Fusion Attention and Physics Correction Long Short-Term Memory Network model

Linlin Lang

***Abstract*— In order to accurately and rationally predict the future trajectories of vehicles, a Fusion Attention and Physics Correction Long Short-Term Memory Network (FAPC-LSTM) model is proposed. It can improve the accuracy and stability of self-driving car trajectory prediction. Traditional methods based on physical models rely on complex parameters and are difficult to adapt to complex scenarios. And purely data-driven models (e.g., LSTM) may output predictions that violate physical laws. FAPC-LSTM integrates vehicle state information through an encoding-decoding structure, utilizes an attention mechanism to capture key timing features, and dynamically constrains the prediction results through a physical correction layer. The experiments are based on the nuScenes dataset and compare the performance of the purely physical model, traditional LSTM and FAPC-LSTM. The results show that FAPC-LSTM significantly outperforms the comparison model in both speed and position prediction. Especially, it exhibits lower cumulative error and higher stability in long-term prediction, which verifies its effectiveness in practical complex driving scenarios.**

***Index Terms*— LSTM; Trajectory prediction; attention mechanism; vehicle engineering; physical correction**

I. INTRODUCTION

Accurate and reasonable prediction of vehicle trajectories can effectively improve the safety of autonomous vehicles. The goal of vehicle trajectory prediction is usually to estimate the trajectory at a specific time point or over a certain period. To achieve this, researchers have used various methods and models. These research approaches can be divided into two main categories.

The first category is physics-based vehicle trajectory prediction. This method uses mathematical formulas or kinematic models to build physical models of vehicle trajectories. The second category is data-driven prediction models. These models mainly rely on neural networks. They learn the patterns of vehicle trajectory changes by training algorithms on large amounts of historical data. Ref.[1] summarizes multiple vehicle trajectory prediction models. It includes both physics-based and data-driven approaches. It also compares the advantages, disadvantages, and suitable scenarios of these methods.

In physics-based motion prediction research, some scholars focus on vehicle dynamic behavior. They build prediction models based on dynamics to describe vehicle motion [2-3]. Ref.[4] combines Kalman filtering and probabilistic simulation for short-term motion prediction. Experimental results show that this method achieves high accuracy in short-term prediction. However, it is not suitable for long-term motion prediction.For long-term motion prediction, Ref.[5] proposes a state estimation method that combines constant yaw rate and acceleration models. It also integrates trajectory prediction based on maneuver recognition. Compared to the method in Ref.[4], this algorithm shows better prediction accuracy. However, the model involves many parameters, which increases complexity in practical applications.

In recent years, with the emergence of new technologies, neural networks have been widely used in state prediction. By training on large amounts of data, the prediction performance of neural networks has improved significantly. Vehicle trajectory prediction relies on massive amounts of driving data. Because of this, neural network-based trajectory prediction methods have gained increasing attention from researchers. For example, ref.[6] proposed a deep feedforward neural network prediction framework. Experiments on the NGSIM driving dataset showed that driver behavior, dynamic environment features, and historical trajectory data are closely related to prediction results. However, the network structure lacks effective sequential modeling, limiting it to fragmented trajectory prediction. Ref.[7] introduced a time-enhanced encoder-decoder network. It integrated real-time vehicle motion parameters and road topology, using a graph attention mechanism for spatial feature extraction. This method shows strong robustness but depends on high-definition maps, which are costly to update. Performance degrades when map data is incomplete. Ref.[8] used LSTM network to predict lateral displacement and longitudinal speed. Results showed improved accuracy compared to earlier methods, but the average error remained high.

In summary, physics-based or trajectory-based prediction methods often require many constraints and parameters. They struggle to provide accurate predictions in complex real-world driving scenarios. Additionally, model complexity limits their practical use. In contrast, data-driven neural networks show greater potential in handling dynamic driving environments. By learning from large datasets, they can automatically identify and predict vehicle behavior with better adaptability and higher accuracy, especially in complex scenarios. Despite progress, neural network-based methods still face challenges. For example, traditional LSTMs often fail to capture long-term dependencies, sometimes producing physically unrealistic predictions. As seen in these studies, vehicle trajectory prediction algorithms are still evolving. Further research is needed to improve prediction accuracy.

To address the above issues, Fusion Attention and Physics Correction Long Short-Term Memory Network (FAPC-LSTM) model is proposed. The model adopts an encoder-decoder structure. Vehicle state information is used as input. Historical state features and control commands during driving are combined for prediction. The trajectory is further refined by the physics-based correction layer.

Manuscript received April 06, 2025
Linlin Lang, Tiangong University, Tianjin, China

II. TRAJECTORY PREDICTION MODEL

The input of the proposed model is defined as $x(t) = (u(t), y(t))$, which consists of control commands and dynamic observation parameters. Here, $u(t)$ includes the acceleration command A_c, braking command D_c, and steering command S_c. $y(t)$ contains the front wheel steering angle τ, longitudinal velocity v_x, lateral velocity v_y, yaw angle Ψ, and vehicle coordinates x_a, y_a. The model outputs the longitudinal velocity v_x, lateral velocity v_y, yaw angle Ψ, and vehicle coordinates x_a, y_a. Although high-frequency sampling can be used to obtain data, the vehicle's dynamic behavior is affected by multiple interacting physical factors in complex driving scenarios such as low-friction roads or emergency obstacle avoidance. The vehicle's motion state is not only related to historical control commands and perception data, but must also strictly follow vehicle dynamics laws.

Therefore, based on the structure in Ref. [9], the proposed model introduces an explicit constraint layer based on vehicle dynamics. The neural network output is corrected in real-time through physical equations, achieving dual assurance of data-driven prediction and physical constraints. The model architecture is shown in Figure 1. The architecture consists of five modules: multi-source data preprocessing, LSTM encoder, attention mechanism, fully connected layer, and physics-based correction module. The multi-source data preprocessing module is responsible for systematically organizing raw collected data and converting it into a format directly readable by the model. The LSTM encoder processes various input factors to extract temporal context features and obtain feature information. The attention mechanism learns attention coefficients and feature information, focusing on key temporal features such as deceleration changes before emergency braking and steering angle saturation. The fully connected layer obtains filtered feature information and provides preliminary prediction results. The physics-based correction module applies dual constraints to the predicted values based on prior knowledge of vehicle dynamics.

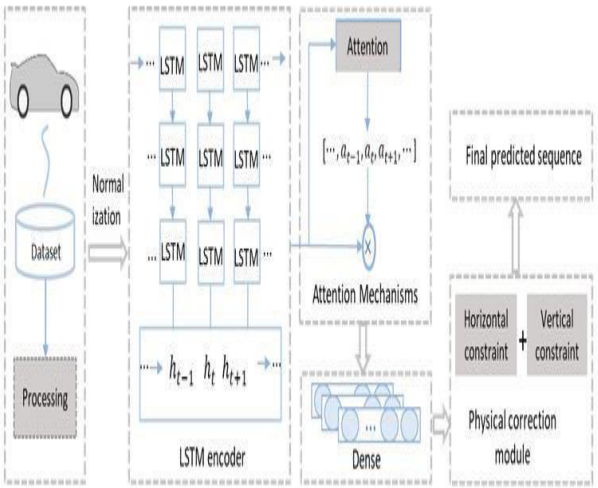

Fig1. Fusion Attention and Physics Correction Long Short-Term Memory Network

A. Multi-source data preprocessing

The system state data obtained from autonomous vehicles requires data processing. The primary objectives of data preprocessing are noise elimination, temporal alignment, and ensuring physical consistency. Raw data often contains significant noise, missing values, and temporal misalignment caused by sensor delays, which directly affect model training and prediction accuracy. Therefore, data cleaning and anomaly handling are essential.As the foundational module of the framework, data preprocessing standardizes raw heterogeneous data to provide reliable input for subsequent analysis. The processing methods are as follows:

Obvious outliers are excluded. For control command data, a threshold filtering method based on physical limits is applied. According to the performance boundaries of vehicle powertrain systems proposed in Ref.[10], the maximum capability of motors and braking systems typically does not exceed these ranges. Therefore, abnormal acceleration commands A_c exceeding the threshold of 3 m/s² are identified as communication interference and removed. For steering commands S_c, a saturation limit strategy is implemented based on the mechanical constraints of the steering mechanism, where the corresponding steering angle is restricted to ≤0.6 rad.

Format-invalid data are discarded. Format errors refer to cases where data does not conform to predefined type requirements. When a dataset should contain numerical values, non-numerical entries are considered invalid. For example, vehicle position data (x_a, y_a) should be time-series numerical values, and each sample must be convertible to a floating-point number. Thus, invalid entries such as "N/A" are removed.

Missing data are imputed. A dual-strategy approach is adopted for missing data: (i) Incomplete records are removed when the missing rate is low and temporal continuity is critical; (ii) Statistical imputation techniques are applied to reconstruct missing values based on observed data. Mean imputation is selected as the primary method due to its ability to maintain parameter distribution stability. Cross-validation is performed to ensure the physical plausibility of imputed values.

Temporal synchronization is calibrated. To address timing misalignment between control signals (e.g. A_c, D_c, S_c) and state parameters (e.g. v_x, τ), signal synchronization optimization is conducted. Considering the inherent CAN bus transmission delay (~50 ms), a time-offset compensation mechanism is introduced. Specifically, the velocity v_x at time t is mapped to the acceleration command A_c at time t-5, based on vehicle dynamics analysis, where the current motion state is essentially the cumulative effect of prior control commands.

B. LSTM encoder

As an important derivative architecture of recurrent neural networks (RNN) [11], the core innovation of LSTM is the introduction of a gating mechanism, which effectively mitigates the problems of gradient vanishing and gradient explosion that are prevalent in the training process of traditional RNN by dynamically adjusting the direction of information flow. As shown in Fig. 2, the LSTM unit consists of three key gate structures-input gate i, oblivion gate f, and output gate o. These gating units work in concert to achieve fine-grained control of the memory state. At each time step t of the temporal data processing, the LSTM updates the internal state through the following mathematical process:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$
$$\check{C}_t = tanh\,(W_c \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \check{C}_t \quad (4)$$
$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (5)$$
$$h_t = o_t \cdot tanh\,(C_t) \quad (6)$$

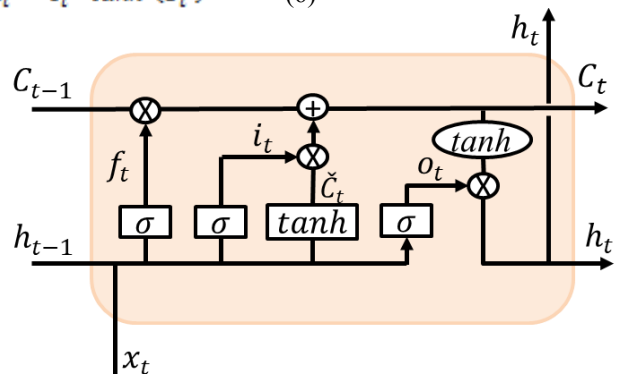

Fig. 2 Schematic diagram of LSTM network

In Eqs. (1) to (6), f_t,i_t,o_t denote the output features of the forgetting gate, the input gate, and the output gate at time t, respectively; $\check{C}$ is the candidate memory state feature; W_f, W_i, W_c,W_o are the weight matrices of the forgetting gate, the input gate, the coupled forgetting gate, and the output gate, respectively; h_{t-1} is the hidden state feature at time $t-1$; $\sigma(\cdot)$ is the $sigmoid$ activation function; b_f, b_i, b_C, and b_o are the bias coefficients of the corresponding gating structure. sigmoid activation function; b_f,b_i,b_c,b_o are the bias coefficients of the corresponding gating structures, respectively. The FAPC-LSTM proposed in this paper uses an LSTM network to encode the vehicle history state information to extract the features, and after the extraction, the network computes to obtain the output of the hidden layer at time t, h_t.

C. Attention mechanisms

The attention mechanism was first identified in the 1990s as a signal processing mechanism in human vision research. It is implemented as a specialized structure in machine learning models, primarily used to automatically learn and compute the influence of input data on output data [12].When integrated into LSTM models, the attention mechanism enables the system to focus on more valuable information while disregarding irrelevant data. As the core module for temporal feature selection, the attention mechanism processes the temporal feature matrix $h_t \in R^{T\times d}$ output from the LSTM encoder, where T represents the timesteps and d denotes the hidden layer dimension.First, the input features are linearly transformed through learnable weight matrices W_Q and W_K to generate query vector Q_t and key vector K_t:

$$Q_t = h_t W_Q, \quad K_t = h_t W_K \quad (7)$$

Subsequently, the attention coefficient matrix α_t for time step t is computed by scaling the dot product attention, and its computation process introduces a normalization factor $\sqrt{d}$ to mitigate the gradient vanishing problem:

$$\alpha_t = \text{softmax}(\frac{Q_t K_t^T}{\sqrt{d}}) \quad (8)$$

This mechanism can adaptively focus on historical features that are closely related to the current dynamic response. For example, in emergency braking scenarios, attention coefficients are significantly enhanced for longitudinal velocity v_x within the short time window before brake commands take effect. This captures the critical transition process from brake force transmission to tire slip ratio changes. Similarly, when the steering angle τ reaches its mechanical limit τ_{max}, the attention mechanism strengthens the temporal correlation between lateral velocity v_y and yaw angle Ψ during steering saturation. This reflects the dynamic characteristics under the kinematic constraints of the steering mechanism. Finally, the weighted context vector C_t is passed to downstream networks through feature fusion along with the current hidden state h_t:

$$C_t = \sum_{j=1}^{n} \alpha_{t,j}\, h_{t,j} \quad (9)$$

Further through the fully connected layer, obtain the preliminary predicted state vector$\check{Y} = \{\check{v}_x, \check{v}_y, \check{\Psi}, \check{x}_a, \check{y}_a, \ldots\}$

D. Physical layer correction

In the FAPC-LSTM model, the physics correction layer is designed to combine vehicle dynamics principles with data-driven prediction results. Physical consistency of predictions is achieved through explicit kinematic constraints.The layer takes the preliminary predicted state vector $\check{x}(t+1)$ from the fully-connected layer as input. Longitudinal velocity v_x and yaw angle Ψ are dynamically adjusted to output physically plausible corrected states. The detailed design is described below:

For vehicle longitudinal velocity v_x , longitudinal acceleration $\dot{v}_x$ is determined by both acceleration and braking commands. According to Newton's second law, the longitudinal force balance equation is expressed as:

$$m\dot{v}_x = F_x - F_f \quad (10)$$
$$F_x = \eta(A_c - D_c) \quad (11)$$

Here, m represents the vehicle mass, and F_f denotes the resistance force, which includes rolling resistance, air resistance, and grade resistance. The resistance force is related to external factors such as friction coefficients and road gradients, and the corresponding parameters can be adjusted according to actual driving conditions. F_x represents the net driving force, which is determined by both the acceleration command A_c and braking command D_c, with η denoting the powertrain efficiency. The neural network's predicted longitudinal velocity $\check{v}_x$ is corrected to $v_{x,cor}$ as follows:

$$v_{x,cor} = \check{v}_x + \alpha \cdot (\frac{F_x - F_f}{m} \cdot \Delta t) \quad (12)$$

In eq.(12), $\alpha\epsilon[0,1]$ represents a learnable parameter that automatically adjusts the strength of physical compensation during training, with Δt denoting the timestep duration.

The yaw angle Ψ: The yaw angle is influenced by both steering mechanism kinematics and tire lateral forces. According to the single-track model [13], the yaw rate $\dot{\Psi}$is primarily determined by the front wheel steering angle τ and longitudinal velocity $\check{v}_x$:

$$\dot{\Psi} = \frac{\check{v}_x \tan(\tau)}{L} \quad (13)$$

Among them, L is the wheelbase of the vehicle, which can be adjusted according to the vehicle's own conditions. Thus, the yaw angle $\check{\Psi}$ output by the neural network is corrected to eq(14)

$$\Psi_{cor} = \check{\Psi} + \gamma(\frac{\check{v}_x \tan(\tau)}{L} \cdot \Delta t) \quad (14)$$

In eq.(14), $\gamma\epsilon[0,1]$is a learnable parameter used to balance data-driven prediction and physical constraints.

III. TESTS AND ANALYSIS OF RESULTS

A. Dataset and Implementation Details

The nuScenes dataset [14], a large-scale autonomous driving public dataset released by Motional, was used in this study. The data collection vehicles were equipped with multiple sensors, including cameras, LiDAR, millimeter-wave radar, inertial measurement units (IMU), and GPS. A total of 15 hours (242 km) of driving data were collected in Singapore and Boston, which are known for their dense traffic and complex scenarios. The traffic scenarios were divided into 1000 segments, with each segment lasting 20 seconds at a sampling frequency of 2 Hz. For the proposed prediction model, vehicle trajectories were predicted for 6 seconds based on 2 seconds of historical data. The experiments were conducted on a computer equipped with an NVIDIA RTX3060 GPU. The PyTorch deep learning framework was employed, with the Adam optimizer used during training. The learning rate was set to 0.0001.

B. Evaluation indicators

The following metrics were adopted to evaluate the prediction model:

(1) Mean Absolute Error (MAE). The MAE measures model accuracy by calculating the average absolute deviation between predicted and true values:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|x_{cor,i} - x_{act,i}| \qquad (15)$$

In eq.(15), $x_{cor,i}$ represents the predicted value of the i-th data point, $x_{act,i}$ denotes the corresponding true value, and N is the sample size.

Due to its linearity, MAE shows strong robustness against outliers and directly reflects the overall prediction deviation level. In vehicle state prediction tasks, MAE is commonly used to evaluate control commands because its uniform error penalty aligns with real-world stability requirements, preventing evaluation distortion from extreme errors.

(2) Root Mean Square Error (RMSE). The RMSE is computed as the square root of the average squared errors:

$$\mathrm{RMAE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{cor,i} - x_{act,i})^2} \qquad (16)$$

The squared term amplifies larger errors, making RMSE more sensitive to significant deviations. In autonomous driving systems, RMSE is typically applied to safety-critical parameters. Its error amplification effect helps identify potentially hazardous states and prioritizes high-risk error reduction during model optimization.

(3) Cumulative Horizon Error (CHE). The CHE assesses error propagation in long-term predictions by summing errors across all time steps in a prediction window:

$$\mathrm{CHE} = \sum_{t=1}^{T} MAE\,(t) \qquad (17)$$

This metric directly indicates model stability over extended periods, particularly useful for long-term state estimation. Diverging errors (e.g., increasing speed errors during emergency braking) lead to significantly higher CHE values. By analyzing CHE's temporal distribution, model improvements can focus on enhancing long-term temporal dependency capture.

Analysis of results

To comprehensively evaluate the model's prediction capability, comparative experiments were conducted on a test set covering complex road conditions including urban streets and highways. Three models were compared: a pure physics-based model (Kalman filter + single-track model) [15], a pure data-driven model (traditional LSTM) [16], and the proposed FAPC-LSTM model.

For vehicle state prediction tasks, the prediction errors of state vector $x(t)$ were calculated in three groups based on physical dimension consistency and functional module relevance principles. This grouping was implemented to enhance the interpretability and practicality of evaluation metrics. The performance comparison of these grouped metrics, obtained through testing with trained models, is presented in Table 3-1.

Table 1 Evaluation of form state prediction results for different models

	unit	Evaluation metrics	physical model	LSTM	FAPC-LSTM
v_x, v_y	**m/s**	MAE	0.25	0.18	**0.12**
		RMSE	0.38	0.27	**0.26**
		CHE (5s)	6.42	4.25	**2.15**
Ψ	**rad**	MAE	0.12	0.08	**0.05**
		RMSE	0.20	0.14	**0.09**
		CHE (5s)	1.58	1.02	**0.48**
x_a, y_a	**m**	MAE	0.45	0.28	**0.18**
		RMSE	0.67	0.42	**0.25**
		CHE (5s)	8.95	5.63	**3.12**

The experimental results show that the pure physics-based model achieved significantly higher CHE (1.82) in the control command group compared to FAPC-LSTM (0.58), indicating its limited capability in temporal modeling of discrete commands. The FAPC-LSTM demonstrates superior performance over both the traditional LSTM and pure physics-based models across all prediction groups.

Fig.3 presents the cumulative error progression of lateral velocity v_x and longitudinal velocity v_y predictions over a 5-second horizon for the three models. The x-axis represents prediction duration while the y-axis shows cumulative error. The blue and red curves correspond to the traditional LSTM and pure physics-based model respectively, with the green curve representing FAPC-LSTM. The pure physics-based model shows smaller initial errors but exhibits significant error growth after 3 seconds. Similarly, the traditional LSTM demonstrates poor stability in long-term predictions. In contrast, the FAPC-LSTM maintains the most gradual error growth in later stages, outperforming both comparison models.

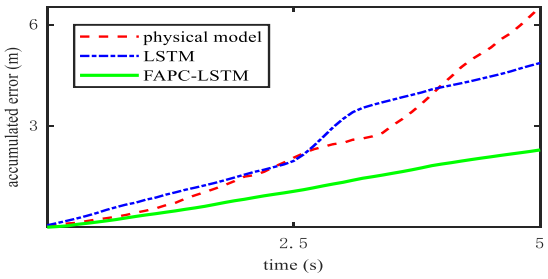

Fig. 3 Comparison of cumulative errors of the three models

The experimental results demonstrate that the FAPC-LSTM model achieves better accuracy and stability in vehicle state prediction tasks. The model effectively captures dynamic changes in vehicle state characteristics while avoiding physically unrealistic predictions that may occur with traditional models. Tests show that FAPC-LSTM maintains

superior stability in long-term predictions, with significantly slower error growth compared to other models. This stability stems from the dual assurance of combining data-driven learning with physical constraints, enabling more reliable trajectory predictions for autonomous driving systems in complex scenarios.

CONCLUSION

The FAPC-LSTM model is proposed to address the limitations of traditional vehicle trajectory prediction methods in terms of accuracy and physical consistency. By integrating data-driven learning with physical constraints, the model is shown to outperform both pure physics-based models and traditional LSTM across multiple evaluation metrics (MAE, RMSE, CHE), particularly demonstrating more gradual error growth in long-term predictions. The attention mechanism enhances the model's ability to capture critical temporal features, while the physics correction layer ensures predictions comply with vehicle dynamics principles. The superior performance of FAPC-LSTM provides reliable support for safety-critical decision making in autonomous driving systems. Future research could further optimize parameter adaptability and explore generalization capabilities in more scenarios.

REFERENCES

[1] Stéphanie Lefèvre, Vasquez D, et al. A survey on motion prediction and risk assessment for intelligent vehicles[J]. Robomech Journal, 2014, 1(1):1-14.

[2] Rajamani R . Vehicle Dynamics and Control[J]. Journal of the japan fluid power system society, 2012, 39(8):111.

[3] Wang P, Wang J, Chan C Y, et al. Trajectory Prediction for Turning Vehicles at Intersections by Fusing Vehicle Dynamics and Driver's Future Input Estimation[J]. Transportation Research Record Journal of the Transportation Research Board, 2016, 2602:68-77.

[4] Barth A, Frabke U. Where will the Oncoming Vehicle Be the next Second[C]. New York: 2008 IEEE Intelligent Vehicle Symposium, IEEE ,2008: 1068-1073.

[5] Houenou A, Bonnifaitp, Cherfaouiv, et al. Vehicle Trajectory Prediction Based on Motion Model and Maneuver Recognition[C]. New York: IEEE,2013 RSJ International Conference on Intelligent Robots and Systems, IEEE ,2013: 4363-4369.

[6] Tomar RS, Verma S, Tomar GS. Prediction of Lane Change Trajectories through Neural Network[C]. New York: 2010 International Conference on Computational Intelligence and Communication Networks(CICN). IEEE, 2010: 249-253.

[7] Liu Y R, Meng Q Y, Guo H Y, et al. Surrounding vehicle trajectory prediction with high-definition map integration using graph attention mechanism[J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(3).

[8] Altché F, Fortelle A. An LSTM network for highway trajectory prediction[C]. New York: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2017:353-359.

[9] Jia P Y, Chen P H, Zhang L, et al 2022 Attention-LSTM based prediction model for aircraft 4-D trajectory. Scientific reports. 12 15533

[10] Zhang T, He X H, Wang Q, et al. Performance study of energy recovery system for wheel-side driven hydraulic hybrid vehicles[J]. Chinese Hydraulics & Pneumatics, 2019(10):90-96.

[11] Wen H, Gao Y. Dual-path RNN with self-attention for automotive radar interference mitigation[J]. Radar Science and Technology, 2022, 20(6):678-687.

[12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[M]. Advances in neural information processing systems, 2017.

[13] Fang Y K, Min H G, et al 2020 A fault detection and diagnosis system for autonomous vehicles based on hybrid approaches. *IEEE Sensors Journal.***20** 9359-9371.

[14] Caesar H, Bankiti V, Lang A H, et al. Nuscenes: a multi-modal dataset for autonomous driving[C] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 11621-11631.

[15] Yang F P. A nonlinear Kalman filtering method for vehicle state measurement and trajectory prediction[J]. Science & Technology Vision, 2024, 14(28):35-37.

[16] Zhang M, He Y J, Yang B Y, et al. High-speed train group trajectory prediction method based on LSTM-KF model[J]. Journal of Traffic and Transportation Engineering, 2024, 24(3):296-310.