The LSTM encoder is responsible for learning the temporal dependence of vehicle motion, the convolutional pooling layer captures the local spatial interaction patterns, and the LSTM decoder synthesizes spatio-temporal features to generate future trajectories. Literature [24] extracted spatial features of the environment from a time-series BEV by CNN based on a hybrid CNN-RNN model, then modeled

temporal features using RNN, and decoded to generate predicted trajectories through a recursive structure. Literature [25] proposed a method for trajectory prediction of multiple intelligent agents in a dynamic scene, which consists of three main components, a GRU-based global spatio-temporal interaction feature extraction network, a CNN-based network for decoding the environment of a dynamic scene, and an LSTM-based network for predicting the trajectories of intelligent agents.

4.4 GNN

In complex driving scenarios, each scenario can be viewed as a graph structure consisting of the traffic participants in the scenario. The nodes in the graph represent traffic participants, and there are interconnected edges between each node to represent the interdependence between traffic participants. GNN, as a deep learning method running on graph-structured data, is able to extract non-Euclidean spatial data features with higher interpretability, so many scholars have tried to use GNN to predict the future trajectories of vehicles.

Graph Convolutional Network (GCN) extends the traditional CNN to graph data convolutional processing to extract graph structural features by integrating feature information from the center node and neighboring nodes. Li et al. [26] proposed a Graph-based Interaction-aware Trajectory Prediction (GRIP) based on GCN and LSTM. This method uses graphs to represent the interactions of close objects, applies multiple graphical convolution modules to extract useful temporal features as input data, and then realizes the prediction of the trajectories of the surrounding vehicle movements using LSTM networks. Chandra et al. [27] performed spatial coordinate prediction and interaction modeling through a two-layer GCN-LSTM network, respectively. The scheme not only predicts future trajectories but also can predict the behavioral intentions of traffic participants, which significantly improves the prediction accuracy in complex urban scenarios.

GNN can also be used to represent vectorized maps. Ziegler et al. [28] pioneered the application of vectorized maps for trajectory prediction, which contains the location of lanes, topology between lanes, and traffic regulations. It predicts the future trajectory of a vehicle through GNN. Liang et al. [29] constructed a lane map using raw map data to preserve the structure of the map and utilized GCN to capture traffic participant interaction with the vector map to predict vehicle trajectories.

4.5 Attention Mechanism

Attention Mechanism is able to quickly and accurately capture high-value information in complex data and environments by mimicking the human mind, which is widely used in visual processing, natural language processing, speech recognition, and other tasks [25, 30].

Vaswani et al. [31] innovatively proposed the Transformer architecture, which is the first sequence transformation model that relies exclusively on Attention Mechanism. Instead of the traditional RNN structure, the model adopts a multi-head self-attention mechanism, which not only significantly improves the sequence processing speed but also computes all the data in the input sequence in parallel. Thanks to the excellent performance of Transformer in the field of natural language processing, the architecture has been gradually introduced into the field of trajectory prediction, which provides a new technical path for time series data processing. For example, Zhao et al. [32] proposed a spatial-channel Transformer network, where they used the Transformer model instead of the traditional RNN to process long-time-domain data. Through spatial embedding techniques, the spatial-temporal features of each traffic participant are captured with the Transformer. Liu et al. [33] proposed an end-to-end multimodal motion prediction framework, mmTransformer (Multimodal Transformer), which utilizes the fusion of environmental and trajectory information to predict future trajectories. Transformer-based multimodal trajectory prediction has higher accuracy compared to multimodal prediction models such as CNN, RNN, and others.

4.6 LLM

In recent years, breakthroughs in LLM have provided a new technological paradigm for trajectory prediction. With self-supervised pre-training and the Transformer architecture, LLM performs well

on time series analysis tasks and demonstrates the ability to model long sequence dependencies, contextual semantic understanding, and zero-sample inference in natural language processing. This technological paradigm provides a new perspective for trajectory prediction research. By mapping spatio-temporal trajectory data into structured linguistic sequences, LLM is expected to break through the traditional model's reliance on explicit physical rules and realize intentional reasoning based on semantic understanding and interactive modeling of multimodal environments.

Liu et al. [34] proposed a Spatial-Temporal Large Language Model (ST-LLM) for traffic prediction, which encodes the location information through a spatio-temporal embedding module and integrates the spatio-temporal features using fusion convolution to efficiently capture the global spatio-temporal dependencies. In order to improve the accuracy and interpretability of long time-domain prediction in trajectory prediction, Peng et al. [35] proposed for the first time an interpretable lane change prediction model for predicting vehicle lane change behavior, which reformulates the lane change prediction task as a language modeling problem, characterizes the driving scenarios through natural language cues, and employs supervised fine-tuning to tailor the LLM. Lan et al. [36] proposed a pre-trained LLM, which converts traffic participant features into an understandable input form for the language model, combines it with the Mamba module for probabilistic modeling, and ultimately outputs scenario-adaptive predictions through a multimodal Laplacian decoder.

4.7 Summary

Compared with traditional trajectory prediction methods, deep learning-based trajectory prediction methods can realize long-time domain prediction of vehicle motion to a greater extent, and can better adapt to complex driving scenarios and handle dynamic interaction information. However, deep learning-based algorithms require a large amount of trajectory data during training. The prediction performance depends on the merit of model training, and problems such as data scarcity or distribution bias, can lead to a decrease in the generalization ability of the model.

5. Outlook of Trajectory Prediction Research

Trajectory prediction technology has made significant breakthroughs in recent years, but the existing deep learning-based trajectory prediction methods have certain limitations, which restrict the further improvement of prediction accuracy. Therefore, based on the current research, this paper puts forward the following outlook for the future development of autonomous driving trajectory prediction technology.

5.1 Multi-agent Prediction

Tightly combining the interaction information of traffic participants and map deter for multi-agent prediction, this approach is more relevant to real-world scenarios. Therefore, future trajectory prediction research can focus more on joint trajectory prediction among multi-agents to achieve the modeling of multi-objective interactions in traffic scenarios. For example, by introducing a global feature interaction framework based on Transformer, the future interactive motion of multiple traffic participants can be predicted.

5.2 Prediction Model Generalizability

To further realize full-scenario prediction and improve prediction accuracy in complex scenarios, it is necessary to enhance the generalization ability of prediction models. For example, a priori knowledge can be injected in conjunction with a large model to enhance the generalization ability of the prediction model. In addition, most of the existing trajectory prediction methods only extract the information from the visual sensors as input data, and the extraction and utilization of the information in other sensors is insufficient, so the extraction and utilization of multi-sensor information is also key to the generalizability of the prediction model.

6. Conclusions

Accurate and efficient trajectory prediction is the basis for autonomous driving systems to realize safe and rational decision-making. This paper reviews the existing research results from three aspects, input representation, output type, and prediction methods. First, it introduces the input representation of the trajectory prediction model from trajectory information and high-precision map information. Additionally, the output types of trajectory prediction are reviewed in terms of driving intention, unimodal trajectory, and multimodal trajectory. Then, this paper provides a detailed overview of deep learning-based trajectory prediction methods. Finally, this paper points out that the future development of self-driving trajectory prediction technology lies in multi-agent prediction and prediction model generalizability. In conclusion, the continuous innovation and development of trajectory prediction will lay a solid foundation for the safety and intelligence of autonomous driving and promote the comprehensive development of the intelligent transportation system.

References

[1] F.Y. Wang, MetaVehicles in the Metaverse: Moving to a New Phase for Intelligent Vehicles and Smart Mobility, IEEE Transactions on Intelligent Vehicles, 7(1) (2022) 1-5.

[2] Liu J, Mao X, Fang Y, et al.,,A survey on deep-learning approaches for vehicle trajectory prediction in autonomous driving,//2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2021: 978-985.

[3] M.S. Shirazi, B.T. Morris, Looking at Intersections: A Survey of Intersection Monitoring, Behavior and Safety Analysis of Recent Studies, IEEE Transactions on Intelligent Transportation Systems, 18(1) (2017) 4-24.

[4] S. Mozaffari, O.Y. Al-Jarrah, M. Dianati, P. Jennings, A. Mouzakitis, Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review, IEEE Transactions on Intelligent Transportation Systems, 23(1) (2022) 33-47.

[5] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, H. Chen, A Survey on Trajectory-Prediction Methods for Autonomous Driving, IEEE Transactions on Intelligent Vehicles, 7(3) (2022) 652-674.

[6] S. Ammoun, F. Nashashibi, Real time trajectory prediction for collision risk estimation between vehicles, 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing 2009, pp. 417-422.

[7] A. Zyner, S. Worrall, E. Nebot, A Recurrent Neural Network Solution for Predicting Driver Intention at Unsignalized Intersections, IEEE Robotics and Automation Letters, 3(3) (2018) 1759-1764.

[8] He Fujian, Jiang Guokai, Ji Guotian, Tian Xiaodi, Wu Feiyan & Tang Shasha, Development Status and Problem Analysis of High-Definition Maps, Auto Electric Parts, (08) (2024) 62-64.

[9] N. Djuric, V. Radosavljevic, H. Cui, T. Nguyen, F.C. Chou, T.H. Lin, N. Singh, J. Schneider, Uncertainty-aware Short-term Motion Prediction of Traffic Actors for Autonomous Driving, 2020 IEEE Winter Conference on Applications of Computer Vision (WACV) 2020, pp. 2084-2093.

[10] W. Zeng, M. Liang, R. Liao, R. Urtasun, LaneRCNN: Distributed Representations for Graph-Centric Motion Forecasting, 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021, pp. 532-539.

[11] Liu Zhiqiang, Wu Xuegang, Ni Jie & Zhang Teng, Driving Intention Recognition Based on HMM and SVM Cascade Algorithm, Automotive Engineering, 40(07) (2018) 858-864.

[12] Zong Changfu, Wang Chang, He Lei, Zheng Hongyu, Zhang Zexing, Driving Intention Identification Based on Double-layer Implicit Markov Model, Automotive Engineering, 33(08) (2011) 701-706.

[13] Ji Xuewu, Fei Cong, He Xiangkun, Liu Yulong, Liu Yahui, Driving Intention Recognition and Vehicle Trajectory Prediction Based on LSTM Network, China Journal of Highway and Transport, 32(06) (2019) 34-42.

[14] K. Min, D. Kim, J. Park, K. Huh, RNN-Based Path Prediction of Obstacle Vehicles With Deep Ensemble, IEEE Transactions on Vehicular Technology, 68(10) (2019) 10252-10256.

[15] X. Xu, W. Liu, L. Yu, Trajectory prediction for heterogeneous traffic-agents using knowledge correction data-driven model, Information Sciences, 608 (2022) 375-391.

[16] N. Kaempchen, K. Weiss, M. Schaefer, K.C.J. Dietmayer, IMM object tracking for high dynamic driving maneuvers, IEEE Intelligent Vehicles Symposium, 2004, 2004, pp. 825-830.

[17] V. Lefkopoulos, M. Menner, A. Domahidi, M.N. Zeilinger, Interaction-Aware Motion Prediction for Autonomous Driving: A Multiple Model Kalman Filtering Scheme, IEEE Robotics and Automation Letters, 6(1) (2021) 80-87.

[18] F. Altché, A.d.L. Fortelle, An LSTM network for highway trajectory prediction, 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 353-359.

[19] L. Xin, P. Wang, C.Y. Chan, J. Chen, S.E. Li, B. Cheng, Intention-aware Long Horizon Trajectory Prediction of Surrounding Vehicles using Dual LSTM Networks, 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 1441-1446.

[20] N. Nikhil, B.T. Morris, Convolutional Neural Network for Trajectory Prediction, in: L. Leal-Taixé, S. Roth (Eds.) Computer Vision – ECCV 2018 Workshops, Springer International Publishing, Cham, 2019, pp. 186-196.

[21] D. Lee, Y.P. Kwon, S. McMains, J.K. Hedrick, Convolution neural network-based lane change intention prediction of surrounding vehicles for ACC, 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 1-6.

[22] S. Casas, W. Luo, R. Urtasun, IntentNet: Learning to Predict Intention from Raw Sensor Data, Proceedings of The 2nd Conference on Robot Learning, PMLR, Proceedings of Machine Learning Research, 2018, pp. 947--956.

[23] N. Deo, M.M. Trivedi, Convolutional Social Pooling for Vehicle Trajectory Prediction, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 1549-15498.

[24] M. Schreiber, S. Hoermann, K. Dietmayer, Long-Term Occupancy Grid Prediction Using Recurrent Neural Networks, 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 9299-9305.

[25] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, Recurrent models of visual attention, Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, MIT Press, Montreal, Canada, 2014, pp. 2204–2212.

[26] X. Li, X. Ying, M.C. Chuah, GRIP: Graph-based Interaction-aware Trajectory Prediction, 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 2019, pp. 3960-3966.

[27] R. Chandra, T. Guan, S. Panuganti, T. Mittal, U. Bhattacharya, A. Bera, D. Manocha, Forecasting Trajectory and Behavior of Road-Agents Using Spectral Clustering in Graph-LSTMs, IEEE Robotics and Automation Letters, 5(3) (2020) 4882-4890.

[28] J. Ziegler, P. Bender, M. Schreiber, et al., Making Bertha Drive—An Autonomous Journey on a Historic Route, IEEE Intelligent Transportation Systems Magazine, 6(2) (2014) 8-20.

[29] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, R. Urtasun, Learning Lane Graph Representations for Motion Forecasting, Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II, Springer-Verlag, Glasgow, United Kingdom, 2020, pp. 541–556.

[30] D. Hu, An Introductory Survey on Attention Mechanisms in NLP Problems, in: Y. Bi, R. Bhatia, S. Kapoor (Eds.) Intelligent Systems and Applications, Springer International Publishing, Cham, 2020, pp. 432-448.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., Long Beach, California, USA, 2017, pp. 6000–6010.

[32] Zhao J, Li X, Xue Q, et al., Spatial-channel transformer network for trajectory prediction on the traffic scenes, arXiv preprint arXiv:2101.11472, 2021.

[33] Y. Liu, J. Zhang, L. Fang, Q. Jiang, B. Zhou, Multimodal Motion Prediction with Stacked Transformers, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 7573-7582.

[34] C. Liu, S. Yang, Q. Xu, Z. Li, C. Long, Z. Li, R. Zhao, Spatial-Temporal Large Language Model for Traffic Prediction, 2024 25th IEEE International Conference on Mobile Data Management (MDM), 2024, pp. 31-40.

[35] M. Peng, X. Guo, X. Chen, K. Chen, M. Zhu, L. Chen, F.-Y. Wang, LC-LLM: Explainable lane-change intention and trajectory predictions with Large Language Models, Communications in Transportation Research, 5 (2025) 100170.

[36] Z. Lan, L. Liu, B. Fan, Y. Lv, Y. Ren, Z. Cui, Traj-LLM: A New Exploration for Empowering Trajectory Prediction With Pre-Trained Large Language Models, IEEE Transactions on Intelligent Vehicles, (2024) 1-14.