Trajectory Prediction with Observations of Variable-Length for Motion Planning in Highway Merging Scenarios*

Sajjad Mozaffari[1], MReza Alipour Sormoli[1], Konstantinos Koufos[1], Graham Lee[1], and Mehrdad Dianati[2]

***Abstract*— Accurate trajectory prediction of nearby vehicles is crucial for the safe motion planning of automated vehicles in dynamic driving scenarios such as highway merging. Existing methods cannot initiate prediction for a vehicle unless observed for a fixed duration of two or more seconds. This prevents a fast reaction by the ego vehicle to vehicles that enter its perception range, thus creating safety concerns. Therefore, this paper proposes a novel transformer-based trajectory prediction approach, specifically trained to handle any observation length larger than one frame. We perform a comprehensive evaluation of the proposed method using two large-scale highway trajectory datasets, namely the highD and exiD. In addition, we study the impact of the proposed prediction approach on motion planning and control tasks using extensive merging scenarios from the exiD dataset. To the best of our knowledge, this marks the first instance where such a large-scale highway merging dataset has been employed for this purpose. The results demonstrate that the prediction model achieves state-of-the-art performance on the highD dataset and maintains lower prediction error w.r.t. the constant velocity across all observation lengths in exiD. Moreover, it significantly enhances safety, comfort, and efficiency in dense traffic scenarios, as compared to the constant velocity model.**

I. Introduction

The successful development of highly automated driving systems in interactive driving scenarios requires understanding the potential evolution of the driving situation in the immediate future. A prime example of such a scenario is merging onto highways, particularly during peak hours. In this context, autonomous vehicles (AVs) must continually adjust their motion based on the trajectory predictions of other road users on the main carriageway. Recent progresses in deep learning have facilitated the development of advanced interaction-aware prediction models that exhibit significantly lower prediction errors and extended prediction horizons [1]. However, there remains a lack of research to investigate the practical limitations and the impact of these methods on downstream motion planning and control in automated vehicles.

One drawback of existing learning-based prediction approaches is their reliance on observing the states of nearby vehicles for a fixed duration, typically two seconds or more [2], [3], [4]. That delays the availability of predictions for the first few seconds of the first time when a vehicle enters the field of view of the AV's perception system. These delays in prediction can hinder timely and safe decision-making by AVs, especially in highly interactive driving scenarios.

In addition, many existing trajectory prediction approaches lack a thorough analysis of how the learning-based prediction impacts downstream motion planning and control tasks. Even though some studies [5], [6], [7] offer qualitative analysis of motion planning using prediction data, they often fall short of providing a comprehensive comparison between learning-based predictions and conventional methods like the constant velocity prediction model. Additionally, these evaluation methods often neglect to adequately assess the influence of prediction quality in diverse driving conditions, such as various traffic densities. To the best of our knowledge, there is a notable absence of studies evaluating the impact of learning-based prediction on motion planning and control for merging onto highways.

To address the aforementioned shortcomings, this paper first proposes a new formulation of vehicle trajectory prediction where a variable-length sequence of past observations is employed as opposed to the fixed duration used in existing methods. Then, we design a novel transformer-based prediction model using a tailored input feature list for highway merging scenarios. The proposed model is specifically trained to handle variable-length observations. The prediction data is then utilised in motion planning and control for vehicles merging onto highways. To this end, the prediction data is encoded into a potential field and fed to a Model Predictive Control (MPC)-based motion planning and control algorithm. We evaluate the performance of the prediction model and the prediction-based motion planning using a large-scale highway merging dataset, namely the exiD dataset [8]. Additionally, we separately evaluate the performance of the prediction model using a benchmark highway driving dataset, namely the highD dataset [9], to facilitate comparison with several state-of-the-art highway trajectory

*This work was part of the Hi-Drive project. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101006664.

[1]S. Mozaffari, M. Alipour Sormoli, K. Koufos, G. Lee are with WMG, University of Warwick, Coventry, U.K. sajjad.mozaffari@warwick.ac.uk

[2]M. Dianati is with School of Electronics, Electrical Engineering and Computer Science, Queen's University of Belfast. m.dianati@qub.ac.uk

prediction models. The contributions of this paper can be summarised as follows:

- A novel formulation for vehicle trajectory prediction that accommodates variable-length observations.
- A novel transformer-based vehicle trajectory prediction model incorporating interaction and map-aware features for highway merging scenarios.
- Statistical evaluation of the prediction model and its impact on downstream motion planning and control in highway merging scenarios.

II. Related Work

This section reviews the existing studies on trajectory prediction, particularly focusing on two key aspects: (1) Learning-based techniques employed for trajectory prediction, and (2) Studies that analyse the impact of nearby vehicle trajectory prediction on the motion planning of the Ego vehicle (EV).

A. Learning-based Trajectory Prediction

Learning-based methods for trajectory prediction have been recently reviewed based on their input representation, prediction model and output type in [1]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memories (LSTMs), have been widely utilized for vehicle trajectory prediction [10], [11], [12], [13]. However, with the introduction of the attention mechanism [14], RNNs are gradually being outperformed in many sequence-to-sequence tasks, such as speech recognition [15] and natural language processing [14]. Additionally, Convolutional Neural Networks (CNNs) are employed in vehicle trajectory prediction due to their advantages in spatial interaction learning [4], [16].

Transformer neural networks have emerged as an alternative to both RNNs and CNNs [17], [18], [19], [20]. Giuliari et al. [17] demonstrated that simple transformer neural networks outperform state-of-the-art LSTM prediction models in human trajectory prediction tasks. In [18], the multimodal vehicle trajectory prediction task was addressed using transformers where each predicted trajectory is conditioned on a separate attention head. Gao et al. [19] employed dual transformers to predict both the intention and the trajectory of a target vehicle leveraging vehicle interaction features and the history of the lateral states of the target vehicle. In [21], a transformer encoder was used to learn latent representation from a list of interaction-aware, context-aware, and dynamic-aware input data. That representation was used to predict multimodal manoeuvres and trajectories for a target vehicle. Similarly, in this paper, we employ a transformer neural network using input features specially designed for merging scenarios. Notably, unlike existing studies that require a fixed observation length of two to three seconds, our proposed method can predict a vehicle's trajectory with a minimum of two time-step observations (equivalent to 0.4 seconds in our setting).

B. Trajectory Prediction for Motion Planning

Wang et al. [6] designed an MPC-based motion planner for risk mitigation leveraging trajectory prediction using LSTMs. The LSTM model was trained and evaluated on the highD dataset [9], however, the planning algorithm was evaluated using two selected scenarios from the highD dataset. Tang et al. [7] expanded upon this work by estimating the uncertainty of prediction using a deep ensemble technique. Then, the effects of uncertainty-aware prediction on motion planning were evaluated in a single cut-in and merging scenario. Building upon these studies, we further extend the research field by conducting a comprehensive statistical analysis of the planning algorithm's performance.

In [5], [22], multimodal joint prediction of nearby vehicles' trajectories was used for contingency planning. Chen et al. in [5] utilised an MPC to plan a trajectory for each prediction mode, and then the results were discussed for a specific driving scenario qualitatively. Cui et al. [22] used a sampling-based planning approach in simulation with reactive agents. The simulation scenarios were initialised with real-world urban data. In [23], an LSTM neural network was used to predict the trajectory of the EV's nearby vehicles in a multi-lane turn intersection, and then the future motion of the EV was planned using the prediction data and MPC planner. The results of the planning algorithm were compared with a path following and constant turning rate and velocity predictions. Different from these studies, we target highway merging scenarios and analyse the impact of trajectory prediction of other vehicles on joint motion planning and control of merging vehicles.

III. Methodology

This section first defines the vehicle trajectory prediction problem and the system model. Then describes the proposed trajectory Prediction model with Observation of Variable Length (POVL). Finally, it discusses the motion planning and control algorithm. Fig. 1 illustrates an overview of the prediction and planning pipelines using an exemplary driving scenario.

A. System Model and Problem Definition

We consider a semi- or fully-automated Ego Vehicle (EV) merging from a single-lane slip road into a main carriageway with an arbitrary number of lanes. All roadways can be straight or curvy. The EV intends to predict the future trajectory of vehicles along both the main carriageway and the slip road, and these predictions are subsequently fed into its motion planning and control module. The tracking data of these vehicles and the lane markings are assumed to be available for the EV, e.g., through onboard perception, V2X communication, or cloud. Assuming that the future trajectories of vehicles are independent, the EV predicts the trajectory of one Target Vehicle (TV) at a time and for all perceived vehicles in the scene.

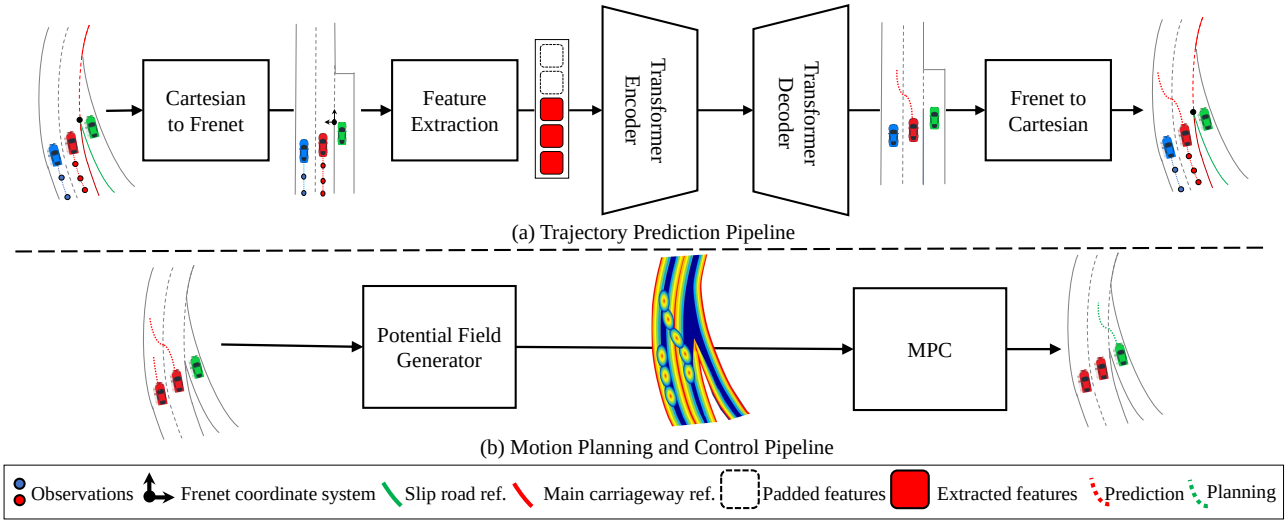

Fig. 1. An overview of the proposed prediction and motion planning pipelines with an illustrative exemplar scenario. The reference paths for coordinate conversion are depicted with solid green and red lines. The predicted and planned trajectories are depicted with dashed red and green lines, respectively. Note that the example output of the Potential Field Generator box is cumulatively drawn over several prediction steps for illustrative purposes.

The problem of trajectory prediction is defined as estimating the sequence of TV's future x-y positions during T_{pred} time-steps of prediction window $Y_{TV} = \{(x_1, y_1), \dots (x_{T_{pred}}, y_{T_{pred}})\}$ given the states of the TV and its surrounding environment during T_{obs} time-steps of the observation window. The observation length for a TV can vary between a minimum and maximum value, i.e., $T_{obs} \in [T_{min}, T_{max}]$.

B. Prediction model

Fig. 1-a illustrates the overview of the prediction framework. The processing steps are summarised below:

1) The TV's tracking data is converted to the Frenet coordinate system.
2) The input features are computed using the tracking and map data.
3) Transformer encoder-decoder neural network is used to predict the future trajectory of the TV.
4) The prediction data is converted back to the Cartesian coordinate system.

1) Coordinate Conversion: The prediction is conducted in the Frenet coordinate system, which describes the movement of vehicles in terms of along-track ($\vec{s}$) and cross-track ($\vec{d}$) components. As the curvature of the slip road may differ from that of the main carriageway, two distinct reference paths are considered for vehicles on each road segment, as illustrated in Fig. 1. The crossing point of the reference paths serves as the origin of the new coordinate system. A vehicle is first associated either with the slip road or the main carriageway reference path[1]. Afterwards, its cross-track component is determined by calculating the distance of the orthogonal projection between the vehicle's centre and the associated reference path. The along-track component is calculated as the arc length along the reference path from the origin to the orthogonal projection point of the vehicle's centre. Similarly, the prediction data is converted back from Frenet to Cartesian coordinate system.

[1]Note that it is assumed vehicles do not deviate from the road boundaries.

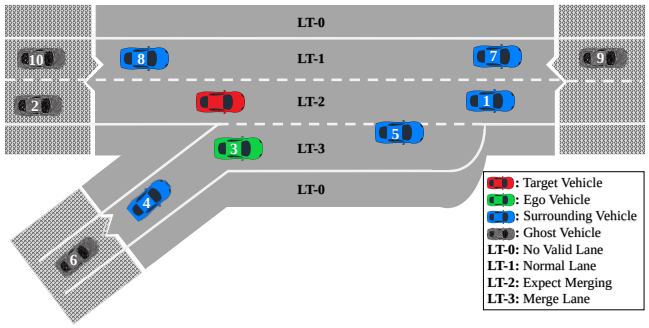

Fig. 2. Illustration of the maximum ten surrounding vehicles for a target vehicle and different lane types (best seen in colour).

2) Input Features: The input to the prediction model consists of a sequence of feature lists with a maximum observation length of T_{max} time steps. If there are less than T_{max} observations available for a vehicle, the rest of the sequence is padded with zero values.

The observation at each time step is represented by a list of 28 features, categorised into three groups. The first group describes the motion of the TV, encompassing features including the TV's lateral position in the lane, longitudinal velocity, and lateral and longitudinal acceleration. The second group characterises the TV's interaction with surrounding vehicles (SVs), considering the lateral and longitudinal distance between the TV and each SV. A maximum of 10 SVs are considered including the preceding/following vehicles on the TV's

lane (1/2), the right close preceding/following vehicles (3/4), right far preceding/following vehicles (5/6), left close preceding/following vehicles(7/8), and left far preceding/following vehicles(9/10). A ghost vehicle at far distances replaces the feature list of each SV that doesn't exist or is not in the perception range of the EV. Fig. 2 shows the numbered labelling of SVs for an example merging scenario. Finally, the third group of features pertains to the driving environment, including lane width and the types of right and left lanes relative to the TV. This study defines four lane types: (0) no lane, (1) normal lanes (where crossings are anticipated into and from them), (2) expect merging lanes (where crossings are expected into them), and (3) merge lane (where crossings are expected from them). Fig. 2 provides an illustration of the surrounding vehicles and the various lane types.

3) Transformer Model: The prediction model includes a transformer encoder and decoder components, initially introduced in [14]. Transformer neural networks utilise multi-head self-attention and cross-attention mechanisms to identify and focus on the informative segment of the input sequence in alignment with the underlying training objectives. Interested readers are referred to [14] for further details about transformers.

The transformer encoder is used to encode the variable length observation into a sequence of latent representations capturing the spatiotemporal dependencies of the input sequence. The transformer decoder then utilises the latent representations to predict the displacement of the TV at each time step during the prediction window. The masking mechanism in transformers is used to filter out the padded sequence of input data.

The encoder and decoder models consist of two transformer layers each with eight attention heads, a model dimension of 512, and a feed-forward hidden layer dimension of 128. To represent the uncertainties in the prediction output, the model estimates the parameters of a bivariate Gaussian distribution of displacement in along-track and cross-track dimensions for each prediction step.

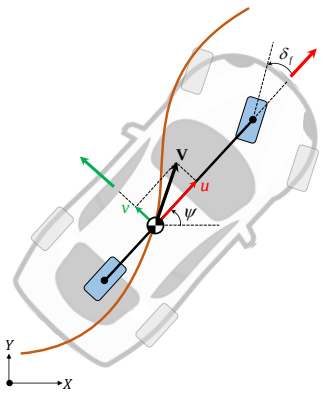

Fig. 3. Kinematic variables and parameters of the bicycle model used in the equation of motion.

C. Motion Planning Algorithm

This section describes the Model Predictive Control (MPC a.k.a receding horizon control) motion planning algorithm for an EV merging into the highway. Given the predicted future trajectories of dynamic objects in the driving environment, MPC calculates the optimised trajectory for the EV by minimising a given cost function [24]. Specifically, the optimisation cost at a time step t is a function of the EV dynamics/constraints (U^{ev}), the perceived environment represented by a potential field (U^{env}), and a reference signal (U^{ref}) over the planning horizon p ($p \leq T_{pred}$). One may write

$$\mathcal{C}_t = \sum_{i=0}^{p} U^{ev}_{t+i} + U^{env}_{t+i} + U^{ref}_{t+i}. \tag{1}$$

Each term of the cost function is further explained in the subsequent sections.

1) Vehicle Dynamics: The (dynamic) bicycle model has been adopted for modelling the dynamic motion of the EV. Accordingly, the EV motion is governed by a set of nonlinear equations [24]:

$$\begin{aligned} \dot{x} &= f\left(x, u_c\right) \\ x &= \left[\ u, v, \psi, \dot{\psi}, (X, Y)\ \right]^T \\ u_c &= \left[\ F_u, \delta_f\ \right]^T, \end{aligned} \tag{2}$$

where x and u_c are the state vector and control inputs respectively, and $f(\cdot)$ represents a set of six nonlinear functions. The state vector x represents the EV's state including its longitudinal and lateral velocities (u and v), the heading and yaw rate (ψ and $\dot{\psi}$), and the Cartesian location of its centre of mass (X and Y). Fig. 3 illustrates these parameters on the bicycle model of the EV. Moreover, the input vector u_c consists of throttle/brake force (F_u) and the front wheel steering angle (δ_f).

The state equations are adaptively linearised around their operating point to facilitate the optimisation process yielding the following simplified form of state equations: $\dot{x} = Ax + Bu_c$, where A and B are the Jacobians of $f\left(x, u_c\right)$ with respect to x and u_c, respectively. Finally, the term U^{ev} in Eq. 1 is defined as the quadratic weighted sum of the control inputs and their rate of change

$$U^{ev} = u_c^T Q_1 u_c + \dot{u}_c^T Q_2 \dot{u}_c, \tag{3}$$

where Q_1 and Q_2 are constant square diagonal matrices to tune the contribution of the control input and its rate in the overall cost.

2) Potential field representation of driving environment: Repulsive potential fields have been widely adopted in path planning of autonomous vehicles (AVs) for modelling the interactions between the EV and other elements in the driving scene such as obstacles and boundaries. A repulsive potential field guides the EV

away from the associated obstacle/boundary. A well-engineered repulsive potential field enables differentiation between diverse types of obstacles, further enhancing driving safety. In a highway merging scenario, the main obstacles that should be avoided can be broadly categorised into three types: Moving/static objects such as other vehicles and road boundaries as non-crossable obstacles and lane markings as crossable obstacles. A repulsive potential is designed for each category of obstacles which contributes to the overall potential field representing the driving context.

The potential field for each of the other vehicles, V_o, can be defined based on its distance to the EV, after it is appropriately weighted in the lateral and longitudinal directions.

$$V_o = \frac{a_o}{\left(\frac{x-x_{obs}}{X_c}\right)^{c_x} + \left(\frac{y-y_{obs}}{Y_c}\right)^{c_y}}, \tag{4}$$

where (x_{obs}, y_{obs}) are the obstacle location, a_o is the obstacle vehicle potential field magnitude, X_c, Y_c are constants determining the rate of change for the potential field in x and y directions, and c_x, c_y is another set of weights allowing to model every vehicle in the 2d plane by an ellipse instead of a circle. One can find in Fig. 4 (top), an example illustration.

The potential field associated with road boundaries, V_r, is defined according to the EV's distance to the boundary, d_r. In that case, the potential should be high when the EV approaches near the boundary but can be degenerated to zero when the distance d_r becomes larger than a threshold, D_r. Specifically,

$$V_r = \begin{cases} a_r(d_r - D_r)^2, & d_r \leq D_r \\ 0, & d_r > D_r, \end{cases} \tag{5}$$

with $a_r > 0$ being the maximum of the potential field V_r.

Finally, the lane-marking potential, V_l, may have a concave shape with its magnitude a_l being lower than the other two categories, i.e., $a_l < a_o$ and $a_l < a_r$ as can also be seen in Fig. 4, since lane crossing should be permitted for overtaking.

$$V_l = a_l \exp\left(-b_l d_l^2\right), \tag{6}$$

where d_l is the distance to the lane marks and $b_l > 0$ determines the rate of change.

The environment cost for each time step in Eq. 1 is obtained by adding up the terms calculated in Eq. (4) - (6):

$$U^{env} = V_r + V_l + \sum\nolimits_o V_o, \tag{7}$$

where the summation is over all obstacle vehicles. Note that obstacle vehicles' positions at each future timestep are obtained from their corresponding predicted trajectory.

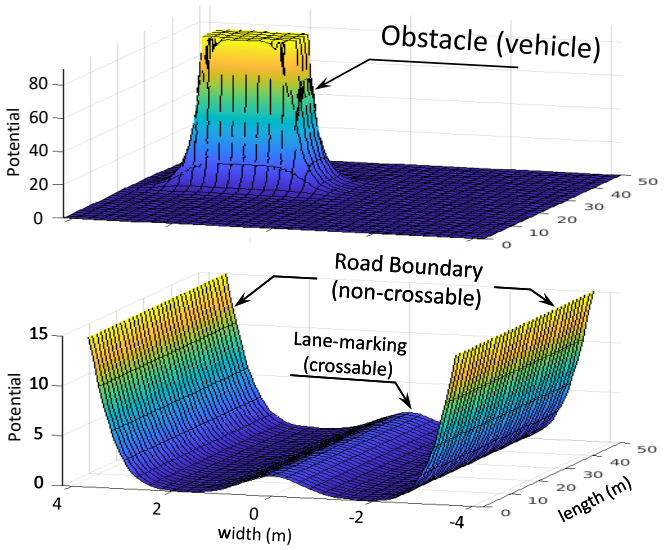

Fig. 4. Potential field for modelling semantic information in the driving environment such as other vehicles, road boundaries and lane markings (the axes have been scaled).

3) Reference signals: Two reference signals are defined for the EV to follow, namely the distance from the centre-line of the desired merging lane (y_{des}) and the nominal speed of that lane (u_{des}). This is to guarantee progress in the EV's trajectory, otherwise, the EV may stop moving to minimise the other terms in the cost function. Accordingly, the contribution of the reference signals in the overall cost is:

$$U^{ref} = Y^T Q_3 Y, \qquad Y = \left[\ y_{des}, u_{des}\ \right]^T, \tag{8}$$

where Q_3 is a constant square diagonal matrix to tune the contribution of the reference signal error in the final cost.

IV. Performance Evaluation

This section offers a performance evaluation of the proposed prediction model and analyses its impact on motion planning and control. To this end, it first presents the datasets used for evaluation, followed by the implementation details. Then, we provide the selected prediction and planning metrics. Subsequently, the results of prediction and planning models are discussed.

A. Dataset

1) exiD Dataset: We utilise the exiD dataset [8] for both trajectory prediction and motion planning evaluation in highway merging scenarios. This dataset comprises a large-scale collection of naturalistic vehicle trajectories (69,172 vehicles) at highway entries and exits, recorded by drones in Germany in 2022. Among seven different locations reported in the dataset, the merging data of four locations are used in this study which matches with the system model defined in section III-A (i.e., a single-lane slip road merging into two- or three-lane main roads). The selected data contains 49

recordings from which 45 are used for training and 4 for evaluation of the prediction model, one per location. Recording number 39, from the prediction test set, is used for evaluating the planning algorithm.

2) highD Dataset:: The prediction model is also separately trained and evaluated on the highD dataset [9], which is a benchmark highway trajectory dataset used in several trajectory prediction studies [13], [19], [1]. The highD dataset contains 110,000 vehicle tracks from six different highways in Germany. The dataset is recorded using a drone from mainly straight highways and includes different levels of traffic densities. The vehicle tracks in highD are exclusively divided into the train, validation, and test set with the ratio of 70%, 10%, and 20%, following the experimental protocol of existing studies [25], [20].Both highD and exiD datasets are reported with 25 frames per second which are reduced to 5 in the prediction task.

B. Implementation Details

The prediction model is trained using Adam optimiser [26] with a learning rate of 0.0001 for a maximum of 10^5 batches, each with a size of 2000 samples. Observations ranging from a minimum of two (0.4 seconds) to a maximum of 15 timesteps (3 seconds) are used to predict the next 25 timesteps (5 seconds). The prediction model is implemented using PyTorch library [27] and is run on GeForce RTX 2080 TI GPU. The source code of this study is available at https://github.com/SajjadMzf/TrajPred

The planning algorithm is evaluated for large-scale highway merging scenarios with realistic data obtained from the exiD dataset. Each scenario is initialised with a merging vehicle at various merging statuses (i.e., different distances to the merging point). Motion planning is performed for the next five seconds during which the corresponding prediction data of other vehicles are being used. In total 97 merging vehicles are used for motion planning evaluation. Sequential quadratic programming is used to optimise the planning cost function (Eq. 1) over the planning horizon. The planning algorithm is implemented in MATLAB.

C. Evaluation Metrics

1) Prediction Metric: The Root Mean Square Error (RMSE) between the predicted trajectory of vehicles and their ground truth is used to evaluate the accuracy of the prediction model:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[(z_i - z_{i,gt})^2 + (w_i - w_{i,gt})^2\right]}, \quad (9)$$

where (z_i, w_i) are the predicted (Cartesian) coordinates at the i-th sample, out of N total samples, and $(z_{i,gt}, w_{i,gt})$ is the associated ground truth.

TABLE I

COMPARISION OF BASELINE STUDIES USING RMSE(M) AT DIFFERENT PREDICTION HORIZONS EVALUATED ON HIGHD AND EXID DATASET. THE BEST AND SECOND BEST ERRORS ARE MARKED AS **bold** AND UNDERLINED.

	Model	1s	2s	3s	4s	5s
highD	CV	0.10	0.33	0.69	1.17	1.76
	CS-LSTM [4]	0.19	0.57	1.16	1.96	2.96
	MHA-LSTM [13]	**0.06**	**0.09**	0.24	0.59	1.18
	Dual Trans. [19]	0.41	0.79	1.11	1.40	-
	MMnTP [19]	0.19	0.36	0.56	0.82	1.19
	POVL (ours)	0.12	0.18	**0.22**	**0.53**	**1.15**
exiD	CV	0.25	0.63	1.19	1.92	2.82
	POVL (ours)	**0.17**	**0.41**	**0.76**	**1.21**	**1.75**

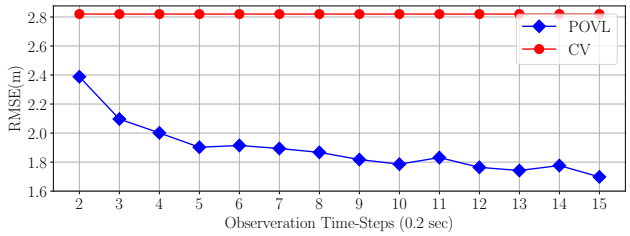

Fig. 5. RMSE of the proposed prediction model in different observation lengths on exiD dataset.

2) Planning Metrics: The motion planning algorithm is evaluated based on three criteria, namely, safety, comfort, and efficiency. The inverse of the average Time-To-Collision, $iTTC$, is reported to evaluate the safety of the planned trajectories, the average Jerk represents the comfort, and the average force is reported for fuel efficiency.

D. Prediction Results

1) Comparison with baselines: Table I compares the performance of the proposed model, some state-of-the-art highway trajectory prediction models, and Constant Velocity (CV) prediction evaluated on highD and exiD datasets. To the best of our knowledge, no trajectory prediction studies have been conducted using the exiD dataset. The selected state-of-the-art models include LSTM-based models such as MHA-LSTM [13], CNN-LSTM-based models such as CS-LSTM [4], and Transformer-based models such as [19], [1]. The results show that the proposed prediction model outperforms SOTA and CV models on both datasets specifically in longer prediction horizons. The CV predictor has a relatively low error and even outperforms one of the baseline learning-based models (e.g., CS-LSTM). We argue that this is because vehicles tend to keep their velocity in highway driving, therefore on average a CV prediction approach can achieve comparable results to learning-based methods on distance-based metrics such as RMSE.

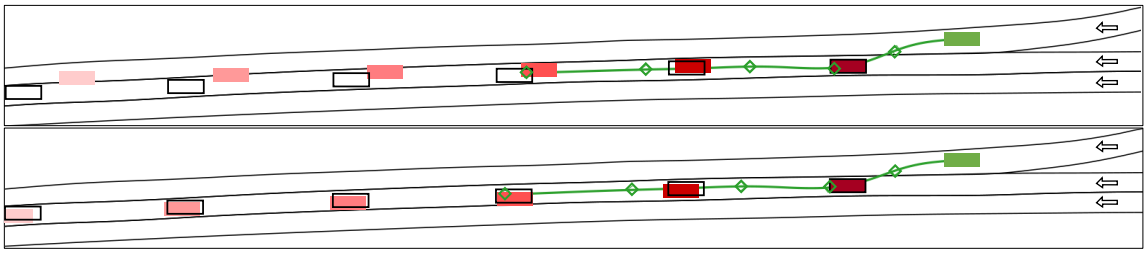

Fig. 6. Trajectory prediction and motion planning for a merging scenario extracted from exiD dataset (top: CV prediction, bottom: POVL prediction). The green line shows the planned trajectory of the EV (green bounding box) with green diamonds representing one-second intervals for the next five seconds (planning horizon). The dark red bounding box with black borders shows the current location of the TV on the main carriageway. The transparent bounding boxes with black borders show the ground-truth position of the TV in the next five seconds. The red bounding boxes with increasing levels of fading show the predicted location of the TV in the next five seconds.

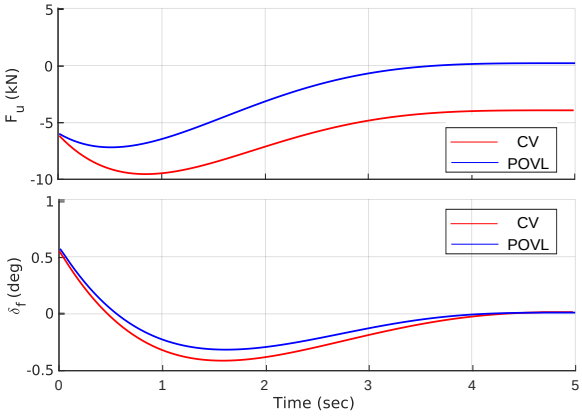

Fig. 7. Control input signal (top: force, and bottom: steering angle) in motion planning using CV and POVL prediction methods for the scenario in Fig. 6. Sharper steering and more force magnitude are observed in motion planning using CV.

2) Performance in different observation lengths: Fig. 5 shows the performance of the proposed prediction model evaluated on the exiD dataset for different observation lengths. The results show that the proposed method even with 0.4 seconds (i.e., 2-time steps) outperforms the CV model. Note that most existing learning-based prediction studies do not provide a prediction for observations less than 2-3 seconds.

E. Prediction Impact on Planning

1) Statistical Analysis: Table II shows the statistical analysis for the impact of trajectory prediction accuracy on motion planning and control metrics. We separately run motion planning algorithm with Ground-Truth (GT) futures of the vehicles as well as the CV and the proposed POVL models. The safety, comfort, and efficiency are reported for different traffic densities from the EV's perspective. To this end, the relevant metrics are calculated in seven sets, based on the max distance of the nearby vehicles to the EV's planned trajectory in each time step of the five-second simulation. The results show that using the proposed prediction model instead of CV can lead to a significant 9.15% improvement in driving safety (i.e., *iTTC*) in total. However, the impact on the ride comfort (i.e., jerk) and fuel efficiency (i.e., Force) is

TABLE II

COMPARISON OF DIFFERENT PREDICTION APPROACHES ON COMFORT, EFFICIENCY, AND SAFETY OF MOTION PLANNING

Metric	Pred. Model	Distance from other vehicles to the EV (m)						
		<3	<5	<7	<10	<20	<30	<50
iTTC $\times 100$ (1/s)	GT	11.71	17.13	16.59	16.74	16.83	17.64	18.54
	CV	14.96	21.12	19.96	18.60	18.63	19.51	20.42
	POVL	12.09	17.95	16.86	16.66	16.86	17.72	18.55
	Rel.%*	19.18	15.00	15.53	10.43	9.50	9.17	9.15
Jerk (m/s^3)	GT	4.32	4.41	4.49	4.76	4.57	4.64	4.80
	CV	5.61	5.23	5.05	4.93	4.65	4.69	4.87
	POVL	4.50	4.50	4.55	4.77	4.57	4.63	4.82
	Rel.%*	19.78	13.95	9.90	3.24	1.72	1.27	1.02
Force (kN)	GT	10.95	8.85	8.88	8.82	8.77	8.89	8.97
	CV	12.65	10.25	9.61	8.82	8.84	8.90	9.13
	POVL	10.98	8.87	9.01	8.81	8.80	8.92	8.99
	Rel.%*	13.20	13.46	6.24	0.11	0.45	-0.2	1.53

* The percentage of relative improvement of POVL compared to CV: (CV-POVL)/CV.

lower with values of 1.02% and 4.80%, respectively. In denser traffic from EV's perspective (e.g., distance from other vehicles to EV less than 3 meters), the impact significantly increases to 19.18%, 19.78%, and 13.20 % for safety, comfort, and efficiency, respectively. We argue that this is because although the average improvement in prediction accuracy of other vehicles is around 1 meter with POVL compared to CV, this improvement becomes relatively smaller in longer distances of the vehicle to the EV compared to shorter distances. Also, note that the magnitude of the potential field component of obstacle vehicles (i.e., V_o) in the motion planning cost function is negligible for far vehicles.

2) Qualitative results: Fig. 6 shows the trajectory prediction and motion planning results for an exemplary driving scenario in two cases i.e., the CV and the POVL (proposed) prediction model. Also, Fig. 7 depicts the corresponding planned control signals for both prediction approaches. In this scenario, the EV is merging behind a

vehicle. The CV model prediction contains a significant longitudinal error and a deviation to the right lane, which consequently results in sharper merging behaviour compared to a more accurate POVL prediction. In addition, the EV has to apply more negative force to avoid collision with the predicted vehicle in case of CV compared to POVL.

V. Conclusion

This paper presents a novel transformer-based prediction model with a variable-length observation window that enhances the accuracy of trajectory prediction in the highway driving scenario. Through statistical analysis conducted on 97 merges, it is illustrated that the improved trajectory prediction accuracy, as compared to conventional prediction methods, can significantly enhance the performance of a model predictive motion planner and improve safety, comfort, and fuel efficiency specifically in dense driving scenarios. Future research endeavours will concentrate on implementing prediction and motion planning algorithms on sensory-equipped vehicles and addressing challenges related to real-world implementation, such as real-time performance and noisy sensor measurements.

References

[1] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 33–47, 2020.

[2] Z. Sheng, Y. Xu, S. Xue, and D. Li, "Graph-based spatial-temporal convolutional network for vehicle trajectory prediction in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17 654–17 665, 2022.

[3] J. Schmidt, J. Jordan, F. Gritschneder, and K. Dietmayer, "Crat-pred: Vehicle trajectory prediction with crystal graph convolutional neural networks and multi-head self-attention," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7799–7805.

[4] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 1468–1476.

[5] Y. Chen, B. Ivanovic, and M. Pavone, "Scept: Scene-consistent, policy-based trajectory predictions for planning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 103–17 112.

[6] H. Wang, B. Lu, J. Li, T. Liu, Y. Xing, C. Lv, D. Cao, J. Li, J. Zhang, and E. Hashemi, "Risk assessment and mitigation in local path planning for autonomous vehicles with lstm based predictive model," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 2738–2749, 2021.

[7] X. Tang, K. Yang, H. Wang, J. Wu, Y. Qin, W. Yu, and D. Cao, "Prediction-uncertainty-aware decision-making for autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 849–862, 2022.

[8] T. Moers, L. Vater, R. Krajewski, J. Bock, A. Zlocki, and L. Eckstein, "The exid dataset: A real-world trajectory dataset of highly interactive highway scenarios in germany," in *2022 IEEE Intelligent Vehicles Symposium (IV)*, 2022, pp. 958–964.

[9] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2118–2125.

[10] F. Altché and A. de La Fortelle, "An lstm network for highway trajectory prediction," in *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)*. IEEE, 2017, pp. 353–359.

[11] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based lstms," in *2018 IEEE intelligent vehicles symposium (IV)*. IEEE, 2018, pp. 1179–1184.

[12] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via lstm encoder-decoder architecture," in *2018 IEEE intelligent vehicles symposium (IV)*. IEEE, 2018, pp. 1672–1678.

[13] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Attention based vehicle trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 175–185, 2020.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[15] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu *et al.*, "Conformer: Convolution-augmented transformer for speech recognition," 2020.

[16] H. Cui, V. Radosavljevic, F.-C. Chou, T.-H. Lin, T. Nguyen, T.-K. Huang, J. Schneider, and N. Djuric, "Multimodal trajectory predictions for autonomous driving using deep convolutional networks," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2090–2096.

[17] F. Giuliari, I. Hasan, M. Cristani, and F. Galasso, "Transformer networks for trajectory forecasting," in *2020 25th international conference on pattern recognition (ICPR)*. IEEE, 2021, pp. 10 335–10 342.

[18] Z. Huang, X. Mo, and C. Lv, "Multi-modal motion prediction with transformer-based neural network for autonomous driving," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2605–2611.

[19] K. Gao, X. Li, B. Chen, L. Hu, J. Liu, R. Du, and Y. Li, "Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[20] S. Mozaffari, K. Koufos, and M. Dianati, "Multimodal manoeuvre and trajectory prediction for autonomous vehicles using transformer networks," *arXiv preprint arXiv:2303.16109*, 2023.

[21] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 33–47, 2022.

[22] A. Cui, S. Casas, A. Sadat, R. Liao, and R. Urtasun, "Lookout: Diverse multi-future prediction and planning for self-driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 16 107–16 116.

[23] Y. Jeong, S. Kim, and K. Yi, "Surround vehicle motion prediction using lstm-rnn for motion planning of autonomous vehicles at multi-lane turn intersections," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 1, pp. 2–14, 2020.

[24] Y. Rasekhipour, A. Khajepour, S.-K. Chen, and B. Litkouhi, "A potential field-based model predictive path-planning controller for autonomous road vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1255–1267, 2016.

[25] C. Tang and R. R. Salakhutdinov, "Multiple futures prediction," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[27] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.