WAKE: Towards Robust and Physically Feasible Trajectory Prediction for Autonomous Vehicles with WAvelet and KinEmatics Synergy

Chengyue Wang*, Haicheng Liao*, Zhenning Li†, and Chengzhong Xu, *Fellow, IEEE*

***Abstract*—Addressing the pervasive challenge of imperfect data in autonomous vehicle (AV) systems, this study pioneers an integrated trajectory prediction model, WAKE, that fuses physics-informed methodologies with sophisticated machine learning techniques. Our model operates in two principal stages: the initial stage utilizes a Wavelet Reconstruction Network to accurately reconstruct missing observations, thereby preparing a robust dataset for further processing. This is followed by the Kinematic Bicycle Model which ensures that reconstructed trajectory predictions adhere strictly to physical laws governing vehicular motion. The integration of these physics-based insights with a subsequent machine learning stage, featuring a Quantum Mechanics-Inspired Interaction-aware Module, allows for sophisticated modeling of complex vehicle interactions. This fusion approach not only enhances the prediction accuracy but also enriches the model's ability to handle real-world variability and unpredictability. Extensive tests using specific versions of MoCAD, NGSIM, HighD, INTERACTION, and nuScenes datasets featuring missing observational data, have demonstrated the superior performance of our model in terms of both accuracy and physical feasibility, particularly in scenarios with significant data loss—up to 75% missing observations. Our findings underscore the potency of combining physics-informed models with advanced machine learning frameworks to advance autonomous driving technologies, aligning with the interdisciplinary nature of information fusion.**

***Index Terms*—Autonomous Driving, Trajectory Prediction, Missing Observations, Physics-Informed Models**

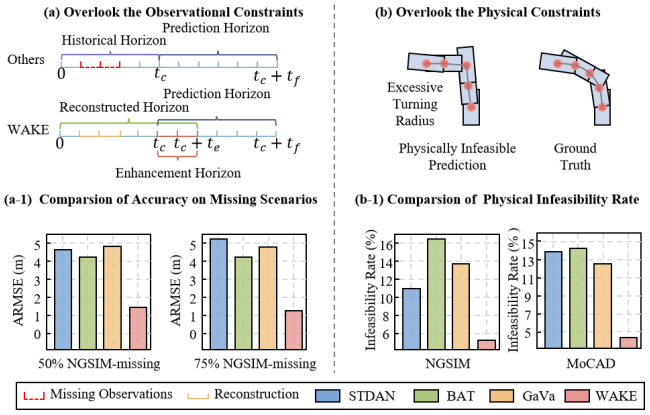

Fig. 1. Illustration of the motivation and key highlights of our proposed model. Existing models overlook the observational constraints that are prevalent in real-world driving scenarios (a). This oversight results in their inability to make accurate predictions in the presence of missing observations. Furthermore, learning-based models tend to prioritize statistical accuracy while neglecting the physical constraints of real driving scenarios (b). Such physically infeasible predictions, like turns with excessive small turning radius, fail to reflect the actual dynamics of vehicles. Motivated by the aforementioned gaps, we propose the WAKE model, which blends physics-based methodologies with advanced learning algorithms. Comparisons with existing models such as STDAN [8], BAT [9], and GaVa [10] demonstrate that our WAKE model not only maintains high accuracy in terms of Average Root Mean Squared Error (ARMSE) metric (a-1) but also ensures the physical feasibility of predictions (b-1).

I. Introduction

THE rise of autonomous vehicles (AVs) promises to revolutionize transportation, offering unprecedented levels of safety and efficiency. At the heart of this revolution lies the critical task of predicting the future trajectories of surrounding vehicles—a necessity for AVs to navigate safely and reliably in the ever-changing tapestry of real-world traffic [1], [2]. Accurate trajectory prediction enables AVs to anticipate potential hazards, make informed decisions, and interact harmoniously with human-driven vehicles, all of which are essential for the seamless integration of AVs into society.

Despite notable advancements, trajectory prediction in real-world driving conditions remains a formidable challenge [3]. One of the primary obstacles is the **imperfection of observational data**. In practical environments, data collected by sensors are often incomplete. Factors such as sensor limitations, occlusions from other vehicles or infrastructure, adverse weather, and dense traffic contribute to significant data loss [4]. These imperfections hinder the performance of models trained on idealized datasets (Fig. 1 (a)), as they struggle to generalize to the unpredictable nature of real-world scenarios [5]–[7]. Without robust mechanisms to handle missing or corrupted data, predictions can become unreliable (Fig. 1 (a-1)), compromising the safety and effectiveness of AV systems. Furthermore, the multi-modal nature of observational data, which encompasses both temporal historical state and Bird's Eye View (BEV) image data, presents a substantial challenge for the trajectory prediction field. This underscores the necessity for the development of resilient techniques capable of adapting to multi-modal data loss.

* Authors contributed equally; † Corresponding author.

C. Wang, H. Liao, Z. Li, and C. Xu are with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau SAR, 999078, China. E-mail: zhenningli@um.edu.mo

This research is supported by 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), State Key Lab of Intelligent Transportation System (2024-B001), Jiangsu Provincial Science and Technology Program (BZ2024055), and University of Macau (SRG2023-00037-IOTSC).

Another significant hurdle is **balancing prediction accuracy with physical feasibility**. Learning-based models, particularly those employing deep neural networks, excel at recognizing patterns within vast datasets and can achieve high levels of predictive accuracy. However, these models often neglect the physical laws governing vehicle dynamics (Fig. 1 (b)), sometimes generating trajectories that are statistically accurate but physically infeasible (Fig. 1 (b-1))—such as sudden, sharp turns or instant speed changes that no real vehicle could perform [11]. On the flip side, physics-based models adhere strictly to kinematic and dynamic principles, ensuring physical realism but often lacking the flexibility to capture the complex, context-dependent behaviors exhibited by human drivers [12]. This dichotomy presents a critical challenge: developing models that are both accurate in prediction and grounded in physical reality.

The third major challenge involves capturing the intricate **interactions among multiple vehicles** on the road. Traffic environments are inherently interactive; the actions of one driver can significantly influence the behaviors of others, leading to a complex web of interdependent dynamics [13]. This complexity is amplified in dense urban settings or during unexpected maneuvers like sudden braking or lane changes. Existing models frequently adopt a rigid approach to modeling these interactions, but such rigidity overlooks the diverse and dynamic nature, resulting in predictions that fail to reflect actual driving behaviors. Accurately modeling these interactions is vital for AVs to anticipate and respond appropriately to the actions of surrounding vehicles [14].

To surmount these challenges, we propose **WAKE**: a robust, two-stage trajectory prediction model that seamlessly integrates physics-informed methodologies with advanced machine learning techniques. WAKE harnesses the strengths of both physical modeling and data-driven learning to deliver trajectory predictions that are highly accurate and physically feasible. Its modular, plug-and-play design ensures easy integration into existing autonomous driving systems, enhancing their ability to navigate complex, real-world conditions.

Overall, the major contributions of this include:

- We introduce a novel Kinematic Bicycle Model integrated Wavelet Reconstruction Network that addresses the challenges of incomplete multi-modal data. By reconstructing missing observations with high accuracy, this network provides a solid foundation for subsequent trajectory prediction, enhancing robustness in real-world conditions fraught with data imperfections.
- We develop an Interaction-aware Module that leverages concepts from quantum mechanics, specifically wave superposition, to model complex traffic dynamics. This innovative approach captures the intricacies of vehicle interactions more effectively than traditional methods, leading to predictions that reflect the true nature of driver behavior in interactive environments.
- Through evaluation of the MoCAD-missing, NGSIM-missing, HighD-missing, INTERACTION-missing, and nuScenes-missing datasets, our study shows the superior performance of WAKE, even in scenes with 75% missing observations. These findings highlight its effectiveness in maintaining both accuracy and physical feasibility and confirm the plug-and-play capability of our wavelet reconstruction technique in handling missing observations.

The structure of this paper is as follows: Section II reviews the relevant literature. Section III defines the trajectory prediction task. Section IV details our proposed approach. Section V presents its performance on various datasets. Finally, Section VI concludes with a summary of the research.

II. Related Work

A. Learning-based Trajectory Prediction

Recent years have witnessed the advent of numerous datasets in the autonomous driving field, including NGSIM [15], HighD [16], and nuScenes [17]. This wealth of data has significantly propelled the evolution of learning-based trajectory prediction methodologies.

Originally, the domain of trajectory prediction witnessed the application of traditional machine learning methodologies, such as Gaussian Processes [18], Support Vector Machines [19], and Hidden Markov Models [20], showing considerable success. The advent of Recurrent Neural Network (RNN) architectures signified a fundamental transformation in the field of trajectory prediction, with deep learning methods subsequently emerging as the dominant paradigm. A key contribution was made by Chandra et al. [21] presented an RNN-based model that utilized historical data of the target vehicle, including coordinates, heading, and velocity, for predicting trajectories. Further, Xu et al. [22] leveraged Long Short-Term Memory (LSTM) networks for forecasting vehicle trajectories at intersections, utilizing speed, heading, and positional data. The development of the spatial-temporal LSTM model by Huang et al. [23] marked a significant advancement, employing the LSTM framework to capture spatial interactions and enhance prediction accuracy. Some works apply RNN-based structures to the field of ship trajectory prediction [24]. The adoption of Convolutional Neural Networks (CNNs) has been pivotal in understanding the spatial relationships between vehicles, with studies often structuring historical and contextual observations into CNN-friendly formats, such as social tensors [25] or rasterized maps [26]. Additionally, the incorporation of attention mechanisms has become a prevalent approach for modeling interactions, allowing for a refined assessment of the influence of surrounding vehicles on the target vehicle [27], [28]. Given the computational intensity of attention mechanisms, innovative methods have been introduced to efficiently characterize inter-vehicle interactions [29]–[33]. Furthermore, Graph Neural Networks (GNNs) and Graph Attention Networks (GATs) have been recognized for their ability to capture complex vehicle interactions, with Gilles et al. [34] and Liao et al. [35] demonstrating how road segments can be modeled as nodes to extract interaction feature through graph-based operations.

B. Physics-based Trajectory Prediction

While learning-driven methods have predominantly focused on data characteristics, they often overlook essential physical

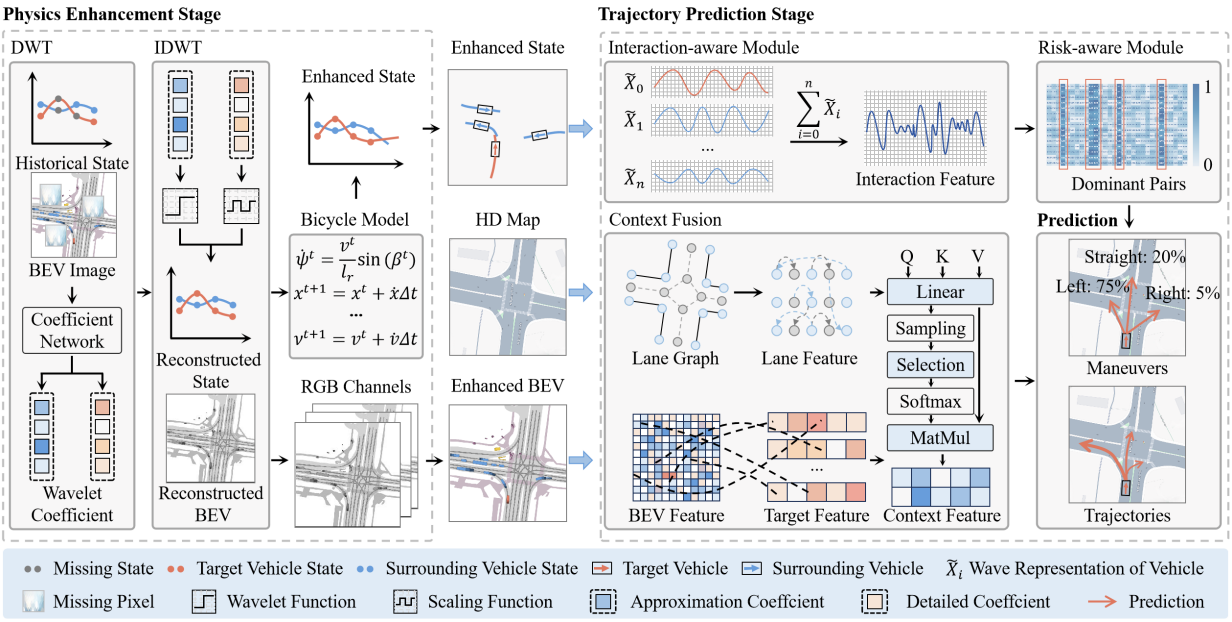

Fig. 2. Overview of the proposed WAKE model. In the **Physics Enhancement Stage**, the WAKE model utilizes physics-based models to enhance the historical observations with missing data, including historical state and BEV image. In the subsequent **Trajectory Prediction Stage**, the WAKE model utilizes these enhanced observations and pre-collected HD map to generate multi-modal predictions.

constraints, leading to trajectories that may not be physically feasible. To address this gap, researchers have sought to integrate physical knowledge into trajectory prediction processes, aiming for outcomes that are more comply with real-world physics. Cui et al. [11] innovatively used a neural network to determine parameters for a biaxial bicycle model, which then served as the output layer to produce trajectories aligned with vehicle kinematics. Model Predictive Control (MPC) has been utilized to employ vehicle dynamics models, in conjunction with current environmental data and objectives, to iteratively optimize and generate the most feasible driving trajectories. MPC methodologies consider various constraints, including vehicle dynamics limitations and traffic regulations, to ensure the physical and safety feasibility of the generated trajectories [36], [37]. The process of experimental investigation has resulted in the formulation of numerous microscopic car following and lane changing models, such as the Intelligent Driver Model (IDM) [38] and Krauss model [39], which are valued for their interpretability due to physically meaningful parameters. Additionally, wavelet transform theory is applied to flight trajectory prediction [40] highlights the innovative use of signal processing techniques in this domain. Geng et al. [41] employed a multi-loss structure to enhance the consistency of predicted trajectories with the IDM model.

III. Problem Formulation

The primary objective is to predict the trajectories of vehicles surrounding an AV. For prediction purposes, each of these vehicles is designated as a *target vehicle*. At a specific time instant t_c, our WAKE model leverages historical observations spanning across the historical horizon t_h and pre-collected high-definition (HD) maps $\boldsymbol{P}$ to predict the target vehicle's trajectory, $\boldsymbol{Y}_0$, up to a future time instant t_c+t_f. The historical observations consist of the historical states—such as coordinates and accelerations—of the target vehicle $\boldsymbol{X}_0$ and its n surrounding vehicles $\boldsymbol{X}_{1:n}$, as well as the BEV image $\boldsymbol{B}$ of the scenario. To reflect the real-world scenario of partial visibility due to various constraints, these historical observations include incomplete data.

Our innovative solution to this challenge involves the integration of a Physics Enhancement Stage, Ω, designed to enrich these incomplete historical observations with physically feasible augmentations. This approach enables our model to formulate trajectory predictions that are not only precise but also rooted in the physical feasibility of vehicle dynamics. Formally, our model is represented as:

$$\boldsymbol{Y}_0 = \Pi\left(\Omega(\boldsymbol{X}_{0:n}, \boldsymbol{B}), \boldsymbol{P}\right) \tag{1}$$

where Π signifies the Trajectory Prediction Stage.

The inherent uncertainty in human driving behavior adds a layer of complexity to the prediction task. To address this challenge and enhance predictive accuracy, we design a multi-modal decoder. Specifically, the first hierarchical level estimates uncertainties stem from different driving maneuvers M_i, while the second level models the uncertainty of vehicle dynamics using probabilistic distributions.

IV. Methodology

A. Overview Architecture of the Model

This study delineates a novel two-stage framework for trajectory prediction, as illustrated in Fig. 2. The initial stage introduces a hybrid model that synergizes a Wavelet Reconstruction Network, including Discrete Wavelet Transform (DWT) and Inverse DWT (IDWT), with a Kinematic Bicycle Model, thereby enhancing incomplete historical observations

through physics-informed methodologies. Following this enhancement, the trajectory prediction stage utilizes an encoder-decoder architecture that incorporates an interaction-aware module, a risk-aware module, context fusion, and a multi-modal decoder, facilitating multi-modal predictions based on the enriched observational dataset.

B. Physics Enhancement Stage

In order to address the dual challenges of observational and physical constraints inherent in real-world driving contexts, we propose a Physics Enhancement Stage. This stage initiates with a Wavelet Reconstruction Network, designed to achieve two-fold objectives: firstly, to reconstruct a comprehensive historical observation dataset, and secondly, to augment these observations with predictions of enhancement horizon t_e. Following this reconstruction, the Kinematic Bicycle Model is applied to generate physically feasible trajectories. This methodological advancement not only ensures the completeness of historical observations but also harnesses the physically feasible trajectories as foundational guides for the subsequent trajectory prediction stage. This integrated approach thus ensures the generation of trajectories that are not only accurate but also adhere to physical constraints.

1) Wavelet Reconstruction Network: This Wavelet Reconstruction Network, based on wavelet transform [40], [42], consists of two components: DWT and IDWT. The DWT is dedicated to generating the wavelet coefficients required by the IDWT, while the IDWT reconstructs observations.

In the DWT, historical states $\boldsymbol{X}$ and BEV images $\boldsymbol{B}$ are decomposed into wavelet coefficients $\boldsymbol{W}$. This process is achieved through the Coefficient Network, which performs convolution operations at multiple scales to extract two types of coefficients: the approximation coefficient, representing global feature, and the detail coefficients, which capture local feature. The decomposition varies between the two modalities. For historical states $\boldsymbol{X}$, the wavelet coefficients consist of an approximation coefficient, denoted as c, and a detail coefficient, d. The procedure is illustrated in Algorithm 1. In the case of the BEV image $\boldsymbol{B}$, the decomposition yields an approximation coefficient, c, along with three detail coefficients: horizontal d^H, vertical d^V, and diagonal d^D for each RGB channel. The generation of reconstructed state $\boldsymbol{X}_r(t)$ is represented as follows:

$$\boldsymbol{X}_r(t) = \sum_k c_J[k]\phi_{J,k}[t] + \sum_{j=0}^{J-1}\sum_k d_j[k]\psi_{j,k}[t] \tag{2}$$

where k represents the translation parameter, and J denotes the maximum level. t is the temporal axis. $\phi_{J,k}[t]$ and $\psi_{j,k}[t]$ are wavelet and scaling function for level j, respectively.

Since the decomposition process for BEV is conducted on a per-channel basis, generating the enhanced BEV $\bar{\boldsymbol{B}}$ requires initially reconstructing the BEV for each channel before merging them. The reconstructed BEV $\boldsymbol{B}_r(x, y)$ can be obtained through the following process:

$$\begin{aligned}\boldsymbol{B}_r(x,y) = &\sum_{k,l} c_J[k,l]\phi_{J,k}[x]\phi_{J,l}[y] \\ &+ \sum_{j=j_0}^{J-1}\sum_{k,l}(d_j^{(H)}[k,l]\psi_{j,k}[x]\phi_{j,l}[y] \\ &+ d_j^{(V)}[k,l]\phi_{j,k}[x]\psi_{j,l}[y] \\ &+ d_j^{(D)}[k,l]\psi_{j,k}[x]\psi_{j,l}[y])\end{aligned} \tag{3}$$

where the indices k and l represent the translation parameters along the two spatial dimensions x and y, respectively.

Algorithm 1: Discrete Wavelet Transform

Input: Historical state $\boldsymbol{X}$, maximum level J, wavelet function $\phi_{J,k}$, and scaling function $\psi_{j,k}$
Output: Approximation coefficients c_J and detail coefficients $\{d_j\}_{j=1}^{J}$

Initialize an empty list for storing detail coefficients as: `detail_coefficients` ← [ ]
Initialize the historical state embedding as: $(\boldsymbol{h_{em}}, \boldsymbol{h_{in}}) \leftarrow$ `GRU_encoder`$(\boldsymbol{X})$
for $j \leftarrow 1$ ***to*** J **do**
  Initialize $\boldsymbol{h_{de}}$ as zero tensor, set $\boldsymbol{h_{de}}[0,:,:] \leftarrow \boldsymbol{h_{in}}[-1,:,:]$
  $\boldsymbol{h_{co}} \leftarrow$ `Convolution`$(\boldsymbol{h_{em}}, \psi_{j,k})$
  $(\boldsymbol{h_{ou}}, \_) \leftarrow$ `GRU_decoder`$(\boldsymbol{h_{co}}, \boldsymbol{h_{de}})$
  $d_j \leftarrow$ `LayerNorm`$(\boldsymbol{h_{ou}})$
  Append: `detail_coefficients.append`(d_j)
  if j == J **then**
    $\boldsymbol{h_{co}} \leftarrow$ `Convolution`$(\boldsymbol{h_{em}}, \phi_{j,k})$
    $(\boldsymbol{h_{ou}}, \_) \leftarrow$ `GRU_decoder`$(\boldsymbol{h_{co}}, \boldsymbol{h_{de}})$
    $c_J \leftarrow$ `LayerNorm`$(\boldsymbol{h_{ou}})$
  end
end
return Approximation coefficients c_J and detail coefficients $\{d_j\}_{j=1}^{J}$

Proposition 1. *High-Fidelity Reconstruction of Historical States with Temporal Precision in 1D DWT-IDWT Operations.*

Proposition 1 posits that the 1D DWT-IDWT operation cycle can achieve the high-fidelity reconstruction of the historical state with a systematic temporal adjustment, ensuring their alignment with the original temporal sequence.

Proof. We initiate by transforming the time-domain original state, $\hat{\boldsymbol{X}}(t)$, into the z-domain utilizing the Z-transformation:

$$X(z) = \sum \hat{\boldsymbol{X}}(t) z^{-t} \tag{4}$$

where $X(z)$ symbolizes the z-domain counterpart of $\hat{\boldsymbol{X}}(t)$. z denotes the complex variable in the z-domain.

Proceeding as depicted in Fig. 3, we execute a Level-1 DWT and a downsampling operation on $X(z)$, facilitating the derivation of approximation and detail coefficients in the z-domain [43]:

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

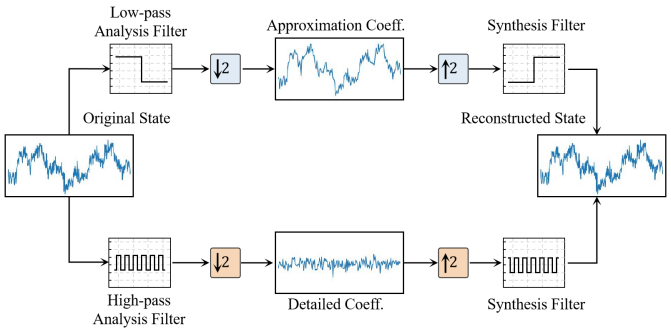

Fig. 3. Illustration of the 1D DWT-IDWT operation process for historical states. ↓ 2 is the downsampling operation. ↑ 2 is the interpolation operation.

$$X_{1,h}(z) = \frac{1}{2}\left[X\left(z^{\frac{1}{2}}\right)H\left(z^{\frac{1}{2}}\right) + X\left(-z^{\frac{1}{2}}\right)H\left(-z^{\frac{1}{2}}\right)\right] \tag{6}$$

where $G(z)$ and $H(z)$ represent the z-domain low-pass and high-pass analysis filters, respectively, while $X_{1,g}(z)$ and $X_{1,h}(z)$ denote the approximate and detailed coefficients in the z-domain after Level-1 DWT.

During the IDWT reconstruction phase, an interpolation operation is first employed to restore the original sampling frequency, followed by the application of the low-pass synthesis filter $G_1(z)$ and the high-pass synthesis filter $H_1(z)$:

$$X_{0,g}(z) = \frac{1}{2}[X(z)G(z) + X(-z)G(-z)]G_1(z) \tag{7}$$

$$X_{0,h}(z) = \frac{1}{2}[X(z)H(z) + X(-z)H(-z)]H_1(z) \tag{8}$$

Merging $X_{0,g}(z)$ and $X_{0,h}(z)$ furnishes the z-domain representation of the reconstructed states $X_0(z)$:

$$\begin{aligned} X_0(z) &= X_{0,g}(z) + X_{0,h}(z) \\ &= \frac{1}{2}\left[G(z)G_1(z) + H(z)H_1(z)\right]X(z) \\ &+ \frac{1}{2}\left[G(-z)G_1(z) + H(-z)H_1(z)\right]X(-z) \end{aligned} \tag{9}$$

Through Eq. 9, it becomes evident that the relationship between $X_0(z)$ and $X(z)$ hinges on the chosen filters. Different wavelets correspond to filters with varying parameter settings; however, their form in the z-domain can be unified as follows:

$$\begin{aligned} H(z) &= \sum_{i=0}^{Z} \lambda_i^h \cdot z^{-i} \qquad G(z) = \sum_{i=0}^{Z} \lambda_i^g \cdot z^{-i} \\ H_1(z) &= \sum_{i=0}^{Z} \lambda_i^{h,1} \cdot z^{-i} \quad G_1(z) = \sum_{i=0}^{Z} \lambda_i^{g,1} \cdot z^{-i} \end{aligned} \tag{10}$$

where the parameters λ_i^h, λ_i^g, $\lambda_i^{h,1}$, $\lambda_i^{g,1}$, and the number of terms Z in the polynomial vary with the type of wavelets.

For this study, we implement the Haar wavelet from PyWavelets [44], substituting the parameter set into Eq. 9 yields:

$$X_0(z) = z^{-1}X(z) + 0 \times X(-z) = z^{-1}X(z) \tag{11}$$

where the z^{-1} represents a one-unit temporal delay.

The Equation 11 demonstrates that the reconstructed states are effectively the original states subjected to a temporal shift of one unit. This Z-domain operation introduces a controllable temporal adjustment z^{-1} rather than distorting or omitting data, thereby proving **Proposition 1**. □

Beyond its reconstruction capabilities, the low computational complexity of the Haar wavelet also plays a crucial role in our decision. Given that the DWT is essentially a convolution operation, its computational complexity can be evaluated similarly to that of convolution. In the context of trajectory prediction, the DWT processes a historical state matrix $X \in \mathbb{R}^{T*N_a}$, where N_a represents the number of historical state features over a time horizon T. By applying a filter of size $[2, 1]$, we obtain the wavelet coefficients $\boldsymbol{W} \in \mathbb{R}^{T/2*N_a}$. Consequently, the computational complexity of the DWT can be derived as $O(T * N_a)$.

2) Kinematic Bicycle Model: To augment the physical feasibility of predicted trajectories, this study employs a two-axle Kinematic Bicycle Model, predicated on the vehicle's center of mass. This model is instrumental in augmenting the reconstructed state $\boldsymbol{X}_r$ over the enhancement horizon, serving as a pivotal reference for the trajectory prediction stage. This model includes crucial parameters such as velocity 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 axles at time t. Importantly, the steering angle is conceptualized as the angle between successive displacement vectors and the vehicle's longitudinal axis. For simplification, l_f and l_r are assumed equivalent.

The state of the vehicle at the next time step, $t + 1$, is determined using state transition equations, derived from the Kinematic Bicycle Model, as articulated by Cui et al. [11]:

$$\begin{aligned} x^{t+1} &= x^t + v^t \cos\left(\psi^t + \tan^{-1}\left(\frac{l_r}{l_f + l_r}\tan\gamma^t\right)\right)\Delta t \\ y^{t+1} &= y^t + v^t \sin\left(\psi^t + \tan^{-1}\left(\frac{l_r}{l_f + l_r}\tan\gamma^t\right)\right)\Delta t \\ \psi^{t+1} &= \psi^t + \frac{v^t}{l_r}\sin\left(\tan^{-1}\left(\frac{l_r}{l_f + l_r}\tan\gamma^t\right)\right)\Delta t \\ v^{t+1} &= v^t + a^t\Delta t \end{aligned} \tag{12}$$

This modeling paradigm enables the iterative synthesis of enhanced state $\bar{\boldsymbol{X}}$ adhering to the principles of physical laws.

C. Trajectory Prediction Stage

In the trajectory prediction stage, the enhanced state, enhanced BEV, and HD map are processed through an encoder-decoder architecture. Within the encoder, three key modules—interaction-aware module, risk-aware module, and context fusion module—extract distinct features: the interaction feature $\mathcal{F}_i$, capturing vehicle-to-vehicle interaction information; the risk feature $\mathcal{F}_r$, reflecting driver risk preferences; and the context feature $\mathcal{F}_c$, embedding contextual information from the environment. These features are subsequently leveraged by the multi-modal decoder to generate predictions.

1) Interaction-aware module: The dynamics of vehicular motion are profoundly influenced by proximate interactions, embodying both temporal and spatial dimensions. Our methodology introduces a novel paradigm, conceptualizing each vehicle as a wave. The essence of this module is encapsulated in

Algorithm 2. This representation facilitates the encapsulation of vehicular attributes within the wave's amplitude, effectively capturing discrete states. Concurrently, the wave's phase embodies the evolution of spatial and temporal dynamics, essential for decoding vehicular motion and interaction patterns. This bi-dimensional representation ensures a mathematically rigorous quantification of vehicular dynamics:

$$\tilde{\boldsymbol{X}}_j = |\boldsymbol{X}_j| \odot e^{i\boldsymbol{\theta}_j}, j = 0, 1, 2, \cdots, n \tag{13}$$

where $\tilde{\boldsymbol{X}}_j$ denotes the wave-form representation of vehicle j, $|\boldsymbol{X}_j|$ is the amplitude, $\odot$ signifies element-wise multiplication, i is the imaginary unit, and $\boldsymbol{\theta}_j$ represents the phase.

Our module architecture bifurcates into amplitude and phase encoders, each tailored to distill and refine vehicular features. The amplitude encoder, through adaptive weighting W^d, enhances amplitude information from vehicular attributes. Simultaneously, the phase encoder, employing weights W^θ, elucidates phase information, pivotal for capturing the spatiotemporal essence of vehicular dynamics. Formally,

$$\begin{aligned} |\boldsymbol{X}_j| &= \text{Amplitude-Encoder}\left(\boldsymbol{X}_j, W^d\right), \\ \boldsymbol{\theta}_j &= \text{Phase-Encoder}\left(\boldsymbol{X}_j, W^\theta\right) \end{aligned} \tag{14}$$

Leveraging the principle of wave superposition, our model adeptly captures the mutual influences among vehicles. This technique quantitatively analyzes the dynamic interactions $\tilde{\boldsymbol{X}}_r$ between any two vehicles, $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$, through the mathematical framework of wave superposition, expressed as:

$$\tilde{\boldsymbol{X}}_r = |\boldsymbol{X}_r| \odot e^{i\boldsymbol{\theta}_r} \tag{15}$$

where $|\boldsymbol{X}_r|$ is the interaction's amplitude. Mathematically,

$$|\boldsymbol{X}_r| = \sqrt{|\boldsymbol{X}_j|^2 + |\boldsymbol{X}_k|^2 + 2\,|\boldsymbol{X}_j| \odot |\boldsymbol{X}_k| \odot \cos\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j\right)} \tag{16}$$

where $\boldsymbol{\theta}_r$ is the interaction's phase, which can be defined as:

$$\begin{aligned} \boldsymbol{\theta}_r = \boldsymbol{\theta}_j + \operatorname{atan}2\,(&|\boldsymbol{X}_k| \odot \sin\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j\right), \\ &|\boldsymbol{X}_i| + |\boldsymbol{X}_j| \odot \cos\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j\right)) \end{aligned} \tag{17}$$

Then, we introduce a Fusion-Encoder that extends the interaction paradigm from pairwise vehicle dynamics to encompass the collective interplay among all vehicles:

$$\mathcal{F}_i = \text{Fusion-Encoder}\left(\tilde{\boldsymbol{X}}, W^a\right) \tag{18}$$

where $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{X}}_0, \tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_n]$ represents the collective wave-based representation of vehicles, and $\mathcal{F}_i$ denotes the interaction feature, and W^a denotes the learnable parameters.

2) Risk-aware Module: Acknowledging the variance in driving behaviors—wherein certain drivers may prefer efficiency through higher-risk maneuvers, while others prioritize safety via lower-risk alternatives—this module is designed to adapt to these divergent tendencies. The introduced risk-aware module integrates a risk factor with a dedicated risk encoder, shedding light on how distinct driving patterns impact vehicular dynamics. The risk factor, I_r, as proposed by Kloeden et al. [45], is articulated through an empirical formula:

$$I_r = \exp\left(0.07039\Delta v + 0.0008617 v^2\right) \tag{19}$$

Algorithm 2: Interaction-aware Module Algorithm

Input : Vehicle states $\{\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$, Amplitude weights W^d, Phase weights W^θ, Fusion weights W^a

Output: Interaction feature $\mathcal{F}_i$

for *each vehicle* $j \in \{0, 1, \ldots, n\}$ **do**
  Obtain amplitude $|\boldsymbol{X}_j| = \text{Amplitude-Encoder}(\boldsymbol{X}_j, W^d)$
  Obtain phase $\boldsymbol{\theta}_j = \text{Phase-Encoder}(\boldsymbol{X}_j, W^\theta)$
  Wave representation $\tilde{\boldsymbol{X}}_j = |\boldsymbol{X}_j| \odot e^{i\boldsymbol{\theta}_j}$
end
for *each pair of vehicles* j, k *where* $j \neq k$ **do**
  Compute interaction amplitude $|\boldsymbol{X}_r| = \sqrt{|\boldsymbol{X}_j|^2 + |\boldsymbol{X}_k|^2 + 2|\boldsymbol{X}_j||\boldsymbol{X}_k|\cos(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j)}$
  Compute interaction phase $\boldsymbol{\theta}_r = \boldsymbol{\theta}_j + \operatorname{atan2}\left(|\boldsymbol{X}_k|\sin(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j), |\boldsymbol{X}_j| + |\boldsymbol{X}_k|\cos(\boldsymbol{\theta}_k - \boldsymbol{\theta}_j)\right)$
  Interaction wave $\tilde{\boldsymbol{X}}_r = |\boldsymbol{X}_r| \odot e^{i\boldsymbol{\theta}_r}$
end
Concatenate all interaction waves $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{X}}_0, \tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_n]$
Obtain interaction feature $\mathcal{F}_i = \text{Fusion-Encoder}(\tilde{\boldsymbol{X}}, W^a)$
return $\mathcal{F}_i$

where v denotes the speed of the target vehicle, and Δv represents the speed differential between the target vehicle and the average traffic speed. The average traffic speed is approximated by the mean speed of surrounding vehicles $\overline{v}_s$.

Initially, observations of the target vehicle are encoded into a high-dimensional latent state. This state is then dynamically enriched with real-time risk awareness, encapsulating the contextual implications of risk on vehicular motion:

$$\mathcal{A}_r = \Phi(\delta(\Theta(\bar{\boldsymbol{X}}_0))) \odot I_{r,0} \tag{20}$$

where $\mathcal{A}_r$ denotes the risk awareness, with Φ representing the LSTM operation, δ symbolizing the LeakyReLU activation function, and Θ denoting a linear transformation.

Understanding the complex relationship between a driver's risk perception and their interactions with surrounding vehicles, our risk-aware module is engineered to refine the accuracy of driver behavior simulation by leveraging an advanced attention mechanism. This mechanism accounts for the significant influence exerted by the immediate surrounding vehicles on driving behavior. Unlike conventional attention mechanisms that compute correlations across all sequence elements—introducing unnecessary computational complexity $O(L_Q L_K)$, where L_Q and L_K denote the query and key sequence lengths, respectively—our approach adopts the ProbSparse Self-attention mechanism [29]. This method aligns with the unique demands of trajectory prediction, efficiently condensing computational complexity to $O(L_K \ln L_Q)$, thereby rendering the module exceptionally suitable for real-time vehicular applications.

The module operates with $\mathbf{Q} = \mathcal{A}_r$, and $\mathbf{K} = \mathbf{V} = \mathcal{F}_i$. Specifically, $\mathcal{A}_r = [a_{r,1}, a_{r,2}, \ldots, a_{r,L_R}]$, and $\mathcal{F}_i =$

$[f_{w,1}, f_{w,2}, \ldots, f_{w,L_W}]$, where $\mathcal{A}_r \in \mathbb{R}^{L_R \times d}, \mathcal{F}_i \in \mathbb{R}^{L_W \times d}$, respectively, with d indicating the dimensionality of the feature.

A pivotal aspect of the ProbSparse mechanism is the identification of dominant query-key pairs, which significantly influence the predictive accuracy without overwhelming computational resources. This efficiency is achieved by calculating the Kullback-Leibler (KL) divergence to evaluate the relevance of each query-key pair $p(f_{w,i}|a_{r,i})$, favoring those with a divergence significantly deviating from a uniform distribution $q = 1/L_W$. Specifically, a higher KL divergence indicates greater importance of the query-key pair. The formulation of KL divergence is expressed as follows:

$$\begin{aligned} KL(q\|p) &= \sum_{j=1}^{L_W} \frac{1}{L_W} \ln \frac{1/L_W}{k\left(a_{r,i}, f_{w,j}\right) / \sum_j k\left(a_{r,i}, f_{w,j}\right)} \\ &= \ln \sum_{j=1}^{L_W} e^{\frac{a_{r,i} f_{w,j}^\top}{\sqrt{d}}} - \frac{1}{L_W} \sum_{j=1}^{L_W} \frac{a_{r,i} f_{w,j}^\top}{\sqrt{d}} - \ln L_W \end{aligned} \tag{21}$$

This streamlined focus allows the risk-aware module to dynamically adjust to varying levels of risk exposure, directly informing the trajectory prediction process:

$$\bar{D}\left(a_{r,i}, \mathcal{F}_i\right) = \max_j \left\{a_{r,i} f_{w,j}^\top\right\} - \operatorname*{mean}_j \left\{a_{r,i} f_{w,j}^\top\right\} \tag{22}$$

where $\bar{D}$ quantifies the significance of each query-key pair, guiding the selection of dominantly interacting vehicles.

To expedite computational processing within the module, an initial selection of a random subset of U query-key pairs is conducted, narrowing down the focus to a manageable subset of interactions. Then, the most pertinent n pairs, pairs, corresponding to the immediate vehicular vicinity of the target, are meticulously selected based on their assessed importance. As a result, the crafted dominant score matrix $\overline{\boldsymbol{F}}$, with dimensions $\mathbb{R}^{n \times d}$, is ingeniously designed to underscore and integrate only those interactions deemed crucial for achieving both accurate and computationally efficient trajectory predictions. The essence of this procedure is encapsulated in Algorithm 3, which delineates the process for obtaining risk feature $\mathcal{F}_r$, thereby underscoring the module's innovative approach to addressing computational efficiency while maintaining accuracy.

3) Context Fusion: The motion patterns of vehicles are fundamentally influenced by the topology of the lane network and its inherent spatiotemporal dynamics. To align the effects of lane network topology and spatiotemporal dynamics, the context fusion module processes data into two streams, which obtain the lane feature $\mathcal{F}_l$ and spatiotemporal feature $\mathcal{F}_s$ respectively. As for the former, we employ GAT to aggregate and propagate information within the lane graph of HD map $\boldsymbol{P}$, we thereby encapsulate both local and global information into the lane feature $\mathcal{F}_l$. BEV serves as the optimal medium to encapsulate temporal and spatial characteristics. We process the BEV image with CNN layers to generate a BEV feature $\mathcal{F}_b$. Subsequently, the spatiotemporal feature $\mathcal{F}_s$, could be obtain as follows:

$$\mathcal{F}_s = \textit{ProbSparse}(Q = \mathcal{F}_t, K = \mathcal{F}_b, V = \mathcal{F}_b) \tag{23}$$

Algorithm 3: 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: Risk Feature $\mathcal{F}_r$

for *Head **in** e Heads* **do**
  $U = L_R \ln L_W$
  Randomly select U query-key pairs $\left[\boldsymbol{Q}\boldsymbol{K}^\top\right]_U$
  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
  Obtain the dominant score matrix $\overline{\boldsymbol{F}} \in \mathbb{R}^{n \times d}$ based on importance
  $\mathbf{H}_i = \text{Softmax}\left(\overline{\boldsymbol{F}}/\sqrt{d}\right) \cdot \mathcal{F}_i$
end
$\mathcal{F}_r = \textit{Concat}[\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_e]$
return $\mathcal{F}_r$

where target feature $\mathcal{F}_t$ is obtained with a LSTM encoder on enhanced states $\bar{\boldsymbol{X}}_0$.

In addition, the context feature $\mathcal{F}_c$ is obtained with:

$$\mathcal{F}_c = \textit{ProbSparse}(Q = \mathcal{F}_s, K = \mathcal{F}_l, V = \mathcal{F}_l) \tag{24}$$

4) Multi-modal Decoder: This decoder is engineered to generate predictions for each potential maneuver, thereby mitigating the uncertainties inherent in human driving behavior. Human drivers exhibit a high degree of variability in their decision-making processes, influenced by a multitude of factors including historical motion patterns, environmental conditions, interactions with surrounding vehicles, and individual risk preferences. To address these, we initially concatenate the target vehicle feature $\mathcal{F}_t$, context feature $\mathcal{F}_c$, interaction feature $\mathcal{F}_i$, and risk feature $\mathcal{F}_r$ generated from previous modules. This combined feature set is then projected into a high-dimensional space through fully connected (FC) and LSTM layers. Finally, we employ a Multi-Layer Perceptron layer to obtain the probability of the corresponding maneuver while simultaneously generating the predicted trajectory $\boldsymbol{Y}_0$ via an FC layer.

V. Experiment

A. Datasets

To comprehensively evaluate the performance of our WAKE model under various conditions, we conducted experiments on the MoCAD [9], NGSIM [46], HighD [16], INTERACTION [47], and nuScenes [17] datasets.

Furthermore, to simulate the variability and randomness of missing data encountered in real-world datasets, we engineered specialized missing versions of each primary dataset: MoCAD-missing, NGSIM-missing, HighD-missing, INTERACTION-missing, and nuScenes-missing. At the strategic level, to cover a wide spectrum of potential data loss scenarios from minimal to severe, each specialized version includes three variants to simulate different levels of missing observation rates. We tailored the missing rate gradients for each dataset based on the number of frames in their original

historical data. Specifically, for MoCAD-missing, NGSIM-missing, HighD-missing, and INTERACTION-missing, we use 25%, 50%, and 75% missing rate variants, while for nuScenes-missing, we employ 20%, 40%, and 60% variants. At the technical level, we implemented a random mechanism to determine which observations are missing, thus avoiding bias and more accurately reflecting the unpredictability inherent in real-world data collection. Specifically, for the historical state data, we use a random function to identify the missing time points, replacing all data at those time points with zeros. For the BEV image data in nuScenes-missing, we use a random function to identify missing pixels, replacing those pixels with pixels of the RGB value (255, 255, 255). This methodological refinement provides a robust framework for evaluating trajectory prediction models under various levels of data completeness, enhancing model robustness and adaptability across different operational environments.

B. Experiments Setups

1) Dataset Segmentation: To facilitate comparability, we align our dataset segmentation framework with previous studies. As for the MoCAD, NGSIM, and HighD dataset, we define a historical observation horizon t_h of 3 seconds and a prediction horizon t_f of 5 seconds. Following the INTERACTION Challenge guidelines [47], we define a historical observation horizon t_h of 1 second and a prediction horizon t_f of 3 seconds for INTERACTION dataset. For the nuScenes dataset, the historical observation horizon t_h is defined as 2 seconds and the prediction horizon t_f is defined as 6 seconds.

2) Implementation Details: We implement the WAKE model using the PyTorch framework [48]. All experiments are conducted on a single Nvidia A40 GPU with 48GB of memory unless stated otherwise. For training, we use the AdamW optimizer [49] with the CosineAnnealingLR scheduler [50]. We set the enhancement horizon t_e to 1 second, with the number of sampled points determined by the dataset's sampling frequency. The batch size is 64. The number of epochs varies between 20 and 120, depending on the dataset. Since only the evaluation on the nuScenes dataset utilized BEV and HD maps, the WAKE model applied to MoCAD, NGSIM, HighD, and INTERACTION datasets does not include the context fusion component and the context feature within the multi-modal decoder during development.

3) Accuracy Evaluation Metrics: To facilitate a comprehensive and equitable assessment of our model's predictive performance relative to state-of-the-art (SOTA) counterparts, we adopt the standard accuracy evaluation metrics prevalent within the respective datasets. This approach ensures consistency in performance evaluation and enables direct comparability of results. This study applies the Root Mean Squared Error (RMSE) [8] as the primary metric for evaluating prediction accuracy across the MoCAD, NGSIM, and HighD datasets. In addition, we employ the Average Displacement Error (ADE) and Final Displacement Error (FDE) metrics [51], [52] for the INTERACTION and nuScenes datasets.

4) Physical Feasibility Evaluation Metrics: We introduce two kinematic metrics: absolute acceleration $\mathcal{S}$, and turning radius $\mathcal{R}$. A predicted trajectory is considered physically feasible if it meets the pre-determined thresholds for all two metrics. Specifically, these thresholds are: maximum acceleration $\mathcal{S}_{max} = 8\ m/s^2$ and minimum turning radius $\mathcal{R}_{min} = 3\ m$ [11]. To assess the model's capability to generate physically feasible trajectories, we calculate the physical infeasibility rate, which measures the proportion of predicted trajectories that fail to meet these thresholds.

C. Training Losses

1) Physics Enhancement Stage: Within the Wavelet Reconstruction Network, we utilize a three-component loss comprising: the reconstruction error of the enhanced state $\mathcal{L}_{st}$, the enhanced BEV image $\mathcal{L}_{bev}$, and the error related to wavelet coefficients $\mathcal{L}_{coe}$ [40].

2) Trajectory Prediction Stage: To effectively train our trajectory prediction model, we have incorporated three distinct loss functions: intention loss $\mathcal{L}_{int}$, kinematic loss $\mathcal{L}_{kin}$, and trajectory loss $\mathcal{L}_{traj}$. The intention loss $\mathcal{L}_{int}$ is designed to accurately capture the underlying intentions of human drivers, while the kinematic loss $\mathcal{L}_{kin}$ is designed to regulate the predictions to align with the kinematic constraints. The trajectory loss $\mathcal{L}_{traj}$ ensures the precision of the predicted trajectories. We employ the cross-entropy loss function as the intention loss to ensure the WAKE model accurately captures driver intent [64]. The kinematic loss $\mathcal{L}_{kin}$ focuses on two aspects of physical feasibility: absolute acceleration $\mathcal{S}$ and turning radius $\mathcal{R}$. When any of these metrics exceed the pre-defined thresholds $\mathcal{S}_{max}$, and $\mathcal{R}_{min}$, a penalty is applied during training. Moreover, the penalties related to these two aspects are combined through a weighted sum, ensuring that each factor contributes appropriately to the kinematic loss $\mathcal{L}_{kin}$. In order to accommodate the diverse evaluation metrics associated with different datasets, we implemented two sets of loss functions during the training phase as the trajectory loss $\mathcal{L}_{traj}$. Specifically, to align with the evaluation standards of the MoCAD, NGSIM, and HighD datasets, we adopted the RMSE as our primary loss function [8]. For the INTERACTION and nuScenes datasets, we employ a weighted sum of ADE and FDE as the trajectory loss [51].

D. Experiment Results

1) Comparison Analysis of Predictive Accuracy: Our comprehensive evaluation rigorously compares our WAKE model against SOTA counterparts, assessed on complete datasets. Our findings highlight the superior performance of our model across a spectrum of scenarios and data completeness levels, underpinning its robustness and adaptability in the face of missing observations. Detailed test results for the MoCAD, NGSIM, and HighD datasets are presented in Table I, while the results for the nuScenes dataset are shown in Table II. According to the evaluation results on the MoCAD dataset, our WAKE model demonstrated outstanding performance across all prediction horizons. Notably, on the 25%-missing dataset, the model achieved significant improvements in prediction accuracy at 1 to 5-second horizons, with respective gains of 28.6%, 41.9%, 25.9%, 21.0%, and 7.6% compared to the

TABLE I
EVALUATION RESULTS FOR OUR WAKE MODEL ON MISSING SCENARIOS ALONGSIDE BASELINE MODEL'S EVALUATION RESULTS ON COMPLETE DATASETS. 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.

Dataset	Model	Prediction Horizon (s)				
		1	2	3	4	5
MoCAD	S-GAN [53]	1.69	2.25	3.30	3.89	4.69
	CS-LSTM [25]	1.45	1.98	2.94	3.56	4.49
	MHA-LSTM [54]	1.25	1.48	2.57	3.22	4.20
	NLS-LSTM [55]	0.96	1.27	2.08	2.86	3.93
	WSiP [30]	0.70	0.87	1.70	2.56	3.47
	CF-LSTM [56]	0.72	0.91	1.73	2.59	3.44
	STDAN [8]	0.62	0.85	1.62	2.51	3.32
	BAT (25%) [9]	0.65	0.99	1.89	2.81	3.58
	BAT [9]	0.35	0.74	1.39	2.19	2.88
	HLTP [35]	0.55	0.76	1.44	2.39	3.21
	WAKE (25%)	**0.25**	**0.43**	1.03	**1.73**	**2.66**
	WAKE (50%)	0.37	0.52	**0.97**	1.75	2.70
	WAKE (75%)	0.46	0.71	1.29	2.17	2.81
NGSIM	S-GAN [53]	0.57	1.32	2.22	3.26	4.40
	CS-LSTM [25]	0.61	1.27	2.09	3.10	4.37
	MATF-GAN [57]	0.66	1.34	2.08	2.97	4.13
	NLS-LSTM [55]	0.56	1.22	2.02	3.03	4.30
	IMM-KF [58]	0.58	1.36	2.28	3.37	4.55
	WSiP [30]	0.56	1.23	2.05	3.08	4.34
	CF-LSTM [56]	0.55	1.10	1.78	2.73	3.82
	MHA-LSTM [54]	0.41	1.01	1.74	2.67	3.83
	TS-GAN [59]	0.60	1.24	1.95	2.78	3.72
	STDAN [8]	0.39	0.96	1.61	2.56	3.67
	iNATran [60]	0.39	0.96	1.61	2.42	3.43
	FHIF [61]	0.40	0.98	1.66	2.52	3.63
	HTPF [62]	0.49	1.09	1.78	2.62	3.65
	DACR-AMTP [63]	0.57	1.07	1.68	2.53	3.40
	BAT [9]	**0.23**	**0.81**	1.54	2.52	3.62
	GaVa [10]	0.40	0.94	1.52	2.24	3.13
	WAKE (25%)	0.36	0.87	**1.43**	**1.98**	2.87
	WAKE (50%)	0.36	0.88	1.44	2.03	**2.86**
	WAKE (75%)	0.41	0.94	1.48	2.07	3.02
HighD	S-GAN [53]	0.30	0.78	1.46	2.34	3.41
	WSiP [30]	0.20	0.60	1.21	2.07	3.14
	CS-LSTM [25]	0.22	0.61	1.24	2.10	3.27
	MHA-LSTM(+f) [54]	0.06	0.09	0.24	0.59	1.18
	NLS-LSTM [55]	0.20	0.57	1.14	1.90	2.91
	CF-LSTM [56]	0.18	0.42	1.07	1.72	2.44
	STDAN [8]	0.19	0.27	0.48	0.91	1.66
	iNATran [60]	**0.04**	**0.05**	**0.21**	0.54	1.10
	DACR-AMTP [63]	0.10	0.17	0.31	0.54	1.01
	GaVa [10]	0.17	0.24	0.42	0.86	1.31
	WAKE (25%)	0.09	0.17	0.35	**0.47**	**0.69**
	WAKE (50%)	0.13	0.19	0.37	0.51	0.71
	WAKE (75%)	0.11	0.23	0.41	0.51	0.74

previous SOTA models. These results highlight the model's effectiveness in urban settings. In the test results on the NGSIM dataset, although our WAKE model achieved the second-best performance in short-term predictions (1-2 seconds), it is important to note that the other models were evaluated on a complete dataset, whereas our model handled varying degrees of missing data. This demonstrates the robustness of our results. Furthermore, our model exhibited excellent performance in long-term prediction horizons, with notable supremacy at the critical 5-second horizon. Against the GaVa model [10], our WAKE (50%) significantly lowers the RMSE from 3.13 meters to 2.86 meters, equivalent to an 8.6% performance improvement. This comparison highlights our model's proficiency in complex highway traffic dynamics. The evaluation results on the HighD dataset indicate that while our WAKE model is slightly less competitive in short-term predictions, its strength becomes evident in long-term prediction windows. For instance, at the 5-second prediction horizon, our WAKE (25%) outperformed the previous SOTA model by 31.7%. The model's ability to accommodate various levels of missing data lends considerable credence to the assertion that our WAKE model is both robust and possesses exceptional predictive capabilities. The evaluation results on the nuScenes dataset demonstrate that our WAKE model outperformed other SOTA models in minADE_5 and minADE_1 metrics. Moreover, when assessing the model's accuracy in predicting the final destination, as measured by minFDE_1, WAKE also achieved the best performance.

2) Comparison Analysis of Physical Feasibility: The physical feasibility of prediction results is a critical factor in assessing the overall quality of trajectory prediction models. To evaluate this, we conducted experiments on the NGSIM and MoCAD datasets, and the infeasibility rates of various models are presented in Table III. As highlighted in the table,

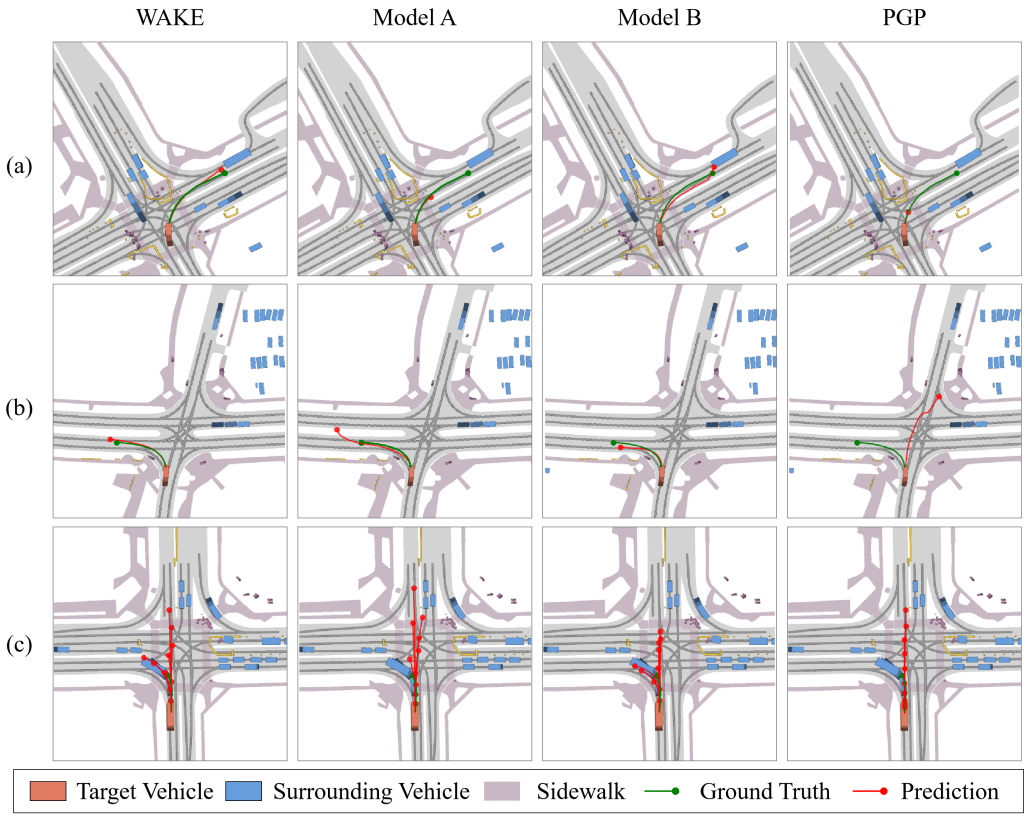

Fig. 4. Qualitative comparison of our WAKE model with various models on the 40% nuScenes-missing dataset. Model A and Model B are ablated variants of the WAKE model in the ablation study, while the PGP model is proposed in [52]. Panels (a) and (b) visualize the most probable prediction by each model. Panel (c) visualizes the multi-modal predictions by each model.

TABLE II
EVALUATION RESULTS FOR WAKE ON MISSING SCENES ALONGSIDE BASELINES' EVALUATION RESULTS ON THE COMPLETE NUSCENES.

Model	Venue	minADE_5	minADE_1	minFDE_1
DLow-AF [65]	ECCV'20	2.11	-	-
Trajectron++ [66]	ECCV'20	1.88	-	9.52
MultiPath [67]	CoRL'20	1.44	3.16	7.69
PGP [52]	CoRL'21	1.30	-	-
LDS-AF [68]	ICCV'21	2.06	-	-
AgentFormer [69]	ICCV'21	1.97	-	-
LaPred [70]	CVPR'21	1.53	3.51	8.12
STGM [71]	TITS'22	-	3.21	9.62
GoHome [34]	ICRA'22	1.42	-	6.99
ContextVAE [72]	RAL'23	1.59	3.54	8.24
EMSIN [73]	TFS'24	1.77	3.56	-
SeFlow [74]	ECCV'25	1.38	-	7.89
WAKE (20%)	-	1.20	3.01	6.92
WAKE (40%)	-	1.24	3.03	7.02
WAKE (60%)	-	1.26	3.13	7.01

our WAKE model not only delivers high prediction accuracy but also ensures a high level of physical feasibility, consistently adhering to real-world driving constraints.

3) Comparison Analysis of Real-time Performance: Driving scenarios are safety-critical, where even momentary hesitation can lead to disastrous outcomes. Therefore, real-time inference performance is a key factor in assessing a trajectory prediction model. To assess the real-time performance of our WAKE model, we compare its inference time with other models in predicting the trajectories of 12 vehicles. Table IV presents a summary of the inference times for various models, as reported in VisionTrap [75], all of which were tested using an Nvidia RTX 3090 Ti GPU. To highlight the real-time inference capability of our model in real-world scenarios, we conducted the tests using a Nvidia RTX 3090 GPU, which has fewer CUDA cores than the Nvidia RTX 3090 Ti GPU. We reported the inference time of the WAKE model by calculating its average inference time on the 40% nuScenes-missing dataset in predicting the trajectories of 12 vehicles. The results in Table IV show that the WAKE model consistently outperforms existing baselines.

Overall, the WAKE model demonstrates exceptional adaptability and robustness across diverse traffic scenarios—including urban streets, campus environments, highways, and dense traffic conditions—as well as across datasets with varying levels of missing data, such as MoCAD-missing, NGSIM-missing, HighD-missing, and nuScenes-missing. Moreover, the model excels in physical feasibility, effectively adhering to real-world driving constraints to ensure the feasibility of predicted trajectories, a capability critical for the safety of AVs. In terms of real-time performance, the WAKE model exhibits significantly faster inference speeds for multi-vehicle trajectory prediction compared to existing baseline models, particularly under resource-constrained hardware environments. These results underscore the WAKE model's high effectiveness and reliability in trajectory prediction, providing substantial support for enhancing the safety of AVs.

TABLE III
KINEMATIC EVALUATION RESULTS ON NGSIM AND MOCAD DATASET.

Model	Infeasibility Rate	
	NGSIM	MoCAD
STDAN [8]	10.8%	13.7%
BAT [9]	16.9%	14.3%
GaVa [10]	13.9%	12.8%
WAKE (50%)	2.3%	3.6%

TABLE IV
INFERENCE TIME COMPARISON ON NUSCENES DATASET.

Model	Inference Time (ms)
P2T [76]	116
Trajectron++ [66]	38
MultiPath [67]	87
AgentFormer [69]	107
PGP [52]	215
LAformer [77]	115
VisionTrap [75]	53
WAKE	21

E. Ablation Study

In the ablation study, our objective is to quantify the contribution of each constituent module within our WAKE model. To this end, we delineate the performance impacts resultant from the exclusion of specific modules, through the lens of six strategically ablated variants of our model—denominated as Model A, Model B, Model C, Model D, Model E, and Model F—contrasted against the integrative prowess of WAKE. The specific configurations for each variant are presented in Table V. The ablation variant Model A omits both the Wavelet Reconstruction Network and the Kinematic Bicycle Model, removing the Physics Enhancement Stage and directly feeding incomplete historical observations into the trajectory prediction stage. Model B removes only the Kinematic Bicycle Model from the Physics Enhancement Stage. Meanwhile, Model C discards the interaction-aware module based on the wave superposition principle. Model D removes the risk-aware module. Model E eliminates the context fusion used for multimodal information integration. Since the context fusion module is specifically designed for the nuScenes dataset, the ablation variant Model E is only applicable to the nuScenes dataset. Model F replaced the proposed multi-modal decoder, crucial for anticipating and mitigating potential prediction uncertainties. Quantitative evaluation, detailed in Table VI and Table VII, demonstrates that WAKE outperforms its ablated variants across a range of datasets. This evidence supports the assertion that the combination of these modules enhances the model's prediction efficacy. In particular, the comparative analysis reveals that Model A exhibits a notable deficiency in performance, as evidenced by its elevated RMSE at a 5-second prediction horizon on the 75% MoCAD-missing dataset, which is 55.7% higher than that of the WAKE model. This significant discrepancy highlights the indispensable role of the Physics Enhancement Stage in addressing the constraints posed by observational constraints. Moreover, the observed decline in the predictive accuracy of each ablated variant serves to reinforce the critical role of each module in bolstering the precision and reliability of the model.

TABLE V
DIFFERENT ABLATED VARIANTS OF ABLATION STUDY. NOTE: CONTEXT FUSION* ONLY EXISTS IN EVALUATION ON NUSCENES. • INDICATES THAT IT IS PRESENT ON THE NUSCENES AND NOT ON OTHER DATASETS.

Components	Ablated variants					
	A	B	C	D	E	F
Wavelet Reconstruction Network	✘	✔	✔	✔	✔	✔
Kinematic Bicycle Model	✘	✘	✔	✔	✔	✔
Interaction-aware Module	✔	✔	✘	✔	✔	✔
Risk-aware Module	✔	✔	✔	✘	✔	✔
Context Fusion*	•	•	•	•	✘	•
Multi-modal Decoder	✔	✔	✔	✔	✔	✘

TABLE VI
ABLATION RESULTS ON 75% MOCAD-MISSING, NGSIM-MISSING, AND HIGHD-MISSING DATASETS. (EVALUATION METRIC: RMSE (M))

Dataset	Time (s)	Ablated variants					WAKE
		A	B	C	D	F	
MoCAD	1	1.93	0.48	0.67	0.71	0.72	**0.46**
	2	2.79	0.79	1.16	1.17	1.22	**0.71**
	3	5.46	1.42	1.58	1.60	1.63	**1.29**
	4	5.98	2.28	2.66	2.63	2.54	**2.17**
	5	6.34	2.93	3.38	3.24	3.31	**2.81**
NGSIM	1	1.77	0.43	0.52	0.48	0.48	**0.41**
	2	3.67	1.03	1.26	1.21	1.13	**0.94**
	3	5.12	1.61	1.84	1.77	1.73	**1.48**
	4	6.88	2.31	2.61	2.58	2.59	**2.07**
	5	8.27	3.14	3.73	3.67	3.75	**3.02**
HighD	1	0.59	0.16	0.34	0.29	0.35	**0.11**
	2	0.83	0.32	0.61	0.62	0.68	**0.23**
	3	1.05	0.52	0.82	0.79	0.75	**0.41**
	4	1.31	0.69	0.94	0.92	0.98	**0.51**
	5	1.99	0.88	1.15	1.09	1.25	**0.74**

TABLE VII
ABLATION RESULTS OF WAKE ON 40% NUSCENES-MISSING DATASETS.

Metrics	Ablated Variants						WAKE
	A	B	C	D	E	F	
$mADE_5$	3.49	1.27	1.41	1.37	1.67	1.53	1.24
$mADE_1$	6.89	3.09	3.40	3.34	3.76	3.38	3.03
$mFDE_1$	10.47	7.13	7.45	7.37	8.13	7.42	7.02

F. Qualitative Comparison Analysis

In addition to quantitative evaluations, qualitative results provide a more intuitive perspective for comparison. To demonstrate how our WAKE model addresses the challenges posed by observational constraints and physical constraints, Fig. 4 presents the predictions of the various models across diverse scenarios on the 40% nuScenes-missing dataset. Model A and Model B are ablated variants of the WAKE model in the ablation study, while the PGP model is proposed in [52]. Fig. 4 (a) illustrates the most probable predictions in the intersection scenario in Boston, USA where the target vehicle is turning right. In this scenario, our WAKE model's prediction aligns perfectly with the ground truth and maintains the physical feasibility. In contrast, the predictions generated by Model A and PGP exhibit considerable deviations from the ground truth trajectory, highlighting the crucial role of our Wavelet Reconstruction Network in ensuring robust predictions under

TABLE VIII
EVALUATION RESULTS FOR SOTA BASELINES ACROSS THREE DATASETS.

Dataset	Model	Prediction Horizons				
		1	2	3	4	5
NGSIM Dataset	STDAN [8]	0.39	0.96	1.61	2.56	3.67
	BAT [9]	0.23	0.81	1.54	2.52	3.62
	GaVa [10]	0.40	0.94	1.52	2.24	3.13
50% Missing	STDAN [8]	1.06	2.46	4.28	6.41	8.72
	BAT [9]	0.96	2.13	3.98	6.35	7.99
	GaVa [10]	1.19	2.55	4.53	6.70	8.38
Reconstructed	STDAN [8]	0.44	1.03	1.73	2.60	3.71
	BAT [9]	0.26	0.99	1.65	2.68	3.84
	GaVa [10]	0.47	1.06	1.63	2.45	3.37

TABLE IX
EVALUATION RESULTS OF WAKE WITH AND WITHOUT WAVELET RECONSTRUCTED NETWORK ON THE INTERACTION DATASET.

Model	Evaluation Metric (m)	
	minADE	minFDE
WAKE (w/o)	2.07	2.47
WAKE (w)	1.58	1.89

incomplete data scenarios. While the predictions of Model B demonstrate high accuracy, the absence of enhancements from the Kinematic Bicycle Model results in overly large turning radius in its trajectory predictions. Fig. 4 (b) depicts the most probable prediction in the intersection scenario in Singapore where the target vehicle is turning left. It is evident that Model A, Model B, and the PGP model produce physically infeasible predictions in the absence of the Kinematic Bicycle Model. Furthermore, without the Wavelet Reconstruction Network, the predictions of Model A and the PGP model diverge significantly from the ground truth. Fig. 4 (c) visualizes the multi-modal prediction results of different models at a complex intersection. We observe that Model A and the PGP model, lacking the Wavelet Reconstruction Network, fail to capture the driver's true left-turn intent under conditions of incomplete data. In contrast, both the WAKE model and Model B accurately predict the driver's left-turn behavior.

Overall, this qualitative comparison experiment, through the analysis of prediction results from different models across various scenarios, intuitively demonstrates the high accuracy and physical feasibility of our proposed WAKE model under conditions of missing data.

G. Validation for Wavelet Reconstruction Network

In our investigation, we designed two distinct experiments to underscore the robustness and versatility of the proposed Wavelet Reconstruction Network within the trajectory prediction domain. The first experiment is crafted to assess the plug-and-play capability of our Wavelet Reconstruction Network. By integrating this network into a suite of pre-established models, we aim to ascertain its efficacy in enhancing predictive accuracy and interoperability across different predictive architectures. This experiment is pivotal in demonstrating the network's utility as a versatile component that can significantly bolster the performance of existing trajectory prediction models. Specifically, we conducted the validation with the STDAN [8], BAT [9], and GaVa models [10] on three

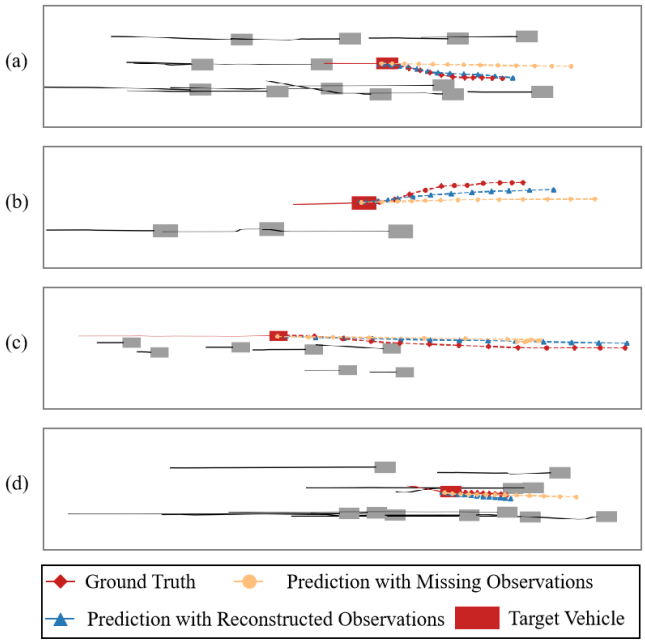

Fig. 5. Illustrate the plug-and-play capability of our proposed Wavelet Reconstruction Network and its impact on predictive performance. Panels (a), (b), (c), and (d) present the comparative prediction results before (Orange) and after (Blue) the Wavelet Reconstruction Network applied to the WAKE, STDAN, BAT, and GaVa models, on 50% NGSIM-missing Dataset.

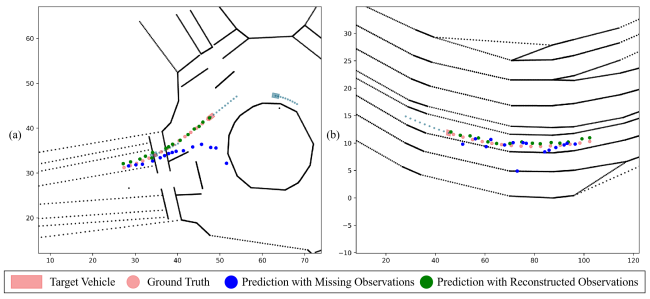

Fig. 6. Illustrate the adaptability of our proposed Wavelet Reconstruction Network and its impact on predictive performance. Panels (a) and (b) present the comparative prediction results before (Blue) and after (Green) the Wavelet Reconstruction Network applied to our WAKE models, on 50% INTERACTION-missing Dataset.

evaluation datasets: the complete NGSIM dataset as a baseline, the 50% NGSIM-missing dataset, and the reconstructed dataset by Wavelet Reconstruction Network. As detailed in Table VIII, the quantitative performance of various models demonstrates a notable decline in accuracy when applied to the 50% NGSIM-missing dataset. Notably, the deployment of the wavelet reconstructed dataset substantially boosts the model's performance, restoring it to levels akin to those achieved with the complete dataset. To intuitively showcase the prowess of our Wavelet Reconstruction Network in mitigating the issue of missing observations, we present Fig. 5 to demonstrate the enhanced accuracy and robustness of our proposed Wavelet Reconstruction Network when applied to the 50% NGSIM-missing dataset. This comparison not only highlights the importance of our Wavelet Reconstruction Network in terms of prediction accuracy but also showcases its plug-and-play ca-

pability to different prediction models. Subsequently, our second experiment shifts focus towards evaluating the network's generalization prowess across a spectrum of driving scenarios that span various countries and driving cultures. To rigorously assess the network's performance, we conducted a experiment on the 50% INTERACTION-missing dataset. As shown in Table IX, we established two distinct groups for comparison: a baseline group, labeled as WAKE (w/o), without the integration of the Wavelet Reconstruction Network, and WAKE (w), which leveraged the network. Moreover, to elucidate the adaptability of our Wavelet Reconstruction Network across diverse scenarios, we visualize the performance in Fig. 6. This comparison underscores the WAKE model's adaptability in diverse cultural and contextual scenarios. The comparative analysis reveals a significant improvement in the predictions across different scenarios upon the inclusion of the Wavelet Reconstruction Network. These findings not only highlight the network's efficacy in enhancing prediction accuracy but also its broad applicability and effectiveness across diverse contexts.

VI. Conclusion

This paper has introduced a novel two-stage trajectory prediction model for AVs, addressing key challenges in the domain, including missing observations and adherence to physical constraints. Our comprehensive approach, integrating the Wavelet Reconstruction Network with the Kinematic Bicycle Model in the Physical Enhancement Stage, has demonstrated a significant advancement in trajectory prediction accuracy and physical feasibility. Our proposed interaction-aware module, inspired by quantum mechanics, represents a paradigm shift in modeling vehicle interactions, offering a novel perspective that enhances our understanding of vehicular dynamics in mixed-traffic environments. This innovation, along with the integration of the risk-aware module, context fusion, and multi-modal decoder, underscores the model's capability to capture the inherent uncertainties of human driving behaviors. The evaluations underscore our model's superior performance, particularly in handling up to 75% missing observations, outperforming existing SOTA models across multiple complex scenarios. In conclusion, this study contributes a significant leap forward in the trajectory prediction field, offering a robust, accurate, and physically feasible model that is well-equipped to tackle the challenges in real-world driving scenarios.

This study primarily focuses on the physical feasibility of vehicle trajectory prediction; however, the physical feasibility of pedestrian trajectory prediction remains unaddressed. Future research should explore methods that draw on human body dynamics to improve the accuracy and realism of pedestrian trajectory predictions. By combining dynamics-based models with learning-based techniques, we can develop more robust and physically feasible predictions, contributing to safer and more reliable trajectory prediction in autonomous driving.

References

[1] F. Marchetti, F. Becattini, L. Seidenari, and A. Del Bimbo, "Multiple trajectory prediction of moving agents with memory augmented networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 6688–6702, 2020.

[2] W. Yu, C. Zhao, H. Wang, J. Liu, X. Ma, Y. Yang, J. Li, W. Wang, X. Hu, and D. Zhao, "Online legal driving behavior monitoring for self-driving vehicles," *Nature communications*, vol. 15, no. 1, p. 408, 2024.

[3] S. Feng, H. Sun, X. Yan, H. Zhu, Z. Zou, S. Shen, and H. X. Liu, "Dense reinforcement learning for safety validation of autonomous vehicles," *Nature*, vol. 615, no. 7953, pp. 620–627, 2023.

[4] X. Yan, Z. Zou, S. Feng, H. Zhu, H. Sun, and H. Liu, "Learning naturalistic driving environment with statistical realism," *Nature Communications*, vol. 14, 04 2023.

[5] Y. Wang, S. Zhao, R. Zhang, X. Cheng, and L. Yang, "Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 236–248, 2020.

[6] X. Wang, J. Liu, H. Lin, S. Garg, and M. Alrashoud, "A multi-modal spatial–temporal model for accurate motion forecasting with visual fusion," *Information Fusion*, vol. 102, p. 102046, 2024.

[7] Z. Li, Z. Chen, Y. Li, and C. Xu, "Context-aware trajectory prediction for autonomous driving in heterogeneous environments," *Computer-Aided Civil and Infrastructure Engineering*, vol. 39, no. 1, pp. 120–135, 2024.

[8] X. Chen, H. Zhang, F. Zhao, Y. Hu, C. Tan, and J. Yang, "Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19 471–19 483, 2022.

[9] H. Liao, Z. Li, H. Shen, W. Zeng, D. Liao, G. Li, and C. Xu, "Bat: Behavior-aware human-like trajectory prediction for autonomous driving," in *AAAI*, vol. 38, no. 9, 2024, pp. 10 332–10 340.

[10] H. Liao, S. Liu, Y. Li, Z. Li, C. Wang, B. Wang, Y. Guan, and C. Xu, "Human observation-inspired trajectory prediction for autonomous driving in mixed-autonomy traffic environments," *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[11] H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, J. Schneider, D. Bradley, and N. Djuric, "Deep kinematic models for kinematically feasible vehicle trajectory predictions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 10 563–10 569.

[12] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, "A survey on trajectory-prediction methods for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 652–674, 2022.

[13] F. Marchetti, F. Becattini, L. Seidenari, and A. Del Bimbo, "Smemo: social memory for trajectory forecasting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[14] J. Sun, Y. Li, L. Chai, and C. Lu, "Modality exploration, retrieval and adaptation for trajectory prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[15] J. Colyar and J. Halkias, "Us highway 101 dataset," *Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-07-030*, pp. 27–69, 2007.

[16] 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)*, 2018, pp. 2118–2125.

[17] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF CVPR*, 2020, pp. 11 621–11 631.

[18] J. Joseph, F. Doshi-Velez, A. S. Huang, and N. Roy, "A bayesian nonparametric approach to modeling motion patterns," *Autonomous Robots*, vol. 31, pp. 383–400, 2011.

[19] G. S. Aoude, B. D. Luders, K. K. Lee, D. S. Levine, and J. P. How, "Threat assessment design for driver assistance system at intersections," in *13th international ieee conference on intelligent transportation systems*. IEEE, 2010, pp. 1855–1862.

[20] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, "Next place prediction using mobility markov chains," in *Proceedings of the first workshop on measurement, privacy, and mobility*, 2012, pp. 1–6.

[21] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions," in *Proceedings of the IEEE/CVF CVPR*, 2019, pp. 8483–8492.

[22] D. Xu, Z. Ding, X. He, H. Zhao, M. Moze, F. Aioun, and F. Guillemard, "Learning from naturalistic driving data for human-like autonomous highway driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7341–7354, 2020.

[23] 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)*, 2022, pp. 2605–2611.

[24] H. Li, H. Jiao, and Z. Yang, "Ais data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods," *Transportation Research Part E: Logistics and Transportation Review*, vol. 175, p. 103152, 2023.

[25] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *Proceedings of the IEEE CVPR workshops*, 2018, pp. 1468–1476.

[26] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, "Home: Heatmap output for future motion estimation," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 500–507.

[27] 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.

[28] K. Zhang, L. Zhao, C. Dong, L. Wu, and L. Zheng, "Ai-tp: Attention-based interaction-aware trajectory prediction for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 73–83, 2022.

[29] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *AAAI*, vol. 35, 2021, pp. 11 106–11 115.

[30] R. Wang, S. Wang, H. Yan, and X. Wang, "Wsip: Wave superposition inspired pooling for dynamic interactions-aware trajectory prediction," in *AAAI*, 2023, pp. 4685–4692.

[31] Y. Tang, K. Han, J. Guo, C. Xu, Y. Li, C. Xu, and Y. Wang, "An image patch is a wave: Phase-aware vision mlp," in *Proceedings of the IEEE/CVF CVPR*, 2022, pp. 10 935–10 944.

[32] R. Girgis, F. Golemo, F. Codevilla, M. Weiss, J. A. D'Souza, S. E. Kahou, F. Heide, and C. Pal, "Latent variable sequential set transformers for joint multi-agent motion prediction," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=Dup_dDqkZC5

[33] Y. Yang, Q. Zhang, T. Gilles, N. Batool, and J. Folkesson, "Rmp: A random mask pretrain framework for motion prediction," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 3717–3723.

[34] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, "Gohome: Graph-oriented heatmap output for future motion estimation," in *2022 international conference on robotics and automation (ICRA)*. IEEE, 2022, pp. 9107–9114.

[35] H. Liao, Y. Li, Z. Li, C. Wang, Z. Cui, S. E. Li, and C. Xu, "A cognitive-based trajectory prediction approach for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, pp. 1–12, 2024.

[36] B. Yi, S. Gottschling, J. Ferdinand, N. Simm, F. Bonarens, and C. Stiller, "Real time integrated vehicle dynamics control and trajectory planning with mpc for critical maneuvers," in *2016 IEEE intelligent vehicles symposium (IV)*. IEEE, 2016, pp. 584–589.

[37] M. Bahari, I. Nejjar, and A. Alahi, "Injecting knowledge in data-driven vehicle trajectory predictors," *Transportation research part C: emerging technologies*, vol. 128, p. 103010, 2021.

[38] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.

[39] S. Krauss, "Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics," DLR Deutsches Zentrum fuer Luft- und Raumfahrt e.V., Koeln (Germany). Abt. Unternehmensorganisation und -information; Koeln Univ. (Germany). Mathematisch-Naturwissenschaftliche Fakultaet, Technical Report DLR-FB-98-08, Apr 1998, tH: Diss.; PBD: Apr 1998. [Online]. Available: https://www.osti.gov/etdeweb/biblio/627062

[40] Z. Zhang, D. Guo, S. Zhou, J. Zhang, and Y. Lin, "Flight trajectory prediction enabled by time-frequency wavelet transform," *Nature Communications*, vol. 14, no. 1, p. 5258, 2023.

[41] M. Geng, J. Li, Y. Xia, and X. M. Chen, "A physics-informed transformer model for vehicle trajectory prediction on highways," *Transportation research part C: emerging technologies*, vol. 154, p. 104272, 2023.

[42] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[43] A. Goel, "Discrete wavelet transform (dwt) with two channel filter bank and decoding in image texture analysis," *Int. J. Sci. Res*, vol. 3, no. 4, pp. 391–397, 2014.

[44] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "Pywavelets: A python package for wavelet analysis," *Journal of Open Source Software*, vol. 4, no. 36, p. 1237, 2019.

[45] C. N. Kloeden, G. Ponte, and J. McLean, *Travelling speed and risk of crash involvement on rural roads*. Australian Transport Safety Bureau, 2001.

[46] J. A. Halkias and J. Colyar, "Us highway i-80 dataset," *Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-06-137*, 2007.

[47] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kümmerle, H. Königshof, C. Stiller, A. de La Fortelle, and M. Tomizuka, "INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps," *arXiv:1910.03088 [cs, eess]*, 2019.

[48] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[49] I. Loshchilov, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[50] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *International Conference on Learning Representations*, 2022.

[51] J. Chen, Z. Wang, J. Wang, and B. Cai, "Q-eanet: Implicit social modeling for trajectory prediction via experience-anchored queries," *IET Intelligent Transport Systems*, vol. 18, no. 6, pp. 1004–1015, 2024.

[52] N. Deo, E. Wolff, and O. Beijbom, "Multimodal trajectory prediction conditioned on lane-graph traversals," in *Conference on Robot Learning*. PMLR, 2022, pp. 203–212.

[53] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE CVPR*, 2018, pp. 2255–2264.

[54] 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, 2021.

[55] ——, "Non-local social pooling for vehicle trajectory prediction," in *IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 975–980.

[56] X. Xie, C. Zhang, Y. Zhu, Y. N. Wu, and S.-C. Zhu, "Congestion-aware multi-agent trajectory prediction for collision avoidance," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 693–13 700.

[57] T. Zhao, Y. Xu, M. Monfort, W. Choi, C. Baker, Y. Zhao, Y. Wang, and Y. N. Wu, "Multi-agent tensor fusion for contextual trajectory prediction," in *Proceedings of the IEEE/CVF CVPR*, 2019, pp. 12 126–12 134.

[58] V. Lefkopoulos, M. Menner, A. Domahidi, and M. N. Zeilinger, "Interaction-aware motion prediction for autonomous driving: A multiple model kalman filtering scheme," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 80–87, 2020.

[59] Y. Wang, S. Zhao, R. Zhang, X. Cheng, and L. Yang, "Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 236–248, 2022.

[60] X. Chen, H. Zhang, F. Zhao, Y. Cai, H. Wang, and Q. 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*, vol. 71, pp. 1–12, 2022.

[61] Z. Zuo, X. Wang, S. Guo, Z. Liu, Z. Li, and Y. Wang, "Trajectory prediction network of autonomous vehicles with fusion of historical interactive features," *IEEE Transactions on Intelligent Vehicles*, 2023.

[62] M. Wang, L. Zhang, J. Chen, Z. Zhang, Z. Wang, and D. Cao, "A hybrid trajectory prediction framework for automated vehicles with attention mechanisms," *IEEE Transactions on Transportation Electrification*, 2023.

[63] P. Cong, Y. Xiao, X. Wan, M. Deng, J. Li, and X. Zhang, "Dacr-amtp: Adaptive multi-modal vehicle trajectory prediction for dynamic drivable areas based on collision risk," *IEEE Transactions on Intelligent Vehicles*, 2023.

[64] H. Liao, X. Li, Y. Li, H. Kong, C. Wang, B. Wang, Y. Guan, K. Tam, Z. Li, and C. Xu, "Characterized diffusion and spatial-temporal interaction network for trajectory prediction in autonomous driving," *International Joint Conference On Artificial Intelligence (IJCAI 2024)*, 2024.

[65] Y. Yuan and K. Kitani, "Dlow: Diversifying latent flows for diverse human motion prediction," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 2020, pp. 346–364.

[66] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer, 2020, pp. 683–700.

[67] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction," in *Conference on Robot Learning*. PMLR, 2020, pp. 86–99.
[68] Y. J. Ma, J. P. Inala, D. Jayaraman, and O. Bastani, "Likelihood-based diverse sampling for trajectory forecasting," 2021.
[69] Y. Yuan, X. Weng, Y. Ou, and K. M. Kitani, "Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9813–9823.
[70] B. Kim, S. H. Park, S. Lee, E. Khoshimjonov, D. Kum, J. Kim, J. S. Kim, and J. W. Choi, "Lapred: Lane-aware prediction of multi-modal future trajectories of dynamic agents," in *Proceedings of the IEEE/CVF CVPR*, 2021, pp. 14 636–14 645.
[71] Z. Zhong, Y. Luo, and W. Liang, "Stgm: Vehicle trajectory prediction based on generative model for spatial-temporal features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 18 785–18 793, 2022.
[72] P. Xu, J.-B. Hayet, and I. Karamouzas, "Context-aware timewise vaes for real-time vehicle trajectory prediction," *IEEE Robotics and Automation Letters*, 2023.
[73] Y. Ren, Z. Lan, L. Liu, and H. Yu, "Emsin: Enhanced multi-stream interaction network for vehicle trajectory prediction," *IEEE Transactions on Fuzzy Systems*, 2024.
[74] Q. Zhang, Y. Yang, P. Li, O. Andersson, and P. Jensfelt, "Seflow: A self-supervised scene flow method in autonomous driving," in *European Conference on Computer Vision*. Springer, 2025, pp. 353–369.
[75] S. Moon, H. Woo, H. Park, H. Jung, R. Mahjourian, H.-g. Chi, H. Lim, S. Kim, and J. Kim, "Visiontrap: Vision-augmented trajectory prediction guided by textual descriptions," in *European Conference on Computer Vision*. Springer, 2025, pp. 361–379.
[76] N. Deo and M. M. Trivedi, "Trajectory forecasts in unknown environments conditioned on grid-based plans," *arXiv preprint arXiv:2001.00735*, 2020.
[77] M. Liu, H. Cheng, L. Chen, H. Broszio, J. Li, R. Zhao, M. Sester, and M. Y. Yang, "Laformer: Trajectory prediction for autonomous driving with lane-aware scene constraints," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 2039–2049.

Chengyue Wang is currently a pursuing Ph.D. degree at the State Key Laboratory of Internet of Things for Smart City and the Department of Civil Engineering at the University of Macau. He holds an MS degree in Civil Engineering from the University of Illinois Urbana-Champaign (2022) and a B.E in Transportation Engineering from Chang'an University (2021). His research primarily focuses on the innovative integration of artificial intelligence with autonomous driving technologies and intelligent transportation systems.

Haicheng Liao (Student Member, IEEE) received the B.S. degree in software engineering from the University of Electronic Science and Technology of China (UESTC) in 2022. He is currently pursuing the Ph.D. degree at the State Key Laboratory of Internet of Things for Smart City and the Department of Computer and Information Science, University of Macau. His research interests include connected autonomous vehicles and the application of deep reinforcement learning to autonomous driving.

Zhenning Li (Member, IEEE) received his Ph.D. in Civil Engineering from the University of Hawaii at Manoa, Honolulu, Hawaii, USA, in 2019. Currently, he holds the position of Assistant Professor at the State Key Laboratory of Internet of Things for Smart City, as well as the Department of Computer and Information Science at the University of Macau, Macau. Over his academic career, he has published over 50 papers. His main areas of research focus on the intersection of connected autonomous vehicles and Big Data applications in urban transportation systems. He has been honored with several awards, including the TRB Best Young Researcher award and the CICTP Best Paper Award, amongst others.

Chengzhong Xu (Fellow, IEEE) received the Ph.D. degree from The University of Hong Kong, in 1993. He is currently the chair professor of computer science and the dean with the Faculty of Science and Technology, University of Macau. Prior to this, he was with the faculty at Wayne State University, USA, and the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. He has published more than 400 papers and more than 100 patents. His research interests include cloud computing and data-driven intelligent applications. He was the Best Paper awardee or the Nominee of ICPP2005, HPCA2013, HPDC2013, Cluster2015, GPC2018, UIC2018, and AIMS2019. He also won the Best Paper award of SoCC2021. He was the Chair of the IEEE Technical Committee on Distributed Processing from 2015 to 2019.