KI-GAN: Knowledge-Informed Generative Adversarial Networks for Enhanced Multi-Vehicle Trajectory Forecasting at Signalized Intersections

Chuheng Wei[1]   Guoyuan Wu[1]   Matthew J. Barth[1]   Amr Abdelraouf[2]

Rohit Gupta[2]   Kyungtae Han[2]

[1] University of California Riverside,   [2] InfoTech Labs, Toyota Motor North America

chuheng.wei@email.ucr.edu, gywu@cert.ucr.edu, barth@ece.ucr.edu,

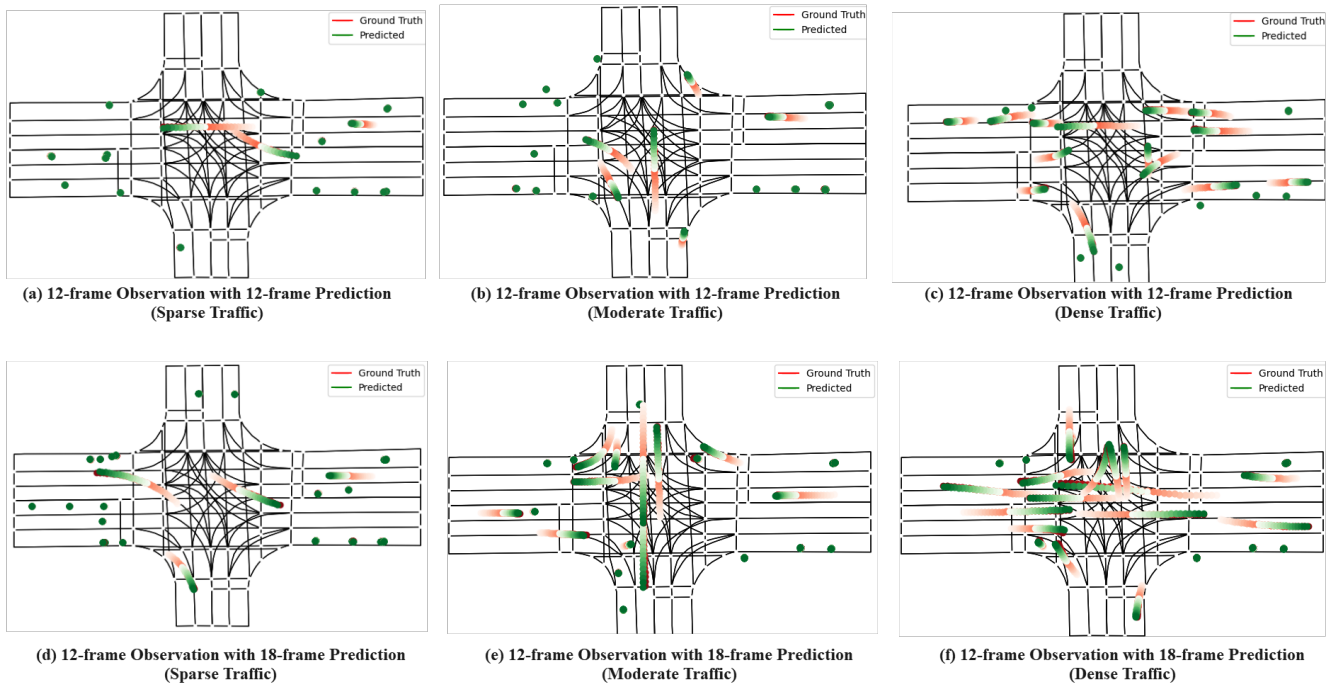{amr.abdelraouf, rohit.gupta, kyungtae.han}@toyota.com

Figure 1. Trajectory Predictions by KI-GAN on SinD Dataset: Comparative Visualization of Ground Truth and Model Predictions. (Light red to dark red dots represent the trajectory's ground truth for both observed and predicted frames. Light green to dark green dots illustrate the KI-GAN model's predictions, highlighting accuracy and temporal progression.)

Abstract

Reliable prediction of vehicle trajectories at signalized intersections is crucial to urban traffic management and autonomous driving systems. However, it presents unique challenges, due to the complex roadway layout at intersections, involvement of traffic signal controls, and interactions among different types of road users. To address these issues, we present in this paper a novel model called Knowledge-Informed Generative Adversarial Network (KI-GAN), which integrates both traffic signal information and multi-vehicle interactions to predict vehicle trajectories accurately. Additionally, we propose a specialized attention pooling method that accounts for vehicle orientation and proximity at intersections. Based on the SinD dataset, our KI-GAN model is able to achieve an Average Displacement Error (ADE) of 0.05 and a Final Displacement Error (FDE) of 0.12 for a 6-second observation and 6-second prediction cycle. When the prediction window is extended to 9 seconds, the ADE and FDE values are further re-

duced to 0.11 and 0.26, respectively. These results demonstrate the effectiveness of the proposed KI-GAN model in vehicle trajectory prediction under complex scenarios at signalized intersections, which represents a significant advancement in the target field. Code is available at https : //github.com/ChuhengWei/KI_GAN.

1. Introduction

In the era of urbanization, traffic management at signalized intersections has become increasingly crucial in order to maintain safety [46]. These intersections serve as epicenters of vehicular movement, wherein the confluence of diverse road users and their interactions exacerbate traffic flow complexity [49]. Multiple vehicle trajectory forecasting in such environments is not only crucial for efficient traffic management systems, but also indispensable for reliable and effective autonomous driving [36]. In this study, we introduce the *Knowledge-Informed Generative Adversarial Network* (KI-GAN), a model specifically designed for predicting vehicle trajectories at signalized intersections. It is imperative to develop such a model to address the inherent challenges presented by urban traffic environments, such as intricate roadway layouts at intersections and traffic conditions and signal operations that are quite dynamic in nature.

1.1. Background

Trajectory prediction is a vital component in the development of efficient and safe transportation systems [36]. In dynamic and interactive environments, forecasting the future path of vehicles based on their current and past states becomes increasingly challenging [34]. In the existing trajectory prediction research, linear scenarios, such as highways or non-interactive environments, have been primarily examined [4, 9, 44]. However, with the growing complexity of urban traffic, particularly at intersections, the need for more sophisticated models has become evident [36]. As opposed to highways and rural thoroughfares, intersections are characterized by their multidirectional flow patterns, divergent vehicular behavior, and traffic control signals, which distinguish them from other junctions [11, 48].

In multi-vehicle scenarios, the interaction between vehicles plays a crucial role [2, 23, 44]. Each vehicle's movement can influence the path of others, creating a dynamic system that is challenging to predict. In addition to physical interactions, drivers are also influenced by their perceptions, decisions, and reactions to surrounding environment, including roadway geometry[30], other vehicles [38], pedestrians [18], and traffic signs and signals [22]. These elements contribute to a certain level of complexity seldom encountered in simpler traffic scenarios. The status of traffic lights, for example, exerts a profound influence on

driving behavior, prompting actions like acceleration, deceleration, or directional changes [13, 41]. Consequently, anticipating vehicular paths in such environments becomes a notably intricate endeavor. Moreover, the heterogeneity in vehicle types and sizes introduces additional complexity layers [38]. For instance, larger vehicles like trucks exhibit distinct maneuvering and acceleration characteristics compared to smaller vehicles [35], underscoring the need for a predictive model capable of assimilating and interpreting these diverse vehicular dynamics for precise trajectory forecasting.

In response to the complexities of intersection scenarios, our research integrates a wealth of intersection-specific information into our trajectory prediction model. We focus on the interactional layers by modeling essential vehicle characteristics such as speed, size, type, and their dynamics, including acceleration. Simultaneously, we incorporate a traffic light model, vital in influencing driver decisions and maneuvers. Additionally, recognizing the distinct nature of vehicle interactions at intersections, we introduce an innovative pooling method designed to capture the unique behavioral patterns of vehicles in these settings. This dual approach of detailed vehicle modeling and advanced interaction techniques marks a significant step towards more accurately predicting vehicle trajectories in the challenging environment of urban intersections.

1.2. Contributions

Addressing the challenges of the trajectory prediction at signalized intersection, our paper contributes significantly to the following fields:

Introduction of KI-GAN: Our primary contribution is the development of Knowledge-Informed Generative Adversarial Networks (KI-GAN), a model that stands out for its ability to integrate diverse and critical data sources. By assimilating diverse vehicle data, traffic signal information, and multi-vehicle interactions, KI-GAN offers a comprehensive understanding of the factors influencing vehicle trajectories at intersections. The core of its innovation lies in the multi-module encoder, a feature that allows the model to process and interpret various types of input data effectively. This encoder is essential for capturing the complex, multi-layered interactions and conditions typical of urban intersections, thereby enabling the model to predict vehicle trajectories with a higher degree of accuracy and reliability.

Specialized Attention Pooling Method: Complementing the KI-GAN, we introduce an advanced attention pooling method specifically designed for analyzing vehicle orientation and proximity in intersection scenarios. This method, known as Vehicle Attention Pooling Net (VAP-Net), marks a significant improvement over traditional interaction modules used in trajectory prediction models. It is adept at discerning the subtle yet critical aspects of vehi-

cle behavior and interaction, such as speed and orientation towards other vehicles and response to nearby traffic elements. Our empirical results demonstrate the superiority of VAP-Net in accurately capturing and predicting the nuanced movements of vehicles in complex urban intersection environments.

2. Related Work

In this review, we focus on trajectory prediction methods tailored for the unique dynamics of signalized intersections and the integral multi-vehicle interactions. Signalized intersections require sophisticated prediction methods due to variables like traffic signals and vehicle maneuvers, while interaction-based multi-vehicle predictions are crucial for a realistic depiction of collective traffic behavior.

2.1. Trajectory Prediction Models for Intersection Scenarios

Some researchers have attempted to tailor deep learning algorithms specifically for trajectory prediction at signalized intersections. Abdelraouf et al. [1] used sequence-to-sequence Long Short-Term Memory (LSTM) networks for intersection-wide vehicle trajectory prediction. Their proposed algorithm predicted future vehicle positions and orientations. Kawasaki and Tasaki [18] designed a model specifically for predicting turning trajectories at intersections, considering the geometry of the crossing and the speeds of observed vehicles. Cao et al. [7] developed a vision-based intersection trajectory prediction model that predicts vehicle maneuvers such as left turns, right turns, or going straight, aiming to prevent traffic collisions. While these models mark progress in intersection-specific prediction, they often overlook the critical role of traffic signals.

As some trial efforts, Roy et al. [38] incorporated vehicle size in their (GAN)-based approach for both signalized and non-signalized intersections, but without directly considering traffic signal status. Oh and Peng [32] focused on the impact of traffic lights on vehicle speed rather than trajectory variation. Zhang et al.'s D2-TPred model [46], using a spatial dynamic interaction graph (SDG) and a behavior dependency graph (BDG), is a notable exception, as it considered traffic light-induced trajectory discontinuities. However, these models still tend to overlook a comprehensive view of vehicle characteristics in their interaction-based trajectory predictions.

In conclusion, while there has been significant progress in trajectory prediction at intersections, existing models often fall short in simultaneously addressing the complex vehicle interactions and the impact of traffic signals. This gap highlights the need for a comprehensive approach like the proposed KI-GAN model, which integrates these critical aspects for enhanced trajectory forecasting at signalized intersections.

2.2. Multi-Vehicle Trajectory Prediction Considering Interactions

Forecasting trajectories at intersections is complex due to significant vehicle interactions and dense traffic conditions. This complexity is mirrored in pedestrian trajectory prediction research [3, 15, 20], where the interplay between multiple agents has also been recognized as a pivotal element. In interactive scenarios, vehicle accidents are riskier than pedestrian incidents, requiring analysis of more complex data including vehicle size, type, and speed, along with historical trajectory Early studies, like those by Pecher et al. [33], recognized the importance of considering neighboring vehicles' trajectories, laying the groundwork for more complex interaction models. Building on this, Lin et al. [24] introduced the concept of Spatio-Temporal Attention in their Long Short-Term Memory (STA-LSTM) model to better understand the influence of historical and nearby vehicle trajectories. Similarly, Mo et al. [29] explored the social interaction behaviors among vehicles, especially on highways, highlighting how the number of vehicles can influence behaviors and trajectories.

In the realm of multi-vehicle interactions, models have become increasingly sophisticated. Salzmann et al. [39] introduced the modular, graph-structured Trajectron++ model, which predicts the trajectories of multiple agents by integrating agent dynamics and semantic maps. Subsequently, other scholars [25–27, 31, 37] advanced this by predicting multiple potential behaviors and future positions of surrounding vehicles, using convolutional neural networks and grid map predictions, respectively.

However, most of these studies did not fully account for the unique complexities at intersections, such as the intricacies of roadway layout, vehicle orientation, and diverse vehicle types. These factors necessitate a multi-parameter model to capture the intricate dynamics at intersections, where the influence of nearby vehicles varies based on proximity and speed.

2.3. Research Gaps

Based on a comprehensive review of existing research, we have found the following key research gaps in trajectory prediction for intersections:

Traffic Light Influence: Most current models do not adequately factor in the influence of traffic light status on vehicle behavior at intersections. Traffic light status has been shown to have significant effect over driving behaviour at intersections [17]. This oversight can lead to significant inaccuracies in trajectory prediction.

Multidimensional Data Integration: Traditional models often rely on simplistic representations of vehicle interactions, primarily based on positional data. However, at intersections, the integration of additional data types such as vehicle speed [28], size [10], type [38], and traffic signal
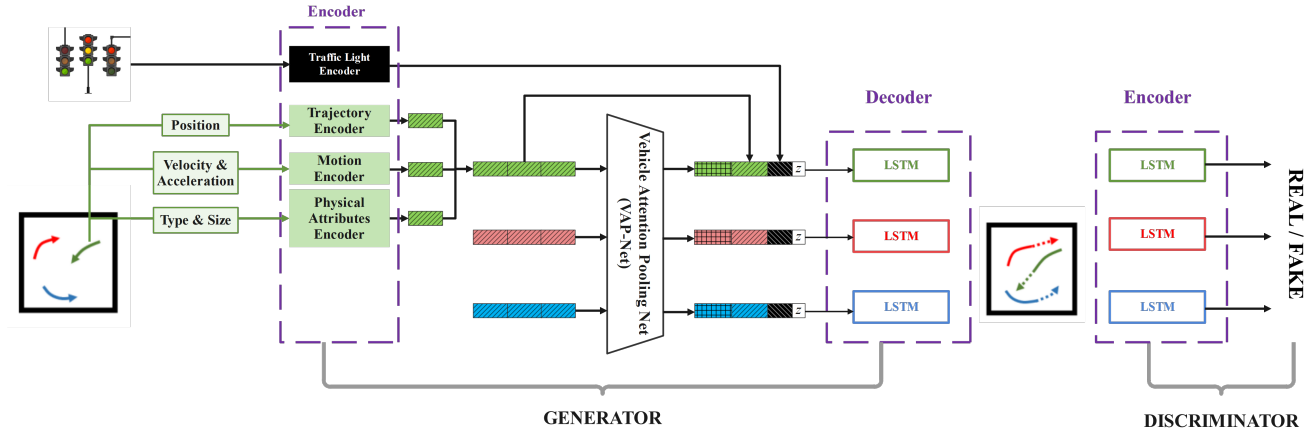
Figure 2. Architecture of Knowledge-Informed Generative Adversarial Networks

status [46] is crucial for a realistic and accurate prediction model.

Advanced Interaction Pooling: Existing models lack a specialized mechanism to handle the complex interactions at intersections. A new approach that can effectively pool interaction data, taking into account the diverse factors influencing vehicle behavior at intersections, is needed.

3. Methodology

3.1. Framework of KI-GAN

The Knowledge-Informed Generative Adversarial Network (KI-GAN) framework is presented in this section, a sophisticated model designed to accurately predict vehicular traffic patterns. Figure 2 illustrates our model's architecture, where the trajectory of a single agent (highlighted in green) is fully processed through multiple encoders to form a combined feature vector. It should be noted that the construction of the combined feature vector for other agents in the scene adheres to the same methodology, ensuring a unified approach to trajectory prediction across all agents. As illustrated in the figure, our framework employs a multi-encoder approach, integrating diverse data streams to capture the complex dynamics of vehicle movement and interaction. This framework allows the generation of realistic and socially-aware trajectory predictions through the structured synthesis of spatial, vehicular, and traffic information.

3.1.1 Encoders

The generator consists of four specialized encoders: Trajectory Encoder, Physical Attributes Encoder, Motion Encoder, and Traffic Encoder, each designed to process distinct types of input data.

- **Trajectory Encoder:** This module focuses on spatial coordinates (x, y) of vehicles. Using Recurrent Neural Net-

work (RNN) architectures, it captures temporal dependencies and movement patterns in trajectory data, offering insight into the vehicles' historical paths.

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

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

- **Motion Encoder:** This encoder deals with dynamic vehicle states, including velocity and acceleration. Similar to the Trajectory Encoder, it also utilizes an RNN to understand the temporal evolution of these state variables, providing a real-time context to the vehicle's movement.

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

where V stands for a vector composed of vehicle speeds, and A represents a vector composed of vehicle accelerations.

- **Physical Attributes Encoder:** This part of the Generator encodes vehicle-specific information such as type and dimensions. By employing separate embedding layers for categorical data (like vehicle type) and linear layers for continuous data (dimensions), it creates a comprehensive feature vector for each vehicle.

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

where T denotes the type of the vehicle, including cars, trucks, buses, etc., and L and W indicate the length and width of the vehicle from a top-down perspective.

- **Traffic Encoder:** Dedicated to processing the traffic state information, this encoder uses an embedding layer to transform discrete traffic states into a higher-dimensional space. For example, we encode the status of traffic lights at intersections in the SinD dataset [42] using numbers 1

to 5 as shown in Table 1. This transformation allows the model to encode and interpret various traffic conditions effectively.

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

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

Table 1. Traffic light status in each direction and the corresponding digital codes (The SinD dataset features an intersection with a single traffic light per direction, lacking dedicated left-turn signals, resulting in a unified signal state per direction. Moreover, the east-west and north-south traffic light states are synchronized respectively.)

Direction				
North	South	East	West	Code
Green	Green	Red	Red	1
Yellow	Yellow	Red	Red	2
Red	Red	Red	Red	3
Red	Red	Green	Green	4
Red	Red	Yellow	Yellow	5

3.1.2 Social Interaction

After encoding the different data streams, the features from the Trajectory, Motion, and Physical Attributes Encoders are concatenated to form a combined feature vector.

$$F_{\text{combined}} = \text{Concat}(E_{\text{traj}}, E_{\text{motion}}, E_{\text{phy}}). \qquad (5)$$

This vector, representing each trajectory, is then fed into a pooling layer along with features from other trajectories. The pooling layer aggregates contextual information from neighboring trajectories, enabling the model to understand the social and spatial interactions among vehicles. For the pooling layer, we propose a **Vehicle Attention Pooling Net (VAP-Net)**, which is designed to effectively calculate and interpret potential interactions and interrelations between vehicle trajectories.

$$P = \text{VAPNet}(F_{\text{combined}}, (X, Y), V), \qquad (6)$$

where (X, Y) denote vectors composed of trajectory coordinates, and V indicates a vector composed of speeds.

3.1.3 Decoder

The output of the pooling layer, when amalgamated with the encoded traffic state and the combined feature vector,

constitutes the comprehensive input to the decoder. This integration is mathematically represented as:

$$F_{\text{recombined}} = \text{Concat}(P, F_{\text{combined}}, E_{\text{traffic}}). \qquad (7)$$

The decoder is realized through an LSTM [16] framework, which utilizes the enriched input $F_{\text{recombined}}$ to forecast future vehicle trajectories. The LSTM, with its recurrent architecture, is adept at generating sequences that are temporally consistent and socially acceptable by reflecting on historical data, present ego-vehicle states, and the surrounding traffic conditions[16]. This dynamic is encapsulated in the equation:

$$\text{Trajectory}_{\text{future}} = \text{LSTM}([F_{\text{recombined}}; z]). \qquad (8)$$

Here, z introduces a degree of randomness, enabling the model to generate a spectrum of possible future trajectories, contributing to its robustness. As a result of the blending of these multifaceted data streams, the model offers a comprehensive insight into vehicular dynamics, paving the way for reliable and precise predictions. This model's versatility is attributed to its ability to synthesize such a wide range of predictive capabilities across a variety of traffic conditions.

3.1.4 Discriminator

The Discriminator assesses the generated paths by comparing them against actual vehicle movements, assigning a score that reflects their authenticity [47]. This evaluation not only categorizes the paths as genuine or fabricated but also enhances the fidelity of the trajectory generation by providing feedback to the Generator.

The Discriminator's objective is formalized within the GAN's loss function, where it seeks to maximize its classification accuracy, thus enabling a refined generation of trajectories. Simultaneously, the model employs an L2 loss to quantify the deviation of the generated paths from the true trajectories, thereby selecting the most plausible path among multiple predictions.

Mathematically, the Discriminator's process is represented as follows:

Objective Function of GAN:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] \\ + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \qquad (9)$$

L2 Loss Metric:

$$L = \min_k \|Y_i - \hat{Y}_i^{(k)}\|_2, \qquad (10)$$

where, G denotes the Generator, D the Discriminator, x the real trajectory data, z the input noise vector, Y_i the actual trajectory, and $\hat{Y}_i^{(k)}$ the k-th predicted trajectory by the Generator.

3.2. Vehicular Attention Pooling Net (VAP-Net)

Generative Adversarial Networks (GANs) [14] employ a pooling layer that is vital for managing information sharing among objects in a scene. This layer calculates agents' relative positions and merges their hidden states to represent their interactions. For vehicle trajectory prediction at intersections, traffic light status is a key factor influencing driver behavior [22], and the proximity between vehicles and to the stop-bar largely dictates driving decisions [10]. Additionally, at intersections, drivers must consider the speed [18] and direction [12] of nearby vehicles. Faster vehicles are monitored more closely due to their higher potential for danger, and vehicles from opposing directions are also critical to note from the collision risk perspective.

Research has shown that vehicle interactions at intersections involve complexities beyond mere positional relationships, as the dynamics of surrounding vehicles, including their speed and direction, significantly influence decision-making during interactions [9, 19, 43]. Considering the particular nature of vehicle trajectory prediction at intersections, we present a novel pooling network architecture known as *Vehicular Attention Pooling Net (VAP-Net)*. By incorporating a mechanism of attention in addition to positional information into VAP-Net, the model enhances its ability to capture dynamic interactions between complex agents. By integrating various streams of information, including positional information, velocity vectors, and historical information, it is designed to intricately process and integrate information. The architecture effectively combines spatial and velocity embeddings with an attention-driven pooling mechanism to create nuanced representations of agent interactions.

3.2.1 Embedding and Interaction Encoding

VAP-Net begins its operation by embedding spatial and velocity information for each agent in the scene:

Relative Position Embedding For each pair of agents i, j, the relative position vector $\mathbf{RP}_{ij} = \mathbf{p}_i - \mathbf{p}_j$ is computed, where $\mathbf{p}_i$ and $\mathbf{p}_j$ denote their respective positional coordinates. This vector is then transformed through a spatial embedding layer:

$$\mathbf{E}_{ij} = f_{\text{spatial}}(\mathbf{RP}_{ij}), \qquad (11)$$

where f_{spatial} represents the spatial embedding function.

Velocity Vector Embedding Simultaneously, the velocity vector $\mathbf{v}_i$ of each agent is embedded:

$$\mathbf{V}_i = f_{\text{velocity}}(\mathbf{v}_i), \qquad (12)$$

with f_{velocity} being the velocity embedding function.

3.2.2 Attention Mechanism

The attention mechanism dynamically assigns weights to features based on their relevance:

Attention Weight Calculation For each agent pair, we concatenate their relative position and velocity embeddings and input this to an attention MLP. The output attention scores are normalized to form weights:

$$\mathbf{W}_{ij} = \text{Softmax}\left(f_{\text{attention}}([\mathbf{E}_{ij}; \mathbf{V}_i])\right), \qquad (13)$$

where $f_{\text{attention}}$ denotes the attention MLP.

3.2.3 Feature Integration and Contextual Pooling

Feature Aggregation The agent's hidden states $\mathbf{H}_i$ are modulated by the attention weights to emphasize significant interactions:

$$\mathbf{H}'_{ij} = \mathbf{H}_i \odot \mathbf{W}_{ij}, \qquad (14)$$

where $\odot$ signifies element-wise multiplication.

Contextual Pooling The weighted hidden states are then concatenated with the relative position embeddings. This concatenated vector is processed through a series of MLPs, and a max-pooling operation is applied to extract a refined representation. We define this output as the Attention Pooling Feature ($\mathbf{APF}_i$), which encapsulates the fused characteristics of individual movement and interactive dynamics:

$$\mathbf{APF}_i = \max\left(\mathbf{MLP}\left([\mathbf{E}_{ij}; \mathbf{H}'_{ij}]\right)\right), \qquad (15)$$

capturing both individual motion characteristics and interaction dynamics.

Through the integration of Pool Hidden Net's feature processing with an advanced attention mechanism, VAP-Net provides a context-aware synthesis of multi-agent interactions. This representation is pivotal for accurate prediction of future trajectories in complex and dynamic environments.

4. Experiments

4.1. Dataset

In the development of our proposed KI-GAN algorithm, a substantial set of vehicular movement parameters is required to accurately simulate and predict the interactions at intersections. These parameters include not only the basic metrics such as the positions, speeds, and accelerations of the vehicles but also extend to more detailed aspects like the vehicle type and size, as well as the status of traffic lights in each direction within the intersections. After a thorough comparison and analysis of a range of available datasets[5, 6, 8, 21, 40, 45], we found that the SinD dataset

[42] stands out as the only dataset that comprehensively meets all our requirements for input data. Its extensive coverage of the aforementioned parameters makes it exceptionally suitable for the verification algorithm in this article, offering the precise blend of data necessary to support the intricate computations and simulations at the heart of our research.

The SinD dataset, originating from a signalized intersection in Tianjin, China, stands as a rich and detailed resource for trajectory analysis, encompassing over 13,000 traffic participants across seven categories, including cars, buses, trucks, motorcycles, bicycles, tricycles, and pedestrians, over a span of seven hours. Besides providing comprehensive information regarding the type, dimensions, and trajectory of each vehicle, this dataset also provides comprehensive data regarding traffic light states and timings, which are crucial for studies related to vehicle dynamics and compliance with traffic signals. Its emphasis on signalized intersections and diverse traffic participant categories, offering a wealth of detailed information, meets the specific needs of our model for comprehensive urban traffic analysis.

4.2. Experimental Setup and Evaluation Metrics

The original SinD dataset captures data at 30 frames per second. For our study, we sampled the data at a step size of 15, resulting in a frequency of 2 frames per second. In situations of minimal vehicle movement or complete stops at intersections influenced by traffic signals, brief observation intervals can result in skewed data, which can negatively affect the model's ability to learn dynamic movements. Thus, we observed data for 6 seconds (12 frames) and predicted trajectories for both 6 seconds (12 frames) and 9 seconds (18 frames).

Regarding hyperparameters, we selected a batch size of 64 and trained the model for 50 epochs. The learning rate for the Generator was set to 0.001, and for the Discriminator, it was set to be 0.0005.

We used Average Displacement Error (ADE) and Final Displacement Error (FDE) as metrics for trajectory accuracy evaluation. These are defined as:

$$\text{ADE} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \sqrt{(x_t^i - \hat{x}_t^i)^2 + (y_t^i - \hat{y}_t^i)^2}, \quad (16)$$

$$\text{FDE} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(x_T^i - \hat{x}_T^i)^2 + (y_T^i - \hat{y}_T^i)^2}, \quad (17)$$

where N is the total number of trajectories, T the prediction horizon, (x_t^i, y_t^i) the ground truth coordinates, and $(\hat{x}_t^i, \hat{y}_t^i)$ the predicted coordinates at time t for the i-th trajectory.

Moreover, experiments for this study were performed using an Nvidia RTX 3090 graphics card, with the code developed in Pytorch.

4.3. Results

Quantitative Results

The SinD dataset encompasses 23 segments, out of which 20 were allocated for training and 3 for validation. Table 2 presents the evaluation metrics on the validation set, showcasing the results for both 12-frame observations with 12-frame predictions and 12-frame observations with 18-frame predictions.

Table 2. Evaluation Metrics for Trajectory Prediction at Signalized Intersection (*In adapting the algorithm from [22] for SinD dataset, which lacks public video information, we have omitted the vision encoding module. Instead, we focus on employing trajectory data, vehicle states, and traffic light status for our comparison models, aligning with the available dataset features.)

Method	12-12		12-18	
	ADE	FDE	ADE	FDE
S-GAN[15]	1.32	2.46	1.53	2.95
S-LSTM[3]	0.87	1.60	0.96	1.78
Trajetron++[39]	0.37	0.93	0.70	1.91
FJMP [37]	0.27	0.68	0.41	1.13
[22]*	0.10	0.21	0.16	0.38
KI-GAN	*0.05*	*0.12*	*0.11*	*0.26*

Overall, the ADE and FDE metrics across all algorithms were relatively low. In addition to the fact that more than half of the vehicles at a traffic-light-controlled intersection were stationary or low-speed, lane-changing behavior was almost completely absent, resulting in universally lower errors at this intersection. For 12-frame predictions, our proposed method outperforms the others by 43%, while for 18-frame predictions, our proposed method outperforms the others by over 31%. As opposed to other methods that rely solely on trajectory information and are less responsive to discontinuous temporal changes [3, 15, 37, 39], our model is much more sensitive to traffic light changes. Furthermore, our approach can identify the type and size of vehicles, employing distinct prediction methods for motor vehicles such as cars and non-motorized vehicles like bicycles and motorcycles. This also contributes positively to the improvement of trajectory prediction accuracy.

Qualitative Results

Figure 1 illustrates selected trajectory prediction segments from the validation set. The top three images correspond to 12-frame observations with 12-frame predictions, while the bottom three images pertain to 12-frame observa-

Table 3. Impact of Encoder Components on Trajectory Prediction Accuracy

Interaction Encoder			Traffic Encoder	12-12		12-18	
Trajectory Encoder	Motion Encoder	Physical Attribute Encoder		ADE	FDE	ADE	FDE
✓	✓	✓	✗	0.52 (↑ 940.0 %)	1.13 (↑ 841.7 %)	0.79 (↑ 618.2 %)	1.54 (↑ 492.3 %)
✓	✓	✗	✓	0.14 (↑ 180.0 %)	0.32 (↑ 166.7 %)	0.22 (↑ 100.0 %)	0.48 (↑ 84.6 %)
✓	✗	✓	✓	0.12 (↑ 140.0 %)	0.27 (↑ 125.0 %)	0.17 (↑ 54.5%)	0.37 (↑ 42.3 %)
✓	✓	✓	✓	0.05	0.12	0.11	0.26

Table 4. Comparison of Pooling Methodologies on Trajectory Prediction Performance

Pooling Method	12-12		12-18	
	ADE	FDE	ADE	FDE
Social Pool Net[3]	0.07 (↑ 40.0 %)	0.15 (↑ 25.0 %)	0.14 (↑ 27.3 %)	0.33 (↑ 26.9 %)
Hidden Pool Net[15]	0.08 (↑ 60.0 %)	0.16 (↑ 33.3 %)	0.19 (↑ 72.7 %)	0.42 (↑ 61.5 %)
VAP Net	0.05	0.12	0.11	0.26

tions with 18-frame predictions. Visually, the congruence between predicted trajectories and the ground truth is remarkably high for both sparse and dense traffic scenarios, affirming the low overall ADE and FDE metrics.

4.4. Ablation Experiments

In the ablation study, to determine the effectiveness of each component of our model, we systematically removed each encoder except the Trajectory Encoder and observed the model performance.

The data in Table 3 illustrates the significant role each encoder plays in the model's performance. For instance, the exclusion of the Traffic Encoder leads to an increase in ADE by 940.0% and FDE by 841.7% for the 12-frame observation and prediction scenario, which underscores the encoder's critical role in capturing traffic light-related dynamics. Similarly, the omission of the Motion Encoder results in a less pronounced but still substantial increase in error rates (ADE by 180.0% and FDE by 166.7% for the 12-frame observation and prediction scenario), indicating its importance in understanding vehicular motion. The combined use of all encoders minimizes prediction errors, achieving the lowest ADE and FDE, which confirms the synergistic effect of the integrated information from all encoders on the model's accuracy.

Furthermore, we assessed how various pooling modules affected the overall model performance. According to Figure 4, the results indicate that the pooling approach is extremely important for trajectory prediction accuracy. The study indicated that attention to vehicle dynamics and interactions significantly enhances the accuracy of prediction with our proposed Vehicular Attention Pooling Net (VAP-Net). Based on Table 4, both the ADE and FDE metrics are lower compared to the Social Pool Net [3] and Hidden Pool Net [15].

5. Conclusion

Our work on the Knowledge-Informed Generative Adversarial Network (KI-GAN) represents a significant stride in trajectory forecasting at signalized intersections. The incorporation of a multi-encoder framework, including the innovative Vehicle Attention Pooling Net (VAP-Net), has proven effective in accurately modeling the complexities of intersection dynamics. A key strength of KI-GAN is the integration of critical traffic signal information and vehicle interactions, which is evident in its performance metrics. Among the key components, the VAP-Net stands out, demonstrating the effectiveness of its attention mechanism in improving trajectory prediction accuracy. As a result of this mechanism's emphasis on vehicle speed and proximity between vehicles, which are crucial in intersection scenarios, the model is able to predict vehicle paths under varying traffic conditions with greater accuracy.

With the promising results of KI-GAN, further advancements in traffic management systems and autonomous intersection navigation can be achieved, emphasizing that the model can be applied to real-world scenarios. For a more in-depth understanding of interactions with other road users (e.g., vulnerable road users) at intersections, future developments could include integrating more contextual data or refining VAP-Net's attention process. In navigating the complexity of urban traffic environments, KI-GAN offers a comprehensive and robust solution, setting a new standard.

References

[1] Amr Abdelraouf, Mohamed Abdel-Aty, Zijin Wang, and Ou Zheng. Trajectory prediction for vehicle conflict identification at intersections using sequence-to-sequence recurrent neural networks. *arXiv preprint arXiv:2210.08009*, 2022. 3

[2] Amr Abdelraouf, Rohit Gupta, and Kyungtae Han. Interaction-aware personalized vehicle trajectory prediction using temporal graph neural networks. *arXiv preprint arXiv:2308.07439*, 2023. 2

[3] 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. 3, 7, 8

[4] Florent Altché and Arnaud de La Fortelle. An lstm network for highway trajectory prediction. In *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)*, pages 353–359. IEEE, 2017. 2

[5] 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

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

[7] Qianxia Cao, Zhongxing Zhao, Qiaoqiong Zeng, Zhengwu Wang, and K. Long. Real-time vehicle trajectory prediction for traffic conflict detection at unsignalized intersections. *Journal of Advanced Transportation*, 2021. 3

[8] 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

[9] Ni Dang, Tim Brüdigam, Zengjie Zhang, Fangzhou Liu, Marion Leibold, and Martin Buss. Distributed stochastic model predictive control for a microscopic interactive traffic model. *Electronics*, 2023. 2, 6

[10] Debargha Dey, Marieke Martens, Berry Eggen, and Jacques Terken. The impact of vehicle appearance and vehicle behavior on pedestrian interaction with autonomous vehicles. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications adjunct*, pages 158–162, 2017. 3, 6

[11] Myungeun Eom and Byung-In Kim. The traffic signal control problem for intersections: a review. *EUROPEAN Transport Research Review*, 12(1), 2020. 2

[12] Fang Fang, Pengpeng Zhang, Bo Zhou, Kun Qian, and Yahui Gan. Atten-gan: Pedestrian trajectory prediction with gan based on attention mechanism. *Cognitive Computation*, 14 (6):2296–2305, 2022. 6

[13] S. A. Fayazi and A. Vahidi. Mixed-integer linear programming for optimal scheduling of autonomous vehicle intersection crossing. *IEEE Transactions on Intelligent Vehicles*, 3: 287–299, 2018. 2

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 6

[15] 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. 3, 7, 8

[16] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–1780, 1997. 5

[17] Zubayer Islam, Mohamed Abdel-Aty, Amrita Goswamy, Amr Abdelraouf, and Ou Zheng. Effect of signal timing on vehicles' near misses at intersections. *Scientific Reports*, 13 (1):9065, 2023. 3

[18] 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, 3, 6

[19] Byeoungdo Kim, C. Kang, Jaekyum Kim, Seung Hi Lee, C. Chung, and J. Choi. Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 399–404, 2017. 6

[20] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, Hamid Rezatofighi, and Silvio Savarese. Socialbigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. *Advances in neural information processing systems*, 32, 2019. 3

[21] 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. 6

[22] Seoyoung Lee, Hyogyeong Park, Yeonhwi You, Sungjung Yong, and Il-Young Moon. Deep learning-based multimodal trajectory prediction with traffic light. *Applied Sciences*, 13 (22), 2023. 2, 6, 7

[23] Siyan Li, Chuheng Wei, Guoyuan Wu, Matthew J Barth, Amr Abdelraouf, Rohit Gupta, and Kyungtae Han. Personalized trajectory prediction for driving behavior modeling in ramp-merging scenarios. In *2023 Seventh IEEE International Conference on Robotic Computing (IRC)*, pages 1–4. IEEE, 2023. 2

[24] Lei Lin, Weizi Li, Huikun Bi, and Lingqiao Qin. Vehicle trajectory prediction using lstms with spatial–temporal attention mechanisms. *IEEE Intelligent Transportation Systems Magazine*, 14:197–208, 2022. 3

[25] Yicheng Liu, Jinghuai Zhang, Liangji Fang, Qinhong Jiang, and Bolei Zhou. Multimodal motion prediction with stacked transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7577–7586, 2021. 3

[26] Yongkang Liu, Xuewei Qi, Emrah Akin Sisbot, and Kentaro Oguchi. Multi-agent trajectory prediction with graph attention isomorphism neural network. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 273–279. IEEE, 2022.

[27] Benedikt Mersch, Thomas Höllen, Kun Zhao, C. Stachniss, and R. Roscher. Maneuver-based trajectory prediction for self-driving cars using spatio-temporal convolutional networks. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4888–4895, 2021. 3

[28] Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, and Fawzi Nashashibi. Attention based vehicle trajectory prediction. *IEEE Transactions on Intelligent Vehicles*, 6(1):175–185, 2020. 3

[29] Xiaoyu Mo, Yang Xing, and Chen Lv. Graph and recurrent neural network-based vehicle trajectory prediction for highway driving. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 1934–1939, 2021. 3

[30] 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. 2

[31] Ingrid Navarro and Jean Oh. Social-patternn: Socially-aware trajectory prediction guided by motion patterns. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9859–9864. IEEE, 2022. 3

[32] Geunseob Oh and H. Peng. Impact of traffic lights on trajectory forecasting of human-driven vehicles near signalized intersections. *arXiv: Robotics*, 2019. 3

[33] P. Pecher, M. Hunter, and R. Fujimoto. Data-driven vehicle trajectory prediction. *Proceedings of the 2016 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, 2016. 3

[34] 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. 2

[35] Dongbo Peng, Guoyuan Wu, and Kanok Boriboonsomsin. Energy-efficient dispatching of battery electric truck fleets with backhauls and time windows. *SAE International Journal of Electrified Vehicles*, 13(14-13-01-0009), 2023. 2

[36] A. Quintanar, D. F. Llorca, I. Parra, R. Izquierdo, and M. Sotelo. Predicting vehicles trajectories in urban scenarios with transformer networks and augmented information. *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 1051–1056, 2021. 2

[37] 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. 3, 7

[38] Debaditya Roy, Tetsuhiro Ishizaka, C. Mohan, and A. Fukuda. Vehicle trajectory prediction at intersections using interaction based generative adversarial networks. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 2318–2323, 2019. 2, 3

[39] 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. 3, 7

[40] 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. 6

[41] 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. 2

[42] 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. 4, 7

[43] 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. 6

[44] W. Zhan, Liting Sun, Yeping Hu, Jiachen Li, and M. Tomizuka. Towards a fatality-aware benchmark of probabilistic reaction prediction in highly interactive driving scenarios. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3274–3280, 2018. 2

[45] 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

[46] 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. 2, 3, 4

[47] Liang Zhao, Yufei Liu, A. Al-Dubai, Albert Y. Zomaya, G. Min, and Ammar Hawbani. A novel generation-adversarial-network-based vehicle trajectory prediction method for intelligent vehicular networks. *IEEE Internet of Things Journal*, 8:2066–2077, 2021. 5

[48] Zhanle Zhao and Chuheng Wei. An analysis of the brake performance of radar-based adaptive cruise control during ramp merging on simulation software. In *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pages 1112–1115. IEEE, 2022. 2

[49] Alex Zyner, Stewart Worrall, and E. Nebot. Naturalistic driver intention and path prediction using recurrent neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 21:1584–1594, 2018. 2