EKI-GAN: Context-Aware Vehicle Trajectory Forecasting with Vehicle Factors and Environmental Information at Signalized Intersections

Chuheng Wei * Guoyuan Wu Matthew J. Barth
University of California, Riverside
chuheng.wei@email.ucr.edu

Abstract

Accurate vehicle trajectory prediction at signalized intersections is crucial for autonomous driving and traffic management systems, yet remains challenging due to complex multi-modal interactions between vehicles, pedestrians, and infrastructure elements. Existing approaches inadequately integrate the diverse factors influencing intersection dynamics, including vehicle characteristics, environmental context, and human-vehicle interactions. We present Enhanced Knowledge-Informed Generative Adversarial Network (EKI-GAN), a comprehensive framework that addresses these limitations through enhanced multi-modal encoding and advanced attention mechanisms. Our approach introduces a dual-stream architecture with six specialized encoders that comprehensively capture vehicle factors and environmental information for robust trajectory prediction. The framework employs Multi-Dimensional Attention Pooling Network (MAP-Net), which extends conventional attention through multi-dimensional feature extraction, stabilized computation, and adaptive gating to model complex vehicular interactions. Extensive experiments on the SinD dataset demonstrate state-of-the-art performance, achieving ADE of 0.09 and FDE of 0.19 for 9-second predictions, with particularly notable improvements in extended temporal prediction scenarios. Code is available at GitHub.

1. Introduction

Signalized intersections represent critical nodes in urban transportation networks, where the complexity of traffic dynamics reaches its peak due to the convergence of multiple transportation modes and regulatory systems. These intersections serve as focal points where vehicles, pedestrians, cyclists, and other road users interact within a constrained spatial environment governed by traffic control signals [33]. As illustrated in Figure 1, vehicle trajectory prediction at signalized intersections is fundamentally influenced by two categories: Vehicle Factors (spatial trajectory patterns, vehicle dimensions, velocity profiles) that drive complex social dynamics through proximity-based influences, and Environmental Information (map topology, traffic signal states, pedestrian positions) that provides critical contextual constraints. The comprehensive integration of both vehicle-centric interaction dynamics and environmental contextual information represents a significant gap in current trajectory prediction methodologies.

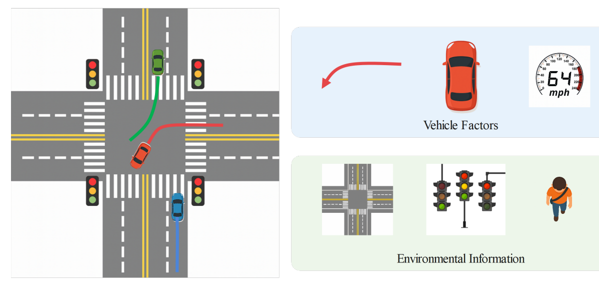


Figure 1. Multi-Modal Factors Influencing Vehicle Trajectory Prediction at Signalized Intersections: Vehicle Factors (trajectory patterns, physical attributes, motion dynamics) and Environmental Information (map topology, traffic signals, pedestrian dynamics) that collectively determine vehicular movement decisions in complex intersection scenarios.

Traditional trajectory prediction models struggle with the inherent complexity of multi-modal traffic environments, where different road users exhibit distinct behavioral patterns and interaction mechanisms [19]. Critical limitations include insufficient consideration of pedestrian dynamics [29], inadequate utilization of spatial context information in road network topology [17], and limited modeling of traffic signal influences in complex intersection environments [25]. While the recently introduced Knowledge-Informed Generative Adversarial Network (KI-GAN) [26] demonstrated advances by integrating vehicle characteristics and traffic signal information through VAP-Net, significant gaps remain in pedestrian interaction modeling, map

*Corresponding Author.

topology utilization, and spatial context exploitation.

To address these challenges, we present Enhanced Knowledge-Informed Generative Adversarial Network (EKI-GAN), an advanced framework that enhances the encoder architecture to explicitly model vehicle-pedestrian interactions [9], integrates comprehensive map topology information to improve spatial awareness [21], and develops Multi-Dimensional Attention Pooling Network (MAP-Net) that synergistically combines pedestrian dynamics, vehicle interactions, and map-based spatial relationships for more accurate trajectory predictions [20].

The primary contributions of this work include:

- **Enhanced Multi-Modal Encoder Architecture**: We introduce a dual-stream framework that systematically integrates vehicle factors (trajectory patterns, physical attributes, motion dynamics) and environmental information (map topology, traffic signals, pedestrian dynamics) through dedicated encoders and hierarchical attention mechanisms.
- **Multi-Dimensional Attention Pooling Network (MAP-Net)**: We develop an advanced pooling mechanism with multi-dimensional feature extraction, stabilized attention computation, and adaptive context-aware gating for sophisticated vehicle interaction modeling.
- **State-of-the-Art Performance**: We achieve 10% improvement in ADE and 9.5% improvement in FDE compared to previous methods on the SinD dataset, establishing new benchmarks for intersection trajectory forecasting.

2. Related Work

2.1. Vehicle Trajectory Prediction at Intersections

Trajectory prediction at intersections plays a vital role in ensuring safety and efficiency in intelligent transportation systems. Early work by Kawasaki and Tasaki [12] used LSTM models to forecast turning trajectories based on vehicle speed and intersection geometry. Tan et al. [22] addressed risky behavior prediction at high-speed intersections using a hybrid approach combining adaptive Kalman filters and k-nearest neighbor models, achieving over 95% accuracy in detecting aggressive passes and red-light violations.

To incorporate driver intent and proactive safety mechanisms, Chen et al. [6] proposed a cooperative scheme in the Internet of Vehicles framework, using a BP neural network for intention recognition and a Kalman filter for trajectory prediction, followed by a collision probability model. Huang et al. [11] introduced a variational neural network that outputs trajectory distributions with confidence scores, enabling uncertainty-aware prediction and integration with physics-based models. Gao et al. [7] combined dynamic Bayesian networks and LSTM to predict cyclists' intentions and future motion at unsignalized intersections.

While recent models incorporate specific signals, intentions, or uncertainty, most are limited to narrow contexts or simplified settings. For example, Zhang et al. [33] and Oh et al. [18] explore traffic light influence, but often ignore broader scene context and multi-modal fusion. Wang et al. [24] used trajectory-based reinforcement learning to suppress speed oscillations in mixed traffic. In contrast, our EKI-GAN integrates signal status, vehicle dynamics, and map context into a unified framework, enhancing prediction accuracy in complex intersection scenarios.

2.2. Multi-Agent Interaction Modeling for Vehicle Forecasting

Accurately modeling vehicle-to-vehicle interactions is essential for reliable trajectory forecasting in complex traffic scenes. Early studies such as Social LSTM [1] and Social GAN [10] introduced pooling mechanisms to aggregate spatial features from nearby agents, enabling socially informed prediction. While effective in pedestrian-dense environments, these methods are less suited for structured vehicular traffic at intersections. More recent models such as Trajectron++ [21] and TPCN [30] apply graph neural networks and temporal message passing to capture dynamic inter-agent dependencies. Despite their improvements, these models often treat all agents uniformly, overlooking differences in maneuvering potential, heading direction, or traffic rules—factors especially important in intersection scenarios.

To better address vehicular dynamics, VAP-Net [26] has been proposed as an interaction pooling mechanism tailored for intersection environments. VAP-Net integrates relative position and velocity features between agents and applies an attention mechanism to dynamically modulate inter-agent influence. Specifically, it embeds each agent's relative location and speed, computes pairwise attention scores based on their motion relevance, and performs pooling to generate socially attentive interaction features. This allows the model to emphasize fast-moving or high-risk vehicles and account for diverse directional movements, which is critical in intersection scenarios with merging or crossing paths.

Nonetheless, existing interaction modules including VAP-Net remain limited in contextual expressiveness. Most fail to incorporate high-level semantics such as road geometry, traffic priority, or control states into their interaction encoding. Motivated by this gap, we propose the Multi-Dimensional Attention Pooling Network (MAP-Net), which extends prior attention-based pooling by injecting context-aware features into interaction modeling. By integrating map-aligned priors and vehicle-specific behavioral cues, MAP-Net enables nuanced representation of multi-agent dynamics under complex intersection constraints.

2.3. Context-Aware and multi-modal Forecasting

Recent advancements in trajectory forecasting increasingly leverage multi-modal context to enhance prediction accuracy beyond historical positional data alone. For instance, Gao et al. introduced VectorNet [8], a hierarchical graph neural network that encodes map and agent trajectories into vectorized representations for improved scene understanding and trajectory prediction. Similarly, Liang et al. proposed LaneGCN [15], employing lane-graph convolutional networks to capture detailed lane topologies and interactions among vehicles within structured road environments, significantly enhancing multi-modal forecasting accuracy in urban driving scenarios.

Other studies incorporate environmental semantics and intent prediction explicitly to enrich trajectory forecasting. Mangalam et al. proposed Y-Net [16], a model that integrates HD maps, agent historical motions, and intermediate intent predictions into a single goal-conditioned framework, allowing for more diverse and realistic multi-modal trajectory predictions. Moreover, Zhao et al. introduced TNT [34], a hierarchical method combining target-driven intent classification with a conditional trajectory predictor, effectively narrowing down feasible trajectory modes based on environmental context and goal likelihoods.

Despite these promising approaches, current multi-modal forecasting models often integrate context features through fixed fusion strategies or early-stage concatenation, limiting their adaptability to dynamic scene changes. Wei et al. recently introduced KI-GAN [26], which utilizes knowledge-informed generative adversarial networks to integrate traffic signal status and multi-vehicle interactions at signalized intersections, achieving notable accuracy improvements in complex scenarios. Motivated by these advances and existing limitations, our proposed EKI-GAN further extends KI-GAN by employing a modular encoder design and the Multi-dimensional Attention Pooling Network (MAP-Net), facilitating adaptive fusion of richer contextual information, vehicle dynamics, and map semantics for robust and accurate intersection trajectory forecasting.

3. Methodology

The Enhanced Knowledge-Informed Generative Adversarial Network (EKI-GAN) framework leverages adversarial training to generate multiple plausible future trajectories through sophisticated modeling of multi-modal contextual information at signalized intersections. The framework integrates six specialized encoders in a dual-stream architecture distinguishing vehicle-centric information requiring social interaction modeling from environmental factors providing contextual constraints. Beyond the framework architecture, we present three key innovations: Map Encoding (Section 3.2) that incorporates road infrastructure topology, Pedestrian-Aware Encoding (Section 3.3) that models human-vehicle interactions, and Multi-Dimensional Attention Pooling Net (Section 3.4) that captures complex vehicular interaction patterns. The encoding methods for Trajectory, Motion, Physical Attributes, and Traffic Light are adapted from KI-GAN framework[26].

3.1. Framework of EKI-GAN

The Enhanced Knowledge-Informed Generative Adversarial Network (EKI-GAN) framework is designed to accurately predict vehicular traffic patterns in complex intersection environments through sophisticated modeling of multi-modal contextual information. Figure 2 illustrates our model's architecture, where the trajectory of a single agent (highlighted in red) is comprehensively processed through multiple specialized encoders to form an enriched feature representation that captures the full spectrum of factors influencing vehicle behavior at signalized intersections.

Our framework categorizes input into two processing streams capturing vehicle movement dynamics and environmental context. This reflects the distinction between intrinsic vehicle characteristics participating in social interactions and extrinsic environmental factors providing contextual constraints. The dual-stream architecture enables realistic trajectory predictions through structured synthesis of social vehicular interactions and comprehensive environmental context.

Vehicle Information Encoders: The first category focuses on vehicle-centric information that directly participates in inter-vehicle social interactions. This includes three specialized encoders: the Trajectory Encoder processes spatial coordinates to capture movement patterns; the Motion Encoder handles dynamic vehicle states including velocity and acceleration; and the Physical Attributes Encoder encodes vehicle-specific characteristics such as type and dimensions. These three encoders form the core of vehicular representation, and their concatenated output undergoes social interaction modeling through our Multi-dimensional Attention Pooling Net (MAP-Net) to capture complex inter-vehicle dynamics.

Environmental Information Encoders: The second category encompasses contextual environmental factors that influence vehicle behavior but do not directly participate in the vehicle-to-vehicle interaction modeling. This includes the Traffic Encoder that processes traffic light states and regulatory information; the Map Encoder that incorporates road infrastructure topology from OpenStreetMap(OSM) data; and the Pedestrian Encoder that models surrounding human agents and their potential influence on vehicle trajectory planning. These environmental encodings provide crucial contextual information that is directly concatenated with the social interaction features without passing through the pooling mechanism.

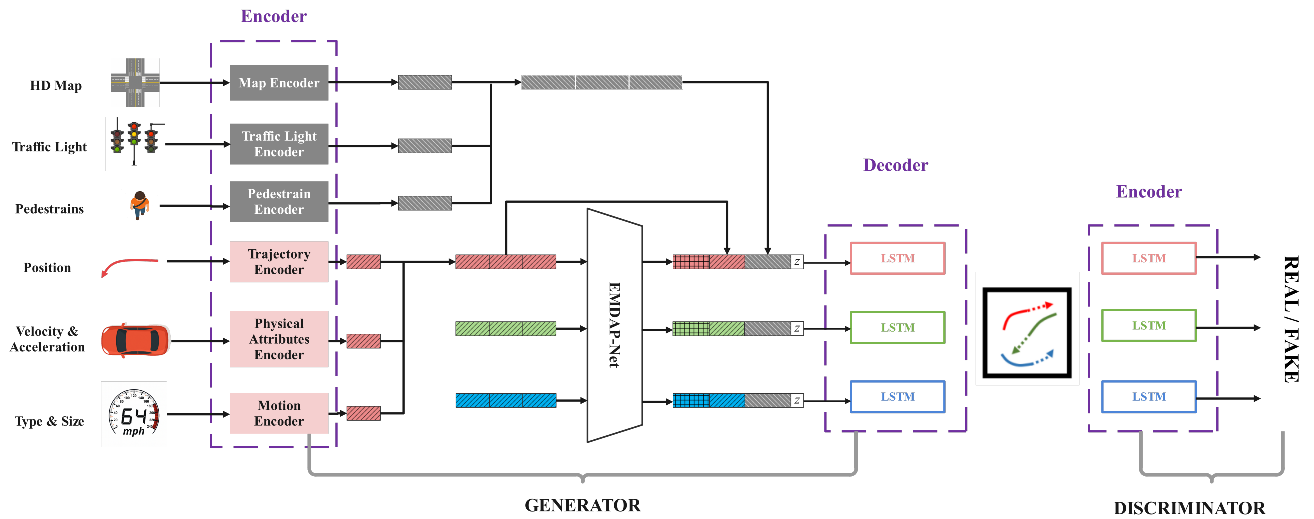


Figure 2. EKI-GAN Architecture Overview: The framework employs a dual-stream approach with six specialized encoders processing vehicle-centric and environmental information to generate socially-aware trajectory predictions at signalized intersections.

This dual-stream architecture enables the generation of realistic and socially-aware trajectory predictions through the structured synthesis of social vehicular interactions and comprehensive environmental context. The framework's strength lies in its ability to distinguish between information that requires social interaction modeling and environmental factors that provide direct contextual constraints.

3.1.1. Encoder Architecture

The generator employs six specialized encoders organized into two distinct processing streams, each designed to handle different types of input data:

Vehicle Information Encoders:

- **Trajectory Encoder:** This module focuses on spatial coordinates (x, y) of vehicles, utilizing Recurrent Neural Network (RNN) architectures to capture temporal dependencies and movement patterns in trajectory data:

$$E_{\text{traj}} = \text{RNN}_{\text{traj}}(X, Y), \quad (1)$$

where X and Y refer to vectors composed of coordinates of trajectory points.

- **Motion Encoder:** This encoder processes dynamic vehicle states, including velocity and acceleration. It employs an RNN to understand the temporal evolution of these state variables:

$$E_{\text{motion}} = \text{RNN}_{\text{state}}(V, A), \quad (2)$$

where V represents vehicle speed vectors and A denotes acceleration vectors.

- **Physical Attributes Encoder:** This component encodes vehicle-specific information including type and dimensions through separate embedding layers for categorical data and linear layers for continuous data:

$$E_{\text{phy}} = \text{Concat}(\text{Embed}_{\text{type}}(T), \text{Linear}_{\text{size}}(L, W)), \quad (3)$$

where T denotes vehicle type, and L, W indicate length and width respectively.

Environmental Information Encoders:

- **Traffic Encoder:** Dedicated to processing traffic state information, this encoder transforms discrete traffic states into higher-dimensional representations:

$$E_{\text{traffic}} = \text{Embed}_{\text{traffic}}(T_{\text{state}}), \quad (4)$$

where T_{state} signifies the traffic light status at the intersection.

- **Map Encoder:** This encoder processes road infrastructure information extracted from OSM data to provide spatial and semantic context:

$$E_{\text{map}} = \text{MapEncoder}(\mathcal{M}_{\text{local}}), \quad (5)$$

where $\mathcal{M}_{\text{local}}$ represents the local map elements around the vehicle.

- **Pedestrian Encoder:** This component models surrounding pedestrians and their potential interactions with vehicles:

$$E_{\text{ped}} = \text{PedestrianEncoder}(\mathcal{P}_{\text{obs}}), \quad (6)$$

where $\mathcal{P}_{\text{obs}}$ denotes the observed pedestrian trajectories and states.

3.1.2. Dual-Stream Social Interaction Modeling

The two processing streams are handled differently based on their roles in vehicular interaction modeling:

Vehicle Information Processing: Features from the Trajectory, Motion, and Physical Attributes Encoders are concatenated to form a combined vehicle feature vector:

$$F_{\text{vehicle}} = \text{Concat}(E_{\text{traj}}, E_{\text{motion}}, E_{\text{phy}}). \tag{7}$$

This vehicle feature vector undergoes social interaction modeling through our Multi-dimensional Attention Pooling Net (MAP-Net), which captures complex inter-vehicle dynamics and spatial relationships:

$$F_{\text{social}} = \text{MAP-Net}(F_{\text{vehicle}}, (X, Y), V), \tag{8}$$

where (X, Y) denote trajectory coordinates and V indicates speed vectors.

Environmental Information Processing: The environmental encoders provide contextual information that directly influences vehicle behavior without requiring social interaction modeling. These features are preserved independently:

$$F_{\text{env}} = \text{Concat}(E_{\text{traffic}}, E_{\text{map}}, E_{\text{ped}}). \tag{9}$$

3.1.3. Decoder and Output Generation

The decoder integrates both processing streams to generate future trajectory predictions. The comprehensive input to the decoder combines the social interaction features with the original vehicle features and environmental context:

$$F_{\text{final}} = \text{Concat}(F_{\text{social}}, F_{\text{vehicle}}, F_{\text{env}}). \tag{10}$$

The decoder, implemented through an LSTM framework, utilizes this enriched input to forecast future vehicle trajectories:

$$\text{Trajectory}_{\text{future}} = \text{LSTM}([F_{\text{final}}; z]), \tag{11}$$

where z introduces randomness for generating diverse plausible future trajectories.

This dual-stream architecture ensures that vehicle-centric information undergoes appropriate social interaction modeling while environmental factors provide direct contextual constraints, leading to more accurate and contextually-aware trajectory predictions.

3.1.4. Discriminator

The discriminator evaluates generated trajectories by comparing them against actual vehicle movements, providing feedback to improve generation quality. The discriminator's objective follows the standard GAN formulation:

$$\begin{aligned}\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] \\ + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \end{aligned} \tag{12}$$

3.2. Map Encoding

Road network topology and infrastructure elements significantly influence vehicle trajectory patterns at signalized intersections. To address this, we propose a lightweight map encoding framework that leverages OpenStreetMap (OSM) data to provide spatial and semantic context for trajectory prediction, enabling predictions that are both socially aware and environmentally consistent.

3.2.1. Vector Map Representation and Feature Extraction

We systematically parse vectorized HD Map data to extract three primary structures: **Nodes** (point locations), **Ways** (road segments), and **Relations** (complex relationships). For each way $w_i \in \mathcal{W}$, we extract geometric and semantic features:

Geometric Features: Given nodes with coordinates, we compute length as the sum of segment distances, direction as the angle between start and end points, and center as the mean coordinate position.

Semantic Features: We extract properties from OSM tags including directional constraints (one-way vs. bidirectional), normalized speed limits, road priority levels, and connectivity information.

The complete feature vector integrates both geometric and semantic attributes:

$$\mathbf{f}_i^{\text{map}} = [L_i, \theta_i, w_{\text{default}}, \mathbf{c}_i, f_{\text{direction}}, f_{\text{priority}}, f_{\text{speed}}, f_{\text{connectivity}}] \tag{13}$$

3.2.2. Spatial Context Encoding

For vehicle trajectory $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$, we encode local map context around current position $\mathbf{x}_T$. We identify the K nearest elements within radius r and compute aggregated features for fixed-size representation:

$$\mathbf{h}_{\text{spatial}} = [f_{\text{avg\_type}}, f_{\text{avg\_dist}}, f_{\text{count}}] \tag{14}$$

where features represent average element type, distance, and count within the local region.

3.2.3. Neural Map Encoding

The spatial context is processed through a learnable encoder:

$$\mathbf{h}_{\text{map}} = \mathbf{W}_{\text{map}} \mathbf{h}_{\text{spatial}} + \mathbf{b}_{\text{map}} \tag{15}$$

This approach provides efficiency through local element selection, generalizability across OSM-compatible data, and interpretability through semantic feature meanings.

3.3. Pedestrian-Aware Encoding

Human-vehicle interactions represent a critical factor in urban trajectory prediction, particularly at intersections where pedestrians and vehicles share spatial environments. We propose a hierarchical pedestrian encoding framework that models dynamic pedestrian behavior and quantifies their influence on vehicle trajectory planning.

3.3.1. Pedestrian State Representation

For each scenario, we consider up to N_{ped} pedestrians, where each pedestrian p_j is characterized by temporal state vectors. The pedestrian data is structured as collections of trajectory sequences, with each state vector encompassing positional coordinates, velocity components, and acceleration information over the observation window.

3.3.2. Multi-Modal Feature Extraction

We employ dual-branch encoding that separately processes spatial trajectory and kinematic dynamics:

Trajectory Branch: Encodes spatial movement patterns through dedicated LSTM networks that process embedded coordinate sequences, capturing long-term spatial patterns and directional intentions.

Kinematic Branch: Processes velocity and acceleration information through separate LSTM encoders focused on dynamic movement characteristics.

The trajectory and kinematic features are fused through multi-layer perceptrons:

$$\mathbf{h}_j^{\text{ped}} = \text{MLP}_{\text{fusion}}([\mathbf{h}_j^{\text{traj}}, \mathbf{h}_j^{\text{state}}]) \tag{16}$$

3.3.3. Hierarchical Attention Mechanism

We implement attention to identify influential pedestrians by computing mean pedestrian features and applying attention MLPs to generate normalized attention weights:

$$\mathbf{h}_{\text{pedestrian}} = \sum\nolimits_{j=1}^{N_{\text{active}}} \alpha_j \mathbf{h}_j^{\text{ped}} \tag{17}$$

where α_j represents the attention weight for pedestrian j, computed through softmax normalization of attention scores.

This framework provides comprehensive modeling through dual-branch architecture, selective attention for relevant pedestrians, and scalability across varying pedestrian numbers.

3.4. Multi-Dimensional Attention Pooling Net (MAP-Net)

Traditional attention pooling mechanisms inadequately capture the multi-faceted nature of vehicular interactions in complex traffic scenarios. MAP-Net extends conventional frameworks by incorporating richer feature representations and enhanced numerical stability to achieve robust trajectory prediction.

3.4.1. Multi-Dimensional Feature Extraction

MAP-Net extracts comprehensive interaction features beyond simple spatial proximity:

Multi-Dimensional Interaction Features For vehicle pairs i and j, we compute distance-based features through learnable embeddings of Euclidean distances, directional features using trigonometric encoding of relative orientations, and velocity differential features capturing relative motion patterns. These are combined with standard spatial and velocity embeddings to form enhanced feature representations:

$$\mathbf{F}_{ij}^{\text{enhanced}} = [\mathbf{E}_{ij}; \mathbf{V}_i; \mathbf{F}_{ij}^{\text{dist}}; \mathbf{F}_{ij}^{\text{dir}}; \mathbf{F}_{ij}^{\text{vel\_diff}}] \tag{18}$$

3.4.2. Enhanced Attention and Adaptive Gating

Stabilized Attention Mechanism To ensure numerical stability, we employ clamped softmax operations that restrict attention logits to stable numerical ranges:

$$\mathbf{W}_{ij} = \frac{\exp(\text{clamp}(f_{\text{attention}}(\mathbf{F}_{ij}^{\text{enhanced}})))}{\sum_{k \in \mathcal{N}_i} \exp(\text{clamp}(f_{\text{attention}}(\mathbf{F}_{ik}^{\text{enhanced}}))) + \epsilon} \tag{19}$$

Context-Aware Gating Adaptive gating enables dynamic emphasis between social interactions and individual dynamics based on vehicle hidden states:

$$\tilde{\mathbf{W}}_{ij} = \mathbf{W}_{ij} \odot \sigma(f_{\text{gate}}(\mathbf{H}_i)) \tag{20}$$

The vehicle hidden states are modulated by normalized gated attention weights to create context-aware representations.

3.4.3. Contextual Pooling

The final pooled representation integrates weighted states with relative position embeddings through max-pooling operations:

$$\mathbf{MAPF}_i = \max\left(f_{\text{pool}}\left([\mathbf{E}_{ij}; \mathbf{H}_i \odot \tilde{\mathbf{W}}_{ij}]\right)\right) \tag{21}$$

Through multi-dimensional feature extraction, stabilized attention computation, and adaptive gating mechanisms, MAP-Net delivers robust agent interaction representation that improves prediction accuracy while maintaining computational efficiency.

4. Experiments

4.1. Dataset

Existing trajectory datasets can be categorized by their scope and focus. Pedestrian datasets such as ETH [19] and UCY [14] capture human movement patterns, while single-vehicle datasets like Argoverse [5] and nuScenes [4] focus on individual vehicle behavior. Multi-vehicle datasets including INTERACTION [32] and inD [2] capture vehicle interactions but lack comprehensive contextual information. These datasets fail to simultaneously provide the detailed vehicle characteristics (type, dimensions, velocity, acceleration), pedestrian dynamics, traffic signal states, and high-definition map topology required for intersection trajectory prediction.

After evaluating available datasets [2, 3, 5, 13, 23, 27, 32], we selected the Signalized INtersection Dataset (SinD) [28] for our experiments. SinD captures data from a two-phase signalized intersection in Tianjin, China, using drone footage to ensure naturalistic behavior. The dataset contains over 13,000 traffic participants across seven categories (cars, buses, trucks, motorcycles, bicycles, tricycles, and pedestrians) recorded over seven hours. SinD provides comprehensive information including vehicle specifications, motion parameters, traffic light states with timing data, and high-definition map topology. The dataset features 62.6% vulnerable road users and documents traffic violations, making it ideal for evaluating complex vehicle-pedestrian interactions at signalized intersections.

4.2. Experimental Setup and Evaluation Metrics

We processed the SinD dataset by downsampling from its original 30 Hz capture frequency to 2 Hz using a step interval of 15 frames. This temporal reduction addresses potential data bias arising from minimal vehicle displacement during traffic signal-induced stops, which could adversely impact the model's capacity to learn meaningful motion patterns. From the 23 available segments in the SinD dataset, we allocated 20 segments for training and 3 segments for validation to ensure robust model evaluation. Our experimental design employs a 6-second observation window (12 frames) followed by prediction horizons of 6 seconds (12 frames) and 9 seconds (18 frames), enabling comprehensive evaluation across different temporal scales.

The training configuration utilizes a batch size of 64 across 100 training epochs. We employed differentiated learning rates with the Generator set to 0.0005 and the Discriminator configured at 0.0005 to ensure stable adversarial training dynamics. Model performance evaluation relies on two standard trajectory prediction metrics: Average Displacement Error (ADE) and Final Displacement Error (FDE), formulated as:

$$\mathrm{ADE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sqrt{(x_t^i-\hat{x}_t^i)^2+(y_t^i-\hat{y}_t^i)^2}, \tag{22}$$

$$\mathrm{FDE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_T^i-\hat{x}_T^i)^2+(y_T^i-\hat{y}_T^i)^2}, \tag{23}$$

where N represents the trajectory count, T denotes the prediction timeframe, (x_t^i, y_t^i) indicates actual position coordinates, and $(\hat{x}_t^i, \hat{y}_t^i)$ represents predicted coordinates at timestep t for trajectory i.

Table 1. Evaluation Metrics for Trajectory Prediction for SinD dataset (underlined values indicate best performance)

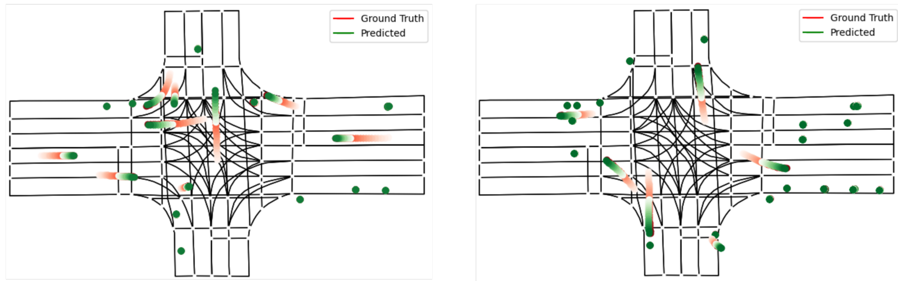
Method	12-12		12-18	
	ADE	FDE	ADE	FDE
S-GAN	1.32	2.46	1.53	2.95
S-LSTM	0.87	1.60	0.96	1.78
Trajetron++	0.37	0.93	0.70	1.91
FJMP	0.27	0.68	0.41	1.13
HP-Net	0.09	0.21	0.11	0.25
KI-GAN	<u>0.05</u>	0.12	0.11	0.26
I2XTraj	0.09	0.12	0.10	0.21
EKI-GAN (Ours)	<u>0.05</u>	<u>0.11</u>	<u>0.09</u>	<u>0.19</u>

4.3. Quantitative Analysis

Table 1 presents the quantitative evaluation results of our EKI-GAN framework compared against state-of-the-art trajectory prediction methods on the SinD dataset. The overall low ADE and FDE values across all methods can be attributed to the nature of intersection prediction, where vehicles frequently remain stationary due to traffic signal constraints, resulting in near-zero prediction errors for stopped vehicles that significantly reduce the average metrics. Our EKI-GAN achieves state-of-the-art performance across both evaluation scenarios (12-12 and 12-18). In the shorter prediction horizon (12-12), EKI-GAN demonstrates competitive performance with KI-GAN in ADE while achieving the best FDE of 0.11, representing an improvement albeit modest due to the already low baseline values. The more substantial improvements are evident in the longer prediction horizon (12-18), where EKI-GAN outperforms all competing methods. Notably, compared to the previous state-of-the-art I2XTraj method [31], our approach achieves a 10% reduction in ADE (from 0.10 to 0.09) and a 9.5% reduction in FDE (from 0.21 to 0.19), demonstrating the effectiveness of our enhanced multi-modal encoding and attention mechanisms for extended trajectory forecasting.

4.4. Qualitative Analysis

The visual evaluation presented in Figure 3 showcases representative prediction examples from our validation dataset. The visualization includes six distinct scenarios, with the upper row demonstrating 12-frame observation and 12-frame prediction configurations, while the lower row presents 12-frame observation with 18-frame prediction settings. Our EKI-GAN model exhibits exceptional alignment between forecasted paths and actual vehicle movements across diverse traffic conditions, ranging from low-density scenarios with minimal vehicle interactions to complex high-density situations involving multiple simultaneous maneuvers. This visual coherence validates the quanti-



(a) 12-Frame Observation with 12-Frame Prediction

(b) 12-Frame Observation with 18-FramePrediction

Figure 3. Trajectory Predictions by EKI-GAN on SinD Dataset: Visual Comparison of Ground Truth and Model Predictions. Light red to dark red dots represent observed and predicted ground truth trajectories, while light green to dark green dots illustrate EKI-GAN predictions, demonstrating temporal progression and prediction accuracy across different forecasting horizons.

tative metrics and demonstrates the model's robust capability to handle varying traffic complexities while maintaining prediction accuracy across different temporal horizons.

4.5. Ablation Study

Our ablation study systematically evaluates each encoder component's contribution through selective removal experiments. Table 2 demonstrates that the Traffic Light Encoder has the most substantial impact, with its removal causing dramatic performance degradation—ADE increasing by 720% in the 12-12 scenario. This reflects traffic signals' fundamental role in governing vehicle behavior at signalized intersections. The Map Encoder shows modest influence, with removal resulting in 22% ADE increase for 12-18 prediction, reflecting the single-intersection nature of SinD dataset where positional encoding can implicitly acquire spatial semantics.

Table 2. Impact of Encoder Components (A: Trajectory Encoder, B: Motion Encoder, C: Physical Attributes Encoder, D: Traffic Light Encoder, E: Map Encoder, F: Pedestrian Encoder)

A	B	C	D	E	F	12-12		12-18	
						ADE	FDE	ADE	FDE
✓	×	✓	✓	✓	✓	0.09	0.23	0.17	0.45
✓	✓	×	✓	✓	✓	0.07	0.15	0.14	0.33
✓	✓	✓	×	✓	✓	0.41	0.97	0.63	1.50
✓	✓	✓	✓	×	✓	0.05	0.11	0.11	0.24
✓	✓	✓	✓	✓	×	0.07	0.13	0.13	0.31
✓	✓	✓	✓	✓	✓	0.05	0.11	0.09	0.19

Table 3 validates our MAP-Net's effectiveness compared to existing pooling approaches. MAP-Net consistently achieves the lowest prediction errors, demonstrating 8.3% FDE improvement over VAP-Net (from 0.12 to 0.11) in 12-12 scenario and 10% ADE improvement (from 0.10 to 0.09) in 12-18 scenario. Performance gains over traditional methods are more substantial, with 44% FDE reduction compared to Social Pool Net (from 0.16 to 0.11), demonstrating our multi-dimensional attention mechanism's superior capability in capturing complex vehicle interactions.

Table 3. Comparison of Pooling Methodologies

Pooling Method	**12-12**		12-18	
	ADE	FDE	ADE	FDE
Social Pool Net	0.07	0.16	0.15	0.38
Hidden Pool Net	0.07	0.19	0.17	0.44
VAP Net	0.05	0.12	0.10	0.25
MAP Net(Ours)	0.05	0.11	0.09	0.19

5. Conclusions

This paper presents the Enhanced Knowledge-Informed Generative Adversarial Network (EKI-GAN), a comprehensive framework for vehicle trajectory prediction at signalized intersections that addresses critical limitations in existing approaches through enhanced multi-modal integration. Our dual-stream architecture with six specialized encoders effectively captures both vehicle factors and environmental information, while the proposed Multi-Dimensional Attention Pooling Network (MAP-Net) enables sophisticated modeling of complex vehicular interactions through multi-dimensional feature extraction and adaptive attention mechanisms. Extensive experimental validation on the SinD dataset demonstrates state-of-the-art performance, establishing new benchmarks for intersection trajectory forecasting accuracy.

Future research directions include exploring the framework's generalizability across diverse geographical regions and extending EKI-GAN to handle more complex traffic scenarios, including congested highway environments, highway on-ramp merging scenarios, and stop sign-controlled intersections. These extensions would further advance the framework's capability to handle the full spectrum of urban and highway traffic prediction challenges.

References

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016. 2

[2] Julian Bock, Robert Krajewski, Tobias Moers, Steffen Runde, Lennart Vater, and Lutz Eckstein. The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1929–1934. IEEE, 2020. 6, 7

[3] Antonia Breuer, Jan-Aike Termöhlen, Silviu Homoceanu, and Tim Fingscheidt. opendd: A large-scale roundabout drone dataset. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2020. 7

[4] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 6

[5] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8748–8757, 2019. 6, 7

[6] Chen Chen, Lei Liu, Tie Qiu, Zhiyuan Ren, Jinna Hu, and Fang Ti. Driver's intention identification and risk evaluation at intersections in the internet of vehicles. *IEEE Internet of Things Journal*, 5(3):1575–1587, 2018. 2

[7] Hongbo Gao, Hang Su, Yingfeng Cai, Renfei Wu, Zhengyuan Hao, Yongneng Xu, Wei Wu, Jianqing Wang, Zhijun Li, and Zhen Kan. Trajectory prediction of cyclist based on dynamic bayesian network and long short-term memory model at unsignalized intersections. *Science China Information Sciences*, 64(7):172207, 2021. 2

[8] Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11525–11533, 2020. 3

[9] 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. 2

[10] 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. 2

[11] Xin Huang, Stephen G McGill, Brian C Williams, Luke Fletcher, and Guy Rosman. Uncertainty-aware driver trajectory prediction at urban intersections. In *2019 International conference on robotics and automation (ICRA)*, pages 9718–9724. IEEE, 2019. 2

[12] A. Kawasaki and T. Tasaki. Trajectory prediction of turning vehicles based on intersection geometry and observed velocities. *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 511–516, 2018. 2

[13] Robert Krajewski, Tobias Moers, Julian Bock, Lennart Vater, and Lutz Eckstein. The round dataset: A drone dataset of road user trajectories at roundabouts in germany. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2020. 7

[14] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. In *Computer graphics forum*, pages 655–664. Wiley Online Library, 2007. 6

[15] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. Learning lane graph representations for motion forecasting. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 541–556. Springer, 2020. 3

[16] Karttikeya Mangalam, Yang An, Harshayu Girase, and Jitendra Malik. From goals, waypoints & paths to long term human trajectory forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15233–15242, 2021. 3

[17] Xiaoyu Mo, Zhiyu Huang, Yang Xing, and Chen Lv. Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):9554–9567, 2022. 1

[18] Geunseob Oh and H. Peng. Impact of traffic lights on trajectory forecasting of human-driven vehicles near signalized intersections. *arXiv: Robotics*, 2019. 2

[19] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, pages 261–268. IEEE, 2009. 1, 6

[20] Luke Rowe, Martin Ethier, Eli-Henry Dykhne, and Krzysztof Czarnecki. Fjmp: Factorized joint multi-agent motion prediction over learned directed acyclic interaction graphs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13745–13755, 2023. 2

[21] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco 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*, pages 683–700. Springer, 2020. 2

[22] Chaopeng Tan, Nan Zhou, Fen Wang, Keshuang Tang, and Yangbeibei Ji. Real-time prediction of vehicle trajectories for proactively identifying risky driving behaviors at high-speed intersections. *Transportation research record*, 2672 (38):233–244, 2018. 2

[23] Hong Wang, Amir Khajepour, Dongpu Cao, and Teng Liu. Ethical decision making in autonomous vehicles: Challenges

and research progress. *IEEE Intelligent Transportation Systems Magazine*, 14(1):6–17, 2020. 7

[24] Shupei Wang, Ziyang Wang, Rui Jiang, Ruidong Yan, and Lei Du. Trajectory jerking suppression for mixed traffic flow at a signalized intersection: A trajectory prediction based deep reinforcement learning method. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):18989–19000, 2022. 2

[25] Chuheng Wei, Ziye Qin, Guoyuan Wu, Matthew J Barth, Amr Abdelraouf, Rohit Gupta, and Kyungtae Han. Dilemma zone: A comprehensive study of influential factors and behavior analysis. In *2024 Forum for Innovative Sustainable Transportation Systems (FISTS)*, pages 1–8. IEEE, 2024. 1

[26] Chuheng Wei, Guoyuan Wu, Matthew J Barth, Amr Abdelraouf, Rohit Gupta, and Kyungtae Han. Ki-gan: Knowledge-informed generative adversarial networks for enhanced multi-vehicle trajectory forecasting at signalized intersections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7115–7124, 2024. 1, 2, 3

[27] Chuheng Wei, Ziye Qin, Siyan Li, Ziyan Zhang, Xuanpeng Zhao, Amr Abdelraouf, Rohit Gupta, Kyungtae Han, Matthew J Barth, and Guoyuan Wu. Pdb: Not all drivers are the same–a personalized dataset for understanding driving behavior. *arXiv preprint arXiv:2503.06477*, 2025. 7

[28] Yanchao Xu, Wenbo Shao, Jun Li, Kai Yang, Weida Wang, Hua Huang, Chen Lv, and Hong Wang. Sind: A drone dataset at signalized intersection in china. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2471–2478. IEEE, 2022. 7

[29] Bo Yang, Song Yan, Zheng Wang, and Kimihiko Nakano. Prediction based trajectory planning for safe interactions between autonomous vehicles and moving pedestrians in shared spaces. *IEEE Transactions on Intelligent Transportation Systems*, 24(10):10513–10524, 2023. 1

[30] Maosheng Ye, Tongyi Cao, and Qifeng Chen. Tpcn: Temporal point cloud networks for motion forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11318–11327, 2021. 2

[31] Huilin Yin, Yangwenhui Xu, Jiaxiang Li, Hao Zhang, and Gerhard Rigoll. Knowledge-informed multi-agent trajectory prediction at signalized intersections for infrastructure-to-everything. *arXiv preprint arXiv:2501.13461*, 2025. 7

[32] Wei Zhan, Liting Sun, Di Wang, Haojie Shi, Aubrey Clausse, Maximilian Naumann, Julius Kummerle, Hendrik Konigshof, Christoph Stiller, Arnaud de La Fortelle, et al. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. *arXiv preprint arXiv:1910.03088*, 2019. 6, 7

[33] Yuzhen Zhang, Wentong Wang, Weizhi Guo, Pei Lv, Mingliang Xu, Wei Chen, and Dinesh Manocha. D2-tpred: Discontinuous dependency for trajectory prediction under traffic lights. In *European Conference on Computer Vision*, pages 522–539. Springer, 2022. 1, 2

[34] Hang Zhao, Jiyang Gao, Tian Lan, Chen Sun, Ben Sapp, Balakrishnan Varadarajan, Yue Shen, Yi Shen, Yuning Chai, Cordelia Schmid, et al. Tnt: Target-driven trajectory prediction. In *Conference on Robot Learning*, pages 895–904. PMLR, 2021. 3