Physics-Informed Trajectory Prediction for Autonomous Driving under Missing Observation

Haicheng Liao[1] , **Chengyue Wang**[1] , **Zhenning Li**[1] , **Yongkang Li**[2] , **Bonan Wang**[1] , **Guofa Li**[3] , **Cheng-Zhong Xu**[1]
[1]University of Macau
[2]UESTC
[3]Chongqing University
{yc27979, chengyuewang, zhenningli, mc35002, czxu}@um.edu.com,
franklin1234560@163.com, hanshan198@gmail.com

Abstract

This paper introduces a novel trajectory prediction approach for autonomous vehicles (AVs), adeptly addressing the challenges of missing observations and the need for adherence to physical laws in real-world driving environments. This study proposes a hierarchical two-stage trajectory prediction model for AVs. In the first stage we propose the Wavelet Reconstruction Network, an innovative tool expertly crafted for reconstructing missing observations, offering optional integration with state-of-the-art models to enhance their robustness. Additionally, the second stage of the model features the Wave Fusion Encoder, a quantum mechanics-inspired innovation for sophisticated vehicle interaction modeling. By incorporating the Kinematic Bicycle Model, we ensure that our predictions align with realistic vehicular kinematics. Complementing our methodological advancements, we introduce MoCAD-missing, a comprehensive real-world traffic dataset, alongside enhanced versions of the NGSIM and HighD datasets, designed to facilitate rigorous testing in environments with missing observations. Extensive evaluations demonstrate that our approach markedly outperforms existing methods, achieving high accuracy even in scenarios with up to 75% missing observations.

1 Introduction

In the field of AVs, the effective prediction of surrounding vehicles' trajectories in dynamic environments encounters two often overlooked yet interrelated challenges. The first challenge arises from real-world observation constraints, such as sensor limitations and environmental factors including obstructions, adverse weather, or traffic congestion. These constraints frequently result in missing observations [Liao *et al.*, 2024b], a scenario that poses a significant hurdle for traditional deep learning models. Despite their effectiveness with ideal datasets, these models typically struggle to adapt to the unpredictable and variable conditions characteristic of real-world driving [Bhattacharyya *et al.*, 2023], an issue that has not been sufficiently addressed in previous studies.

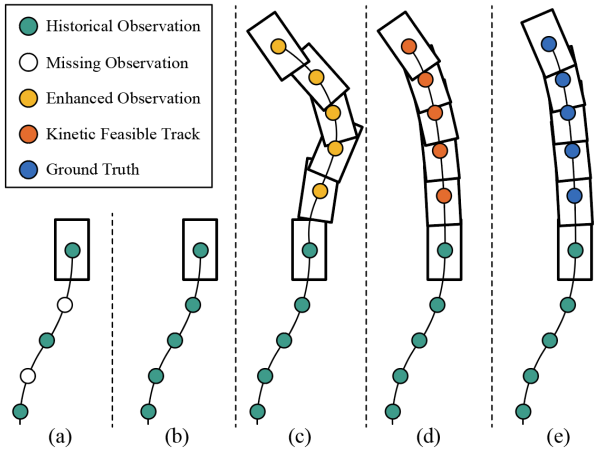

Figure 1: Physics Enhancement Stage. (a) The input of Wavelet Reconstruction Network — incomplete historical observations. (b) The reconstructed complete historical observations. (c) Enhanced observations for immediate future by Wavelet Reconstruction Network. (d) Kinematic feasible trajectory regularized by Kinematic Bicycle Model. (e) The ground truth trajectory for immediate future.

Second, equally overlooked challenge involves ensuring that these models adhere to physical laws in trajectory prediction. Many current models inadequately consider the kinematic constraints of vehicular motion, leading to predictions that are statistically accurate but kinematic infeasible. This limitation can potentially compromise the safety and reliability of AVs' motion plans [Huang *et al.*, 2022; Shen *et al.*, 2023], yet it has not been given due attention in the existing literature.

To bridge these gaps, our study introduces a novel two-stage trajectory prediction approach that synergizes data-driven and physics-based methodologies with a Physics Enhancement Stage, as shown in Figure 1, and a Trajectory Prediction Stage. This approach combines the robustness of deep learning with physics-informed principles, ensuring realistic and robust trajectory predictions even in scenarios with missing observations. Demonstrating its effectiveness, our approach consistently outperforms state-of-the-art (SOTA)

models in comprehensive evaluations, even with up to 75% missing observations. The key contributions of our study can be summarized as follows:

- We introduce a pioneering Physics Enhancement Stage which includes a Wavelet Reconstruction Network and a Kinematic Bicycle Model. This integration fosters a significant advancement in the trajectory prediction domain, markedly enhancing the approach's robustness against missing observations and augmenting the kinematic feasibility of the predicted trajectory.
- We significantly enrich trajectory prediction research by introducing MoCAD-missing, a dataset curated from a fully (Level 5) AV in real-world traffic scenarios, and by enhancing the established NGSIM and HighD datasets with scenarios featuring missing observations. These contributions expand the research scope in this domain, providing robust testing grounds for advanced models.
- Our innovative Wave Fusion Encoder, inspired by quantum mechanics, revolutionizes interaction modeling. By conceptualizing vehicle features as waveforms, this module facilitates a novel approach to model vehicle interactions using the principle of wave superposition.
- We conduct comprehensive experiments on MoCAD-missing, NGSIM-missing, and HighD-missing datasets. Results show that i) Our approach can outperform other SOTA models across various scenarios even in the face of 75% missing observations, and ii) The wavelet reconstruction network can integrate seamlessly with other SOTA models to enhance the robustness against missing observations.

2 Related Work

Initially, the application of physics-based models played a significant role in the realm of trajectory prediction. Notably, various classical physics models, such as the Constant Velocity model [Miller and Huang, 2002] and the Bicycle model [Brännström *et al.*, 2010], garnered substantial attention due to their computational efficiency. Additionally, Kalman filtering is used in tracking algorithms to predict the vehicle's position [Kaempchen *et al.*, 2004]. Moreover, the application of Monte Carlo simulations, as exemplified in the works of Danielsson et al. [Danielsson *et al.*, 2007] offered a means of simulating future vehicle states.

However, while these physics-based models exhibit commendable performance in physical feasibility, they falter in long-term prediction accuracy, particularly in complex and dynamic traffic environments. This limitation has catalyzed the shift towards learning-based methodologies, aiming to enhance prediction accuracy. The introduction of Recurrent Neural Network (RNN) architectures, notably by the work [Zyner *et al.*, 2019; Zyner *et al.*, 2017; Liao *et al.*, 2024d; Chen *et al.*, 2022b], marked a significant advancement in this field. These models [Cong *et al.*, 2023; Zuo *et al.*, 2023; Liao *et al.*, 2024b], harnessed historical vehicle observations to enhance prediction accuracy. Subsequent enhancements in the trajectory prediction field included the application of Convolutional Neural Networks (CNNs), which significantly revolutionized the processing of contextual features. This was exemplified in the use of social tensors [Deo and Trivedi, 2018] and rasterized maps [Gilles *et al.*, 2021].

The learning-based methods mentioned above are purely data-driven methods that lack interpretability and will lead to some impractical trajectories with issues like excessive acceleration. Therefore, researchers have turned their attention to integrating physics knowledge into trajectory prediction. Drawing inspiration from the wave superposition principle in quantum mechanics [Arndt *et al.*, 1999], Wang et al. introduced Wave-MLP [Tang *et al.*, 2022] into the trajectory prediction model to replace the multi-head attention mechanism to obtain information about interactions between vehicles [Wang *et al.*, 2023]. However, their approach was limited to capturing only spatial interaction at the current moment, neglecting the temporal interaction. In [Cui *et al.*, 2020], Cui et al. used a neural network to learn the parameters required for a biaxial bicycle model, and then the biaxial bicycle model was used as the output layer, and the output was consistent with the vehicle kinematics to learn the trajectory. Zhang et al. [Zhang *et al.*, 2023] used wavelet transform theory to predict the trajectory of the aircraft.

3 Problem Formulation

The main objective of this study is to predict the trajectory of the *target vehicle*, referring to the selected vehicle that lies within the sensing range of the AV. At each time t_c, the AV aims to predict the target vehicle's trajectory $\boldsymbol{Y}_0$ for future t_f time based on the historical observations $\boldsymbol{X}$. Our proposed method consists of two stages: trajectory repair and trajectory prediction. It uses historical observations from time $t_c - t_h$ to t_c, including observations for the target vehicle $\boldsymbol{X}_0$ and all its observed surrounding vehicles $\boldsymbol{X}_{1:n}$, as input. To be more consistent with real-world, the input has missing observations. Formally, our approach could be formulated as:

$$\boldsymbol{Y}_0 = F\left(K(\boldsymbol{X})\right) \tag{1}$$

where the input observations $\boldsymbol{X} = \{\boldsymbol{X}_0, \boldsymbol{X}_{1:n}\}$ consist of the position coordinates, speed, acceleration, vehicle type, and lane number of the target vehicle and all its surrounding vehicles. Correspondingly, F denotes the trajectory prediction stage, while K denotes the physics enhancement stage.

Additionally, we introduce a multi-modal decoder with a hierarchical hierarchical multi-modal prediction framework to model the uncertainty in the driving behavior of surrounding human drivers. At first, the uncertainty of human drivers is reflected in the variety of potential maneuvers. In the real-world, the longitudinal maneuvers are divided into acceleration, deceleration and constant speed, and the lateral maneuvers are divided into left lane change and right lane change, and keep constant. In this framework, the possibility $P\left(M_i|\boldsymbol{X}\right)$ of each maneuver M_i is first estimated based on the historical observations of the target vehicle and surrounding vehicles. And then, the human driver's uncertainty is reflected in the trajectory for each potential maneuver, so the trajectory is modeled as a Bivariate Gaussian distribution. Specifically, at time t_c+f the prediction is characterized with means $\mu_x^{t_c+f}$ and $\mu_y^{t_c+f}$, variances ${\sigma_x^{t_c+f}}^2$ and ${\sigma_y^{t_c+f}}^2$, and

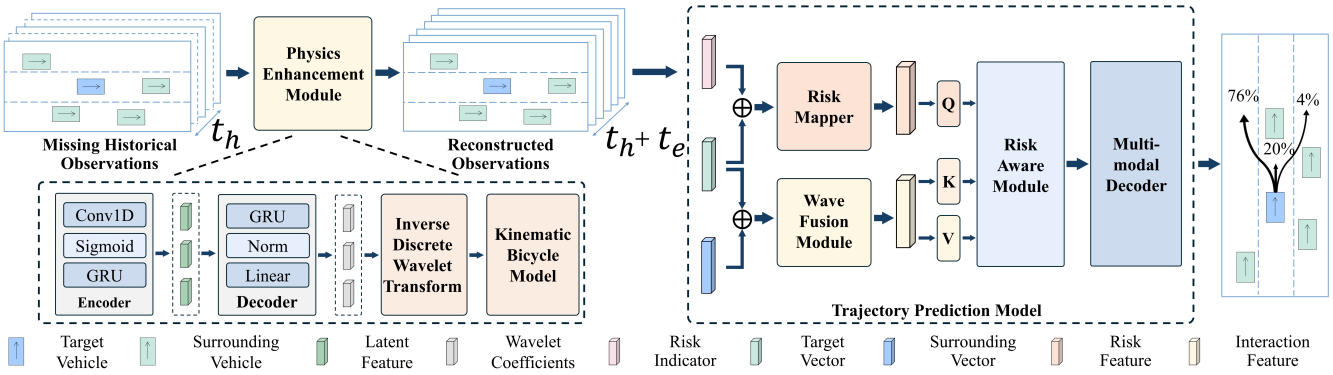

Figure 2: The framework of the proposed approach.

the correlation coefficient ρ_{t_c+f}.

$$P(\boldsymbol{Y}_0|\boldsymbol{X}) = \sum_{\forall i} P\left(M_i|\boldsymbol{X}\right) P\left(\boldsymbol{Y}_0|M_i, \boldsymbol{X}\right) \tag{2}$$

4 Methodology

As shown in Figure 2, our approach is divided into two main stages, including the physics enhancement stage and the trajectory prediction stage. The physics enhancement stage is designed to enhance missing historical observations and provide a robust foundation. Moreover, the trajectory prediction stage makes multimodal predictions based on these observations.

4.1 Stage 1: Physics Enhancement

In this stage, we enhance trajectory prediction accuracy by addressing the challenge of missing observations. We utilize the wavelet reconstruction network to restore the full spectrum of historical observations, thereby providing a robust foundation for predicting future trajectories. Furthermore, we prompt the trajectory prediction model by inputting a kinematic feasible immediate future trajectory, thereby improving the kinematic feasibility of the final output trajectory. To this end, the wavelet reconstruction network is also used to enhance immediate future observations. Then, the kinematic model regularizes the enhanced observations to generate the kinematic feasible trajectory.

Wavelet Reconstruction Network

This network includes a pair of reversible operations: Discrete Wavelet Transform (**DWT**) decomposes the original observations into wavelet coefficients, and Inverse Discrete Wavelet Transform (**IDWT**) uses the decomposed wavelet coefficients to reconstruct the original observations. However, in the evaluation stage we only have missing historical observations. In order to reconstruct complete historical observations, IDWT cannot use the wavelet coefficients decomposed by DWT from missing historical observations. To address this, a wavelet coefficient extractor is used to extract the wavelet coefficients corresponding to complete historical observations from missing historical observations. The DWT and complete historical observations are used to supervise the training of the wavelet coefficient extractor and the detailed training strategy is shown in Figure 3.

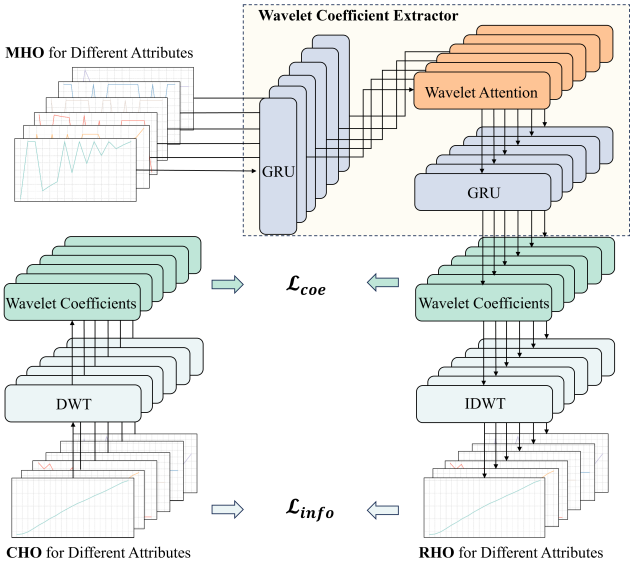

Figure 3: The training strategy of Wavelet Coefficient Extractor. MHO stands for missing historical observations, CHO stands for complete historical observations, RHO stands for reconstructed historical observations.

Specifically, we use Level-3 DWT and its corresponding IDWT in this network. For clarity and consistency, historical longitudinal coordinates observations $\boldsymbol{x}_i$ serving as illustrative examples for subsequent analysis, and the complete longitudinal coordinates observations are denoted as $\hat{\boldsymbol{x}}_i$. Level-3 DWT decomposes the historical longitudinal coordinates observations into detail $c_i(j,k)$ and approximation coefficients $d_i(j,k)$ at different level j and translation k. As shown in Figure 4, this decomposition is achieved through an iterative process involving high-pass filter h and low-pass filter g, followed by downsampling.

After obtaining the $c_i(j,k)$ and $d_i(j,k)$ of the $\hat{\boldsymbol{x}}_i$, the IDWT is employed to reconstruct the $\hat{\boldsymbol{x}}_i$:

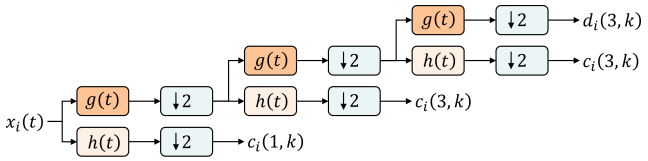

Figure 4: Illustration of the Level-3 Discrete Wavelet Transform process. $\downarrow 2$ is the downsampling operation.

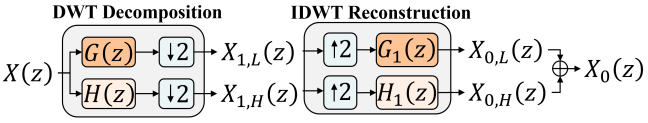

Figure 5: The process encompasses pipeline in z-domain from the decomposition of historical longitudinal coordinates observations using Discrete Wavelet Transform to their reconstruction using Inverse Discrete Wavelet Transform. $\uparrow 2$ is the interpolation operation.

$$\hat{\boldsymbol{x}}_i = \sum_j \sum_k c_i(j,k) \cdot 2^{j/2} \cdot h\left(2^{-j}t - k\right) + \sum_j \sum_k d_i(j,k) \cdot 2^{j/2} \cdot g\left(2^{-j}t - k\right) \tag{3}$$

Theorem 1. *Complete preservation of observations in DWT decomposition and IDWT reconstruction.*

Theorem 1 is the cornerstone of the wavelet reconstruction network, which ensures that the observations reconstructed by IDWT contain the original observations. Following the procedure in Figure 5, we will give a proof for this theorem.

Proof. Apply the Z-transformation to transform the time-domain observations of $\hat{\boldsymbol{x}}_i(t)$ into its z-domain representation $X(z) = \sum \hat{\boldsymbol{x}}_i(t) z^{-t}$. Following [Goel, 2014], Level-1 DWT decomposes $X(z)$ to obtain the approximate $X_{1,L}(z)$ and detailed coefficients $X_{1,H}(z)$ in the z-domain:

$$X_{1,L}(z) = \frac{1}{2}\left[X\left(z^{\frac{1}{2}}\right) G\left(z^{\frac{1}{2}}\right) + X\left(-z^{\frac{1}{2}}\right) G\left(-z^{\frac{1}{2}}\right)\right] \tag{4}$$

$$X_{1,H}(z) = \frac{1}{2}\left[X\left(z^{\frac{1}{2}}\right) H\left(z^{\frac{1}{2}}\right) + X\left(-z^{\frac{1}{2}}\right) H\left(-z^{\frac{1}{2}}\right)\right] \tag{5}$$

where $G(z)$ and $H(z)$ are the low-pass filter and high-pass filter in z-domain.

In the IDWT, the interpolation is first used to restore the original frequency and then passes the low-pass reconstruction filter $G_1(z)$ and the high-pass reconstruction filter $H_1(z)$. By combining $X_{0,L}(z)$ and $X_{0,H}(z)$, the reconstructed observations in z-domain $X_0(z)$ could be obtain as:

$$\begin{aligned} X_0(z) &= X_{0,L}(z) + X_{0,H}(z) \\ &= \frac{1}{2}\left[G(z)G_1(z) + H(z)H_1(z)\right] X(z) + \\ &\quad \frac{1}{2}\left[G(-z)G_1(z) + H(-z)H_1(z)\right] X(-z) \end{aligned} \tag{6}$$

By analyzing Eq. 6, we can conclude that the relationship between $X_0(z)$ and $X(z)$ is determined by the filters. We use the Haar wavelet and its corresponding filters from PyWavelet [Lee *et al.*, 2019], and in the Z-domain:

$$\begin{aligned} H(z) &= 0.7071067811865476 + 0.7071067811865476 \cdot z^{-1} \\ G(z) &= -0.7071067811865476 + 0.7071067811865476 \cdot z^{-1} \\ H_1(z) &= 0.7071067811865476 + 0.7071067811865476 \cdot z^{-1} \\ G_1(z) &= 0.7071067811865476 - 0.7071067811865476 \cdot z^{-1} \end{aligned} \tag{7}$$

Substituting the values from Eq. 7 into Eq. 6 we get:

$$X_0(z) = z^{-1} X(z) \tag{8}$$

Eq. 8 indicates that the reconstructed observations is a unit-delay version of the original observations, where z^{-1} represents one unit of delay. In the Z-domain, z^{-1} represents a unit delay, which does not lead to observations loss but merely indicates a temporal shift of the observations. □

The wavelet reconstruction network needs to provide immediate future observations for kinematic bicycle model to generate the kinematic feasible immediate future trajectory. Therefore, in the wavelet reconstruction network, DWT not only needs to decompose complete historical observations but also decompose immediate future observations to supervise the training of wavelet coefficients extractor, enabling this network could enhance immediate future observations.

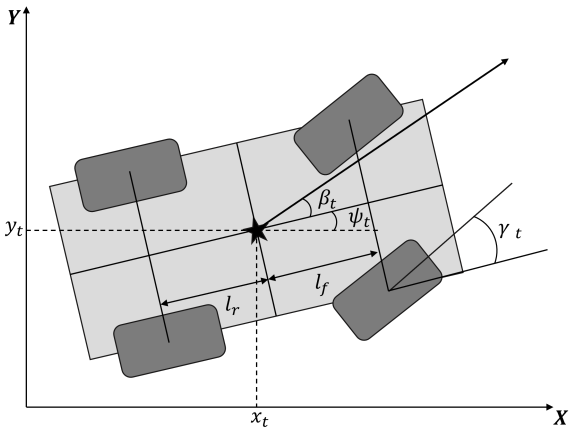

Figure 6: Illustration of various parameters in the bicycle model.

Kinematic Bicycle Model

A two-axis kinematic bicycle model, centered on the vehicle's center of mass, is employed to generate the trajectory for the immediate future t_e based on the observations enhanced by wavelet reconstruction network. As shown in Figure 6, critical variables including speed v^t, acceleration a^t, steering angle γ^t, coordinates (x^t, y^t), heading ψ^t, as well as the distances from the center of mass to the front l_f and rear l_r wheels at time t.

The vehicle's state at the next time $t+1$ is determined using state transition equations, derived from the kinematic bicycle model detailed by Cui et al. [Cui *et al.*, 2020]:

$$\begin{aligned} x^{t+1} &= x^t + \dot{x}^t \Delta t, \quad y^{t+1} = y^t + \dot{y}^t \Delta t \\ \psi^{t+1} &= \psi^t + \dot{\psi}^t \Delta t, \quad v^{t+1} = v^t + \dot{v}^t \Delta t \end{aligned} \tag{9}$$

where Δt denotes the time difference between consecutive times, while the derivatives $[\dot{x}^t, \dot{y}^t, \dot{\psi}^t, \dot{v}^t]$ at time t are calculated as per the following expressions:

$$\begin{aligned} \dot{x}^t &= v^t \cos\left(\psi^t + \beta^t\right), \quad \dot{\psi}^t = \frac{v^t}{l_r} \sin\left(\beta^t\right) \\ \dot{y}_i^t &= v^t \sin\left(\psi^t + \beta^t\right), \qquad\qquad \dot{v}_i^t = a^t \end{aligned} \tag{10}$$

where $\beta^t = \tan^{-1}\left(\frac{l_r}{l_f + l_r} \tan\gamma^t\right)$ signifies the angle between the vehicle's velocity vector and its heading direction.

4.2 Stage 2: Trajectory Prediction

Wave Fusion Encoder

Our module employs a conceptual framework from quantum mechanics wherein a vehicle's features are conceptualized as a wave. In this framework, the vehicle features are encoded within the amplitude of the wave, while the vehicle's temporal and spatial dynamics are encapsulated in the phase. Mathematically, this conceptual framework can be defined as:

$$\tilde{\boldsymbol{X}}_j = |\boldsymbol{X}_j| \odot e^{i\boldsymbol{\theta}_j}, j = 0, 1, 2, \cdots, n \tag{11}$$

where $\tilde{\boldsymbol{X}}_j$ is the wave representation of vehicle j. $|\boldsymbol{X}_j|$ is the amplitude learning from the vehicle features. The phase is denoted as $\boldsymbol{\theta}_j$. And $\odot$ is the element-wise multiplication, i is the imaginary unit.

With this conceptual framework, we can use the superposition of waves to model the effects of interactions between vehicles. Leveraging the superposition principle of wave, the interaction between two vehicles, $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$, can be rigorously expressed by the defined mathematical formula:

$$|\boldsymbol{X}_r| = \sqrt{|\boldsymbol{X}_j|^2 + |\boldsymbol{X}_k|^2 + 2\,|\boldsymbol{X}_j| \odot |\boldsymbol{X}_k| \odot \cos\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j\right)} \tag{12}$$

$$\begin{aligned} \boldsymbol{\theta}_r = &\boldsymbol{\theta}_j + \operatorname{atan2}\left(|\boldsymbol{X}_k| \odot \sin\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j\right),\right. \\ &\left.|\boldsymbol{X}_j| + |\boldsymbol{X}_k| \odot \cos\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j\right)\right) \end{aligned} \tag{13}$$

$$\tilde{\boldsymbol{X}}_r = |\boldsymbol{X}_r| \odot e^{i\boldsymbol{\theta}_r} \tag{14}$$

where $\tilde{\boldsymbol{X}}_r$ is the interaction features. By employing Euler expansion, interaction features could be expressed as:

$$\tilde{\boldsymbol{X}}_j = |\boldsymbol{X}_j| \odot \cos\boldsymbol{\theta}_j + i\,|\boldsymbol{X}_j| \odot \sin\boldsymbol{\theta}_j \tag{15}$$

For convenience of representation, the final output $\boldsymbol{o}_j$ is projected into the real number domain:

$$\boldsymbol{o}_j = \sum_k W_{jk}^a \boldsymbol{X}_k \odot \cos\boldsymbol{\theta}_k + W_{jk}^i \boldsymbol{X}_k \odot \sin\boldsymbol{\theta}_k \tag{16}$$

where W^a and W^i are learnable weights.

Risk Aware Module

The Risk Aware Module is architected with two core components: a risk encoder and a probsparse attention module. The former is tasked with the translation of the risk vector into the target vector space, thereby yielding the real-time risk features pertinent to the target vehicle. The risk vector, foundational to this module, is conceptualized as per Kloeden et al. [Kloeden *et al.*, 2001] and is quantified by the equation:

$$I_r = \exp\left(0.07039\Delta v + 0.0008617 v^2\right) \tag{17}$$

Algorithm 1 Risk Aware Algorithm

Input: $\boldsymbol{Q} \in \mathbb{R}^{L_R \times d}, \boldsymbol{K} \in \mathbb{R}^{L_W \times d}, \boldsymbol{V} \in \mathbb{R}^{L_W \times d}$, number of surrounding vehicles n
Output: Feature map F

1: **for** *Head* **in** *P Heads* **do**
2: $U = L_R \ln L_W$.
3: Randomly select U query-key pairs $\left[\boldsymbol{Q}\boldsymbol{K}^\top\right]_U$.
4: Compute the importance $\bar{D} = \max(\left[\boldsymbol{Q}\boldsymbol{K}^\top\right]_U) - \operatorname{mean}(\left[\boldsymbol{Q}\boldsymbol{K}^\top\right]_U)$ for each query.
5: Obtain the dominant score matrix $\overline{\boldsymbol{F}} \in \mathbb{R}^{n \times d}$ based on importance.
6: $\mathbf{F}_i = \text{Softmax}\left(\overline{\boldsymbol{F}}/\sqrt{d}\right) \cdot \mathbf{V}$.
7: **end for**
8: $\mathbf{F} = \textit{Concat}[\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_P]$.
9: **return F**

where I_r is the injury crash rate, v is the speed of the target vehicle, and Δv is the difference between the target vehicle speed and average traffic speed.

Next, the probsparse attention module is employed in this module to integrate real-time risk considerations into the dynamics between the target vehicle and its surrounding vehicles. It operates under the premise that the target vehicle's interactions are predominantly influenced by surrounding vehicles within a specific proximity. This notion resonates with the findings of Zhou et al. [Zhou *et al.*, 2021], which articulate that only a limited array of salient query-key pairs substantially contribute to the computation of attention. Based on this finding, probsparse attention module is devised to specifically accentuate the interactions among surrounding vehicles. The operation of this module is detailed in Algorithm 1.

5 Experiments

5.1 Experiment Setup

The performance of our approach is evaluated using three well-regarded real-world trajectory prediction datasets: Macao Connected Autonomous Driving (MoCAD) [Liao *et al.*, 2024b], NGSIM [Deo and Trivedi, 2018], and HighD [Krajewski *et al.*, 2018]. These datasets provide longitudinal and lateral coordinates of traffic agents, which are systematically divided into training, validation, and test sets. The trajectory data from these datasets are segmented into 8-second intervals. The first 3 seconds are used as trajectory history for model input ($t_h = 3$), and the following 5 seconds are used as ground truth for model prediction ($t_f = 5$). This segmentation enables a comprehensive evaluation of various traffic scenes, including urban streets, campuses, and highways. To measure prediction accuracy, the Root Mean Square Error (RMSE) is used as the evaluation metric, where lower RMSE values indicate higher prediction accuracy. Additionally, the immediate future t_e for the kinematic bicycle model in the physics enhancement module is set to 1 second. The model is trained on a single Nvidia A40 48GB GPU to convergence, using a learning rate of 0.0005 and a batch size of 64. We employ Mean Square Error (MSE) and Negative Log Likelihood (NLL) as loss functions.

Dataset	Model	Prediction Horizon (s)				
		1	2	3	4	5
MoCAD	S-GAN [Gupta *et al.*, 2018]	1.69	2.25	3.30	3.89	4.69
	CS-LSTM [Deo and Trivedi, 2018]	1.45	1.98	2.94	3.56	4.49
	MHA-LSTM [Messaoud *et al.*, 2021]	1.25	1.48	2.57	3.22	4.20
	NLS-LSTM [Messaoud *et al.*, 2019]	0.96	1.27	2.08	2.86	3.93
	WSiP [Wang *et al.*, 2023]	0.70	0.87	1.70	2.56	3.47
	CF-LSTM [Xie *et al.*, 2021]	0.72	0.91	1.73	2.59	3.44
	STDAN [Chen *et al.*, 2022b]	0.62	0.85	1.62	2.51	3.32
	BAT (25%) [Liao *et al.*, 2024b]	0.65	0.99	1.89	2.81	3.58
	BAT [Liao *et al.*, 2024b]	0.35	0.74	1.39	2.19	2.88
	HLTP [Liao *et al.*, 2024a]	0.55	0.76	1.44	2.39	3.21
	Ours (25%)	**0.28**	**0.47**	1.08	1.84	**2.70**
	Ours (50%)	0.40	0.63	**0.99**	**1.72**	2.74
	Ours (75%)	0.32	0.67	1.33	2.20	2.76
NGSIM	S-GAN [Gupta *et al.*, 2018]	0.57	1.32	2.22	3.26	4.40
	CS-LSTM [Deo and Trivedi, 2018]	0.61	1.27	2.09	3.10	4.37
	MATF-GAN [Zhao *et al.*, 2019]	0.66	1.34	2.08	2.97	4.13
	NLS-LSTM [Messaoud *et al.*, 2019]	0.56	1.22	2.02	3.03	4.30
	IMM-KF [Lefkopoulos *et al.*, 2020]	0.58	1.36	2.28	3.37	4.55
	WSiP [Wang *et al.*, 2023]	0.56	1.23	2.05	3.08	4.34
	CF-LSTM [Xie *et al.*, 2021]	0.55	1.10	1.78	2.73	3.82
	MHA-LSTM [Messaoud *et al.*, 2021]	0.41	1.01	1.74	2.67	3.83
	DRBP[Gao *et al.*, 2023]	1.18	2.83	4.22	5.82	-
	TS-GAN [Wang *et al.*, 2022]	0.60	1.24	1.95	2.78	3.72
	STDAN [Chen *et al.*, 2022b]	0.39	0.96	1.61	2.56	3.67
	iNATran [Chen *et al.*, 2022a]	0.39	0.96	1.61	2.42	3.43
	FHIF [Zuo *et al.*, 2023]	0.40	0.98	1.66	2.52	3.63
	DACR-AMTP [Cong *et al.*, 2023]	0.57	1.07	1.68	2.53	3.40
	BAT (25%) [Liao *et al.*, 2024b]	0.31	0.85	1.65	2.69	3.87
	BAT [Liao *et al.*, 2024b]	**0.23**	**0.81**	1.54	2.52	3.62
	GaVa [Liao *et al.*, 2024c]	0.40	0.94	1.52	2.24	3.13
	Ours (25%)	0.42	0.92	**1.45**	**2.07**	**2.89**
	Ours (50%)	0.42	0.93	**1.45**	2.08	2.90
	Ours (75%)	0.42	0.93	1.46	2.08	2.91
HighD	S-GAN [Gupta *et al.*, 2018]	0.30	0.78	1.46	2.34	3.41
	WSiP [Wang *et al.*, 2023]	0.20	0.60	1.21	2.07	3.14
	CS-LSTM(M) [Deo and Trivedi, 2018]	0.23	0.65	1.29	2.18	3.37
	CS-LSTM [Deo and Trivedi, 2018]	0.22	0.61	1.24	2.10	3.27
	MHA-LSTM [Messaoud *et al.*, 2021]	0.19	0.55	1.10	1.84	2.78
	NLS-LSTM [Messaoud *et al.*, 2019]	0.20	0.57	1.14	1.90	2.91
	DRBP[Gao *et al.*, 2023]	0.41	0.79	1.11	1.40	-
	EA-Net [Cai *et al.*, 2021]	0.15	0.26	0.43	0.78	1.32
	CF-LSTM [**?**]	0.18	0.42	1.07	1.72	2.44
	STDAN [Chen *et al.*, 2022b]	0.19	0.27	0.48	0.91	1.66
	iNATran [Chen *et al.*, 2022a]	**0.04**	**0.05**	**0.21**	0.54	1.10
	DACR-AMTP [Cong *et al.*, 2023]	0.10	0.17	0.31	0.54	1.01
	GaVa [Liao *et al.*, 2024c]	0.17	0.24	0.42	0.86	1.31
	Ours (25%)	0.13	0.24	0.36	**0.49**	0.73
	Ours (50%)	0.14	0.26	0.39	0.53	0.75
	Ours (75%)	0.14	0.24	0.37	0.50	**0.72**

Table 1: Evaluation results for our approach alongside baseline comparisons within the test datasets across varying temporal horizons. RMSE (m) serves as the metric for evaluation. Instances where values are not available are marked with a dash ("-"). **Bold** and underlined values represent the best and second-best performance.

Acknowledging the prevalence of varying missing rates in real-world scenarios, we meticulously engineer three specialized versions of each dataset: MoCAD, NGSIM, and HighD. These enhanced iterations, designated as MoCAD-missing, NGSIM-missing, and HighD-missing, are specifically tailored to embody missing rates of 25%, 50%, and 75%, where 25%, 50%, and 75% of observations are randomly missing per sample. This deliberate design ensures that each dataset is represented across a spectrum of missing conditions, providing a robust framework for a thorough and nuanced evaluation. By integrating these variants into our experimental setup, we aim to extensively assess and validate the resilience and adaptability of our approach. Correspondingly, our approach's performance is benchmarked as Ours (25%), Ours (50%), and Ours (75%), reflecting tests on the datasets with missing rates of 25%, 50%, and 75%, respectively. To address the randomness in missing datasets, we mitigate potential biases by averaging 10 experimental performances with different random seeds.

Components	Ablated variants				
	A	B	C	D	E
Physics Enhancement Stage	✘	✔	✔	✔	✔
Wave Fusion Encoder	✔	✘	✔	✔	✔
Risk Aware Module	✔	✔	✘	✔	✔
Multi-modal Decoder	✔	✔	✔	✘	✔

Table 2: Different ablated variants of ablation study.

Dataset	Horizon (s)	Model				
		A	B	C	D	E
MoCAD	1	1.35	0.59	0.42	1.03	**0.40**
	2	2.68	0.99	0.82	1.86	**0.63**
	3	5.01	1.60	1.31	2.44	**0.99**
	4	5.81	2.41	1.83	2.81	**1.72**
	5	6.12	3.27	2.89	3.22	**2.74**
NGSIM	1	1.58	0.47	0.44	0.44	**0.42**
	2	3.48	1.10	0.95	1.03	**0.93**
	3	5.05	1.80	1.49	1.72	**1.45**
	4	6.75	2.63	2.13	2.59	**2.08**
	5	8.06	3.64	2.98	3.68	**2.90**
HighD	1	0.47	0.53	0.21	0.22	**0.14**
	2	0.78	0.55	0.33	0.41	**0.26**
	3	0.86	0.77	0.49	0.62	**0.39**
	4	1.07	0.80	0.62	0.81	**0.53**
	5	1.56	1.03	0.85	1.19	**0.75**

Table 3: Ablation results for our model on 50% MoCAD-missing, NGSIM-missing, and HighD-missing datasets.

5.2 Experimental Results

Our evaluation, as delineated in Table 1, benchmarks our approach against prevailing SOTA models, which are conventionally evaluated on datasets devoid of missing observations. Leveraging the MoCAD dataset, which encapsulates a diverse range of urban and campus scenarios, our approach demonstrates a consistent superiority over the incumbent SOTA model across all prediction horizons. The performance enhancements are pronounced, ranging from 18.7% to 54.8%, underscoring the robustness of our approach in real-world conditions. On the NGSIM dataset, our approach achieves a substantial 21.3% reduction in RMSE at the pivotal 5-second prediction horizon, when juxtaposed with the STDAN model. A similar trend of improved performance is observed in the HighD dataset, where our approach achieves a remarkable 28.8% reduction in RMSE at the 5-second prediction horizon. These findings show that our approach can effectively handle missing observations in various scenarios.

5.3 Ablation Study

In this ablation study, we systematically examine the necessity of each component within our proposed approach on 50% NGSIM-missing dataset. The comparative analysis involves four ablated variants (detailed in Table 2) of the model

Evaluation Datasets	Prediction Horizon (s)				
	1	2	3	4	5
NGSIM Dataset	0.39	0.96	1.61	2.56	3.67
50% NGSIM-missing Dataset	1.06	2.46	4.28	6.41	8.72
Reconstructed Dataset	0.44	1.03	1.73	2.60	3.71

Table 4: Performance of STDAN model with three datasets.

juxtaposed against the comprehensive Model E. Specifically, Model A does not perform the physics enhancement stage. Model B omits the integration of the wave fusion encoder. Model C eschews the incorporation of risk features in the risk-aware module. Lastly, Model D operates without the multi-modal decoder. The performance of these ablated variants is detailed in Table 3. Our findings demonstrate that the comprehensive Model E consistently outperforms the ablated variants across various temporal horizons. This empirical evidence highlights the contribution of each module to the model's effectiveness, affirming their integral roles in achieving optimal predictive performance. Model A, in particular, exhibits the lowest performance. In the NGSIM dataset, its prediction error at a 5-second horizon is 64.0% higher than that of Model E. The robustness of the physics enhancement stage in mitigating missing observations is demonstrated by this stark contrast. Additionally, the performances of Models B and D validate the contributions of the wave fusion encoder and the multi-modal decoder in improving accuracy.

5.4 Validation of Wavelet Reconstruction Network

This experiment investigates the plug-and-play capability of the wavelet reconstruction network with established models, exemplified by the STDAN model. We conduct the validation with three evaluation datasets: complete NGSIM dataset as a baseline (STDAN), 50% NGSIM-missing dataset (STDAN (50%)), and reconstructed dataset by wavelet reconstruction network (STDAN (Reconstructed)). Detailed performance is presented in Table 4. The outcomes depict a notable decline in STDAN's performance with 50% NGSIM-missing dataset. Remarkably, the reconstructed dataset elevates the model's performance to near-baseline levels. To understand the comparison results more intuitively, we visualize the performance of the model with three evaluation datasets in Figure 7. This finding highlights the effectiveness of wavelet reconstruction networks in reconstructing observations and their potential to improve robustness against missing observations.

5.5 Qualitative Results

Figure 8 (a) and Figure 8 (b) depict a comparative analysis of trajectory predictions, highlighting the superior accuracy and robustness of our approach (50% NGSIM-missing dataset) over the STDAN model in both simple and complex traffic scenarios. Figure 8 (c) and Figure 8 (d) further contrast the kinematic feasibility of predicted trajectories.

6 Conclusion

In this study, we develop a two-stage physics-informed trajectory prediction approach, grounded in the foundation of physical theories including wavelet transforms, kinematic bicycle model, and quantum mechanics. This approach adeptly addresses the challenges of missing observations and enhances kinematic feasibility while effectively modeling interactions. Our experiment demonstrates the robustness and accuracy of our approach, showing its superior performance in dealing with missing observations. Furthermore, the MoCAD-missing dataset facilitates the evaluation of the model's robustness in real-world missing observation conditions.

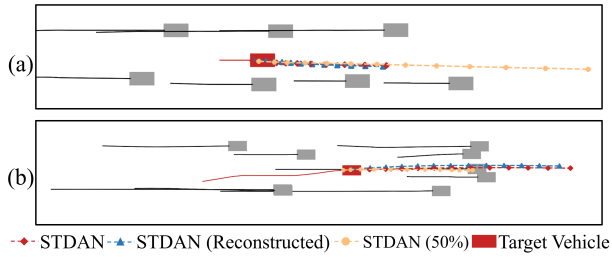

Figure 7: Visualization of predictions in three evaluation datasets. The target vehicle is marked in red while surrounding vehicles are marked in gray. (a) Evaluation across three datasets in a simple scene. (b) Evaluation across three datasets in a complex scene.

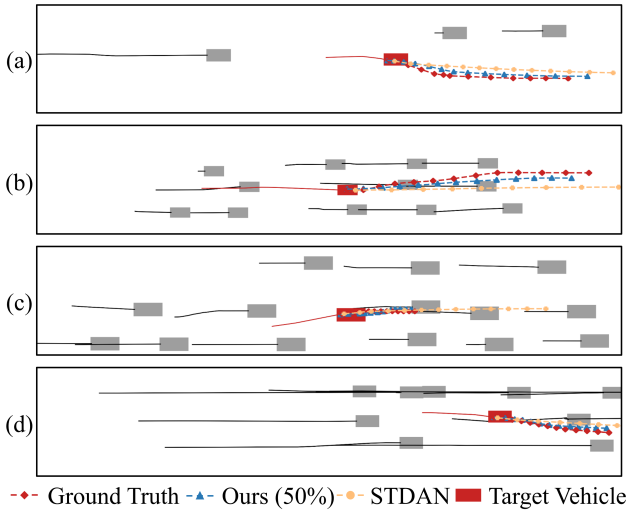

Figure 8: Qualitative accuracy comparison of our approach and STDAN model. (a) Accuracy comparison in the simple scenario. (b) Accuracy comparison in the complex scenario. (c) Excessive acceleration trajectory predicted by STDAN. (d) Implausibly small turning radius trajectory predicted by STDAN.

Acknowledgements

This research is supported by the Science and Technology Development Fund of Macau SAR (File no. 0021/2022/ITP, 0081/2022/A2, 001/2024/SKL), Shenzhen-Hong Kong-Macau Science and Technology Program Category C (SGDX20230821095159012), and University of Macau (SRG2023-00037-IOTSC). Haicheng Liao and Chengyue Wang contributed equally to this work. Please ask Dr. Zhenning Li (zhenningli@um.edu.mo) for correspondence.

References

[Arndt *et al.*, 1999] Markus Arndt, Olaf Nairz, Julian Vos-Andreae, Claudia Keller, Gerbrand Van der Zouw, and Anton Zeilinger. Wave–particle duality of c60 molecules. *nature*, 401(6754):680–682, 1999.

[Bhattacharyya *et al.*, 2023] Prarthana Bhattacharyya, Chengjie Huang, and Krzysztof Czarnecki. Ssl-lanes: Self-supervised learning for motion forecasting in autonomous driving. In *Conference on Robot Learning*, pages 1793–1805. PMLR, 2023.

[Brännström *et al.*, 2010] Mattias Brännström, Erik Coelingh, and Jonas Sjöberg. Model-based threat assessment for avoiding arbitrary vehicle collisions. *IEEE Transactions on Intelligent Transportation Systems*, 11(3):658–669, 2010.

[Cai *et al.*, 2021] Yingfeng Cai, Zihao Wang, Hai Wang, Long Chen, Yicheng Li, Miguel Angel Sotelo, and Zhixiong Li. Environment-attention network for vehicle trajectory prediction. *IEEE Transactions on Vehicular Technology*, 70(11):11216–11227, 2021.

[Chen *et al.*, 2022a] Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yingfeng Cai, Hai Wang, and Qiaolin Ye. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2022.

[Chen *et al.*, 2022b] Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yu Hu, Chenkai Tan, and Jian Yang. Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):19471–19483, 2022.

[Cong *et al.*, 2023] Peichao Cong, Yixuan Xiao, Xianquan Wan, Murong Deng, Jiaxing Li, and Xin Zhang. Dacramtp: Adaptive multi-modal vehicle trajectory prediction for dynamic drivable areas based on collision risk. *IEEE Transactions on Intelligent Vehicles*, 2023.

[Cui *et al.*, 2020] Henggang Cui, Thi Nguyen, Fang-Chieh Chou, Tsung-Han Lin, Jeff Schneider, David Bradley, and Nemanja Djuric. Deep kinematic models for kinematically feasible vehicle trajectory predictions. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10563–10569, 2020.

[Danielsson *et al.*, 2007] Simon Danielsson, Lars Petersson, and Andreas Eidehall. Monte carlo based threat assessment: Analysis and improvements. In *2007 IEEE Intelligent Vehicles Symposium*, pages 233–238, 2007.

[Deo and Trivedi, 2018] Nachiket Deo and Mohan M Trivedi. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1468–1476, 2018.

[Gao *et al.*, 2023] Kai Gao, Xunhao Li, Bin Chen, Lin Hu, Jian Liu, Ronghua Du, and Yongfu Li. Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[Gilles *et al.*, 2021] Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, and Fabien Moutarde. Home: Heatmap output for future motion estimation. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 500–507. IEEE, 2021.

[Goel, 2014] Akash Goel. Discrete wavelet transform (dwt) with two channel filter bank and decoding in image texture analysis. *Int. J. Sci. Res*, 3(4):391–397, 2014.

[Gupta *et al.*, 2018] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2255–2264, 2018.

[Huang *et al.*, 2022] Yanjun Huang, Jiatong Du, Ziru Yang, Zewei Zhou, Lin Zhang, and Hong Chen. A survey on trajectory-prediction methods for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 7(3):652–674, 2022.

[Kaempchen *et al.*, 2004] N. Kaempchen, K. Weiss, M. Schaefer, and K.C.J. Dietmayer. Imm object tracking for high dynamic driving maneuvers. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 825–830, 2004.

[Kloeden *et al.*, 2001] Craig Norman Kloeden, Giulio Ponte, and Jack McLean. *Travelling speed and risk of crash involvement on rural roads*. Australian Transport Safety Bureau, 2001.

[Krajewski *et al.*, 2018] Robert Krajewski, Julian Bock, Laurent Kloeker, and Lutz 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)*, pages 2118–2125, 2018.

[Lee *et al.*, 2019] Gregory Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

[Lefkopoulos *et al.*, 2020] Vasileios Lefkopoulos, Marcel Menner, Alexander Domahidi, and Melanie N Zeilinger. Interaction-aware motion prediction for autonomous driving: A multiple model kalman filtering scheme. *IEEE Robotics and Automation Letters*, 6(1):80–87, 2020.

[Liao *et al.*, 2024a] Haicheng Liao, Yongkang Li, Zhenning Li, Chengyue Wang, Zhiyong Cui, Shengbo Eben Li, and Chengzhong Xu. A cognitive-based trajectory prediction approach for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, pages 1–12, 2024.

[Liao *et al.*, 2024b] Haicheng Liao, Zhenning Li, Huanming Shen, Wenxuan Zeng, Dongping Liao, Guofa Li, and

Chengzhong Xu. Bat: Behavior-aware human-like trajectory prediction for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 10332–10340, 2024.

[Liao *et al.*, 2024c] Haicheng Liao, Shangqian Liu, Yongkang Li, Zhenning Li, Chengyue Wang, Bonan Wang, Yanchen Guan, and Chengzhong Xu. Human observation-inspired trajectory prediction for autonomous driving in mixed-autonomy traffic environments. *arXiv preprint arXiv:2402.04318*, 2024.

[Liao *et al.*, 2024d] Haicheng Liao, Huanming Shen, Zhenning Li, Chengyue Wang, Guofa Li, Yiming Bie, and Chengzhong Xu. Gpt-4 enhanced multimodal grounding for autonomous driving: Leveraging cross-modal attention with large language models. *Communications in Transportation Research*, 4:100116, 2024.

[Messaoud *et al.*, 2019] Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, and Fawzi Nashashibi. Non-local social pooling for vehicle trajectory prediction. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 975–980. IEEE, 2019.

[Messaoud *et al.*, 2021] Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, and Fawzi Nashashibi. Attention based vehicle trajectory prediction. *IEEE Transactions on Intelligent Vehicles*, 6(1):175–185, 2021.

[Miller and Huang, 2002] R. Miller and Qingfeng Huang. An adaptive peer-to-peer collision warning system. In *Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No.02CH37367)*, volume 1, pages 317–321 vol.1, 2002.

[Shen *et al.*, 2023] Ping Shen, Shilan Wei, Huabin Shi, Liang Gao, and Wan-Huan Zhou. Coastal flood risk and smart resilience evaluation under a changing climate. *Ocean-Land-Atmosphere Research*, 2:0029, 2023.

[Tang *et al.*, 2022] Yehui Tang, Kai Han, Jianyuan Guo, Chang Xu, Yanxi Li, Chao Xu, and Yunhe Wang. An image patch is a wave: Phase-aware vision mlp. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10935–10944, 2022.

[Wang *et al.*, 2022] Yu Wang, Shengjie Zhao, Rongqing Zhang, Xiang Cheng, and Liuqing Yang. Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion. *IEEE Transactions on Intelligent Transportation Systems*, 23(1):236–248, 2022.

[Wang *et al.*, 2023] Renzhi Wang, Senzhang Wang, Hao Yan, and Xiang Wang. Wsip: Wave superposition inspired pooling for dynamic interactions-aware trajectory prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4685–4692, 2023.

[Xie *et al.*, 2021] Xu Xie, Chi Zhang, Yixin Zhu, Ying Nian Wu, and Song-Chun Zhu. Congestion-aware multi-agent trajectory prediction for collision avoidance. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13693–13700. IEEE, 2021.

[Zhang *et al.*, 2023] Zheng Zhang, Dongyue Guo, Shizhong Zhou, Jianwei Zhang, and Yi Lin. Flight trajectory prediction enabled by time-frequency wavelet transform. *Nature Communications*, 14(1):5258, 2023.

[Zhao *et al.*, 2019] Tianyang Zhao, Yifei Xu, Mathew Monfort, Wongun Choi, Chris Baker, Yibiao Zhao, Yizhou Wang, and Ying Nian Wu. Multi-agent tensor fusion for contextual trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12126–12134, 2019.

[Zhou *et al.*, 2021] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

[Zuo *et al.*, 2023] Zhiqiang Zuo, Xinyu Wang, Songlin Guo, Zhengxuan Liu, Zheng Li, and Yijing Wang. Trajectory prediction network of autonomous vehicles with fusion of historical interactive features. *IEEE Transactions on Intelligent Vehicles*, 2023.

[Zyner *et al.*, 2017] Alex Zyner, Stewart Worrall, James Ward, and Eduardo Nebot. Long short term memory for driver intent prediction. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1484–1489. IEEE, 2017.

[Zyner *et al.*, 2019] Alex Zyner, Stewart Worrall, and Eduardo Nebot. Naturalistic driver intention and path prediction using recurrent neural networks. *IEEE transactions on intelligent transportation systems*, 21(4):1584–1594, 2019.