

Article

Trajectory Prediction with Attention-Based Spatial–Temporal Graph Convolutional Networks for Autonomous Driving

Hongbo Li ^{1,2,*} , Yilong Ren ^{1,2,*} , Kaixuan Li ¹ and Wenjie Chao ¹

¹ School of Transportation Science and Engineering, Beihang University, Beijing 100191, China; lihongbo@buaa.edu.cn (H.L.); kxuanli@163.com (K.L.)

² Zhongguancun Laboratory, Beijing 100094, China

* Correspondence: yilongren@buaa.edu.cn

Abstract: Accurate and reliable trajectory prediction is crucial for autonomous vehicles to achieve safe and efficient operation. Vehicles perceive the historical trajectories of moving objects and make predictions of behavioral intentions for a future period of time. With the predicted trajectories of moving objects such as obstacle vehicles, pedestrians, and non-motorized vehicles as inputs, self-driving vehicles can make more rational driving decisions and plan more reasonable and safe vehicle motion behaviors. However, due to traffic environments such as intersection scenes with highly interdependent and dynamic attributes, the task of motion anticipation becomes challenging. Existing works focus on the mutual relationships among vehicles while ignoring other potential essential interactions such as vehicle–traffic rules. These studies have not yet deeply explored the intensive learning of interactions between multi-agents, which may result in evaluation deviations. Aiming to meet these issues, we have designed a novel framework, namely trajectory prediction with attention-based spatial–temporal graph convolutional networks (TPASTGCN). In our proposal, the multi-agent interaction mechanisms, including vehicle–vehicle and vehicle–traffic rules, are meticulously highlighted and integrated into one homogeneous graph by transferring the time-series data of traffic lights into the spatial–temporal domains. Through integrating the attention mechanism into the adjacency matrix, we effectively learn the different strengths of interactive association and improve the model’s ability to capture critical features. Simultaneously, we construct a hierarchical structure employing the spatial GCN and temporal GCN to extract the spatial dependencies of traffic networks. Profiting from the gated recurrent unit (GRU), the scene context in temporal dimensions is further attained and enhanced with the encoder. In such a way, the GCN and GRU networks are fused as a features extractor module in the proposed framework. Finally, the future potential trajectories generation tasks are performed by another GRU network. Experiments on real-world datasets demonstrate the superior performance of the scheme compared with several baselines.

Keywords: autonomous vehicle; trajectory prediction; intersection scene; graph convolutional networks; NGSIM



Citation: Li, H.; Ren, Y.; Li, K.; Chao, W. Trajectory Prediction with Attention-Based Spatial–Temporal Graph Convolutional Networks for Autonomous Driving. *Appl. Sci.* **2023**, *13*, 12580. <https://doi.org/10.3390/app132312580>

Academic Editor: Juan-Carlos Cano

Received: 6 October 2023

Revised: 27 October 2023

Accepted: 29 October 2023

Published: 22 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Autonomous vehicles have become increasingly popular recently and are expected to remarkably improve transportation efficiency in innovative ways, which is enabled by a series of emerging technologies [1,2]. The fact is that the premise of deploying autonomous vehicles is the need to ensure the absolute safety of all traffic participants. One of the alternatives is to have sufficient awareness of the other road sharers’ movements for future periods. Evidently, if the future motion status of nearby agents can be accurately predicted, the autonomous vehicles would conduct progressive decision making to minimize risk and avoid potential collisions. However, due to the traffic environment’s highly interdependent and dynamic characteristics, especially in urban driving scenes, accurate trajectory prediction is a challenging task for autonomous vehicles.

To alleviate the matter raised above, multitudes of researchers in the field of autonomous vehicles have demonstrated diversified schemes involving physics-based, maneuver-based, and interaction-aware designs [3,4]. Among them, the interaction-aware proposals have become widely appealing for their outstanding performance in handling the prediction issues [5,6]. In the interaction-aware methods, the tracks of nearby vehicles, together with target agent states, are concurrently considered, which significantly enhances the ability of prediction. The performance improvements of these methods are due to feeding the vehicle–vehicle interaction information [7,8]. In practice, the traffic rules are another critical concern for vehicles when traveling across intersection and play significant roles in reshaping trajectories. However, there are rare existing works that take the traffic-rule factors into account and furthermore consider the interaction of vehicle–traffic rules in trajectory prediction problems, which may lead to evaluation bias. In addition, the traffic rules and vehicle trajectory data are heterogeneous, where the first one is time-series data and the latter belongs to spatial–temporal sequences, thus inducing another key problem of how to integrate them through a united network.

In reality, the trajectory anticipation task hinges on how to effectually extract representations from historical movements in complex scenarios. As in conventional traffic prediction, the spatial–temporal dependencies are emphasized to execute anticipation of the trajectory [9–11]. Recently, machine learning techniques have offered a promising strategy for integrating multi-modal features in virtue of their competence in data mining, wherein time-series-based approaches, such as the CNN and RNN, have been broadly adopted in the setting of trajectory prediction [12–14]. Nevertheless, these methods are not adequately suitable for non-Euclidean domains, especially for intersection scenes, where interactive operations frequently occur. Inspired by the topology structure, the graph convolutional network (GCN) is introduced to depict vehicle relationships, thus solving the problem of trajectory prediction [3,15,16]. These initial attempts regard the connection intensity as the function of the distance between nodes in the graph, which leads to a lack of growing learning ability. In fact, the interactive concentration is a dynamic variable and depends on the degree of impact, which cannot be expressed briefly by space distance. As shown in Figure 1, when $T = 1$ s, vehicle3 exerts more influence on vehicle1 than vehicle4 since vehicle3 and vehicle1 are in the same lane, despite the longer distance between vehicle1 and vehicle3 than that between vehicle1 and vehicle4. Nevertheless, as $T = 2$ s, vehicle1 approaches the intersection, while vehicle4 leaves the research region, in which case the traffic rules dominate. From the above appearance-based observations, it can be concluded that the interactive relationships cause unremitting changes over the time windows and need to be properly distinguished.

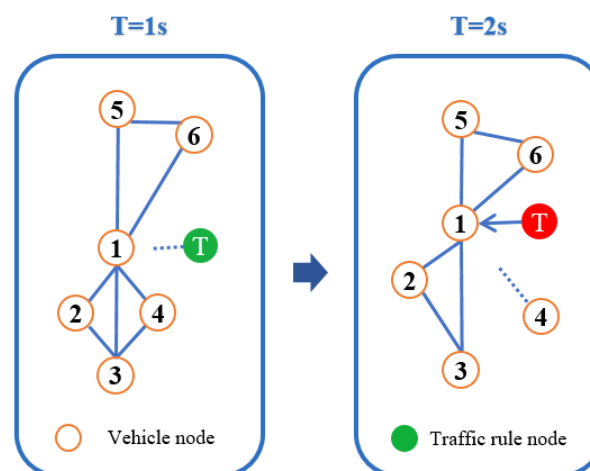


Figure 1. Illustration of the dynamic interaction.

In response to the problems mentioned above, a novel framework is developed in this work, namely trajectory prediction with attention-based spatial-temporal graph convolutional networks (TPASTGCN). In our scheme, the multi-agent interaction mechanisms of vehicle-vehicle and vehicle-traffic rules are carefully highlighted and effectively learned by a united graph, and they are proven to contribute significantly to the enhancement of the model performance. To capture the spatial dependencies of traffic networks with higher accuracy, the spatial GCN and the temporal GCN are utilized to construct a hierarchical structure. Additionally we embed the attention mechanisms, which help to focus on the vital information, into the GCN to learn the heterogeneous interaction relationships raised above. Moreover, we utilize a gated recurrent unit (GRU) encoder to extract and refine the scene context over time. This enables us to seamlessly incorporate both GCN and GRU networks into a unified feature extraction module. Ultimately, we leverage another GRU network to predict future potential trajectories with high fidelity yet low computational overhead. The key contributions of our work are highlighted below.

- The interdependent correlations of the interactions between vehicles and traffic rules in trajectory prediction problems are elaborated for the first time in our work. To further address the issue of data heterogeneity from diverse or varied sources, we also provide an innovative solution to transfer the time-series data of traffic rules into the spatial-temporal ones and produce a homogeneous joint graph along with trajectory data, which provides a fresh idea for the self-driving vehicle to predict the trajectory of nearby vehicles more accurately.
- The interactive intensities between agents are effectively learned and distinguished. We introduce an attention mechanism to learn about discrepancies in multi-agent interactions, which impels the graph dynamic in the spatial domain with growing learning ability. In such a way, we boost the model performance in capturing the spatial-temporal dependencies combined with the higher efficiency of GCN and GRU, which is conducive to the improvement of the prediction precision.
- A novel framework is developed for trajectory anticipation. It integrates the multi-agent interaction mechanisms and the scene context. Based on the real-world dataset tests, the effectiveness and superiority of our proposal have been verified through quantitative and qualitative analyses.

The rest of the paper is organized as follows. Section 2 provides the related works that investigate the vehicle trajectory prediction. Section 3 states the framework of the proposed proposal. The method is described in Section 4, and in Section 5, we present experimental results based on a real-world dataset and discuss their implications. Lastly, we summarize the contributions and future directions of our research in Section 6.

2. Related Work

With the continued information perception, autonomous vehicles are required to speculate road sharers' movements and then conduct safe path planning and decision making in real time to realize a collision-free path. However, in highly complex traffic architectures, especially for the intersection scenes, implementing trajectory anticipation is not a trivial task since there are several challenges. Firstly, in contrast to intention prediction, the trajectory anticipation problem requires generating more exhaustive information over a duration of time, which increases the demand for model performance. Secondly, there are assorted traffic participants and the degree of mutual influence varies, resulting in the difficulties of how to distinguish the effective impact or not and then how to properly evaluate their interactions. Additionally, traffic rules can reshape vehicles' driving behavior and should be taken into account, which also adds to the complexity of prediction. A number of works have been devoted to partly addressing these tricky issues, and such approaches can be broadly partitioned according to their input representations.

The first thread is mainly regarded as the conventional method with the target vehicle information as the input, including vehicle positions, speed, operation elements, driving styles and so on [17,18]. To extract the contextual attributes of road networks, a Bayesian

method combining uncertain observations was proposed for achieving the vehicle maneuvering intention evaluation [19]. Thomas and Karl introduced the Hidden Markov Models into a trajectory prediction framework and showed that the learning accuracy depends on both the size of the data and the data quality, which transformed the prediction task to the classifying problem [20]. Based on the autoregressive moving average model, Qian et al. developed a two-level quantified adaptive Kalman filter algorithm to predict vehicle location and velocity and then broadcasted the vehicle states to each other via the road unit [21]. These schemes only leverage the target vehicle features and ignore the movements of other nearby agents, which can lead to errors in the process of trajectory evaluation in crowded driving situations.

Inspired by the social interaction, recently the tracks of surrounding vehicles have been incorporated into studies, and a number of interaction-aware models have been developed based on the fact that the vehicles in the traffic environments are interplaying [22]. Here, the machine learning methods are widely adopted and demonstrate preferable performance in dealing with the trajectory prediction problem [23,24]. Phillips et al. employed an LSTM model to classify the observed vehicles maneuvers considering the neighboring features, which outperforms the baselines of conditional probability table and multilayer perceptron models [25]. Similarly, Nachiket et al. proposed an LSTM model for vehicle motion prediction on freeways by utilizing an NGSIM dataset [26]. To improve the estimation accuracy, Zhao et al. introduced a Multi-Agent-Tensor Fusion network and evaluated its performance in both highway and pedestrian crowd scenes, incorporating scene context constraints and vehicle interactions [27]. Nevertheless, the above proposals neglect the spatial features in conducting the trajectory evaluation. To tackle the drawback, Mercat et al. designed an improved LSTM vehicle motion forecasting scheme embedded with multi-head attention aiming to realize the interactions and uncertainty estimation [28]. Using the clustering method and LSTM model, Zhang et al. integrated the intention prediction and trajectory prediction tasks, where the statistical law of trajectory was used to provide prior knowledge for latter anticipation [29]. However, as the number of stacked layers increases, the performance of the RNN-based prediction models is gently ameliorated while the computation sources exponentially grow, which is obviously not suitable for the real-time requirements of autonomous vehicles [30]. In addition, the RNN works in the Euclidean domain, leading to the trickiness to extract and describe several critical features of traffic networks. In contrast, the GCN is an expert in dealing with the topology structure similarly to the multi-agent interactions in the intersection scene, and its excellent function has been proven [31–33]. In the field of traffic flow prediction, attention-based spatio-temporal graph convolutional network models are beginning to be used to tackle the deficiency of spatio-temporal correlation of traffic data in traffic flow prediction methods [34–36]. Therefore, our proposal benefits the GCN network, and proposes a novel framework for enhancing trajectory anticipation tasks.

In summary, while existing works have attempted to perform the trajectory predictions via diverse solutions, there are several remaining challenges which require attention. Traffic rules such as traffic lights and stop lines have rarely been incorporated into the proposal. Despite the focus on interactions among vehicles, the multi-agent influence degree or the intensity also needs to be learned, which is conducive to enhancing the capacity of capturing crucial features. In this work, we provide a new solution for the intersection scene trajectory predictions for the purpose of tackling the above challenges.

3. Framework Statement

In this section, we first describe the problem statement and then elaborate on our proposed framework construction process for outlining the overall architecture.

3.1. Problem Definition

The task of trajectory anticipation is to model a vehicle movement sequence that simultaneously embeds temporal and spatial characteristics. The general proposal is

to predict the potential position of ego-vehicle for the T upcoming time steps based on historical information. Specifically, our approach not only considers history attributes but also substantially integrates multi-agent interaction mechanisms. The interaction features are derived by the interaction graph G , where the relationships of vehicle–vehicle and vehicle–traffic rules are aligned at the same time. Thus, the prediction problem input I can be extended to the following:

$$I = [X_t, G_t] \quad (1)$$

where

$$X_t = [\mu_t^1, \mu_t^2, \dots, \mu_t^n] \quad (2)$$

$$\mu_t^i = [(x_{(t-\tau+1)}^i, y_{(t-\tau+1)}^i), \dots, (x_{(t-1)}^i, y_{(t-1)}^i), (x_t^i, y_t^i)] \quad (3)$$

where X_t denotes the historical status μ of n vehicles before moment t . The τ is the trace-back horizon, and the μ_t^i normally shows the coordinates of vehicle i .

The output O presenting future potential trajectory in prediction length T after observed time T_{obs} can be defined as shown below:

$$O = [Y_{(T_{obs}+1)}^i, Y_{(T_{obs}+2)}^i, \dots, Y_{(T_{obs}+T)}^i] \quad (4)$$

where $Y_{(t)}^i$ is the vehicle i anticipation position at moment t .

As a result, the crucial challenge of the problem is how to effectively learn the rule function $f(\cdot)$ according to the prior knowledge extracted from X and G in time windows $[T_{obs} - \tau, T_{obs}]$. Once the $f(\cdot)$ is determined, the potential position of the vehicle in the short term can be obtained by the following:

$$\begin{aligned} & [Y_{(T_{obs}+1)}^i, Y_{(T_{obs}+2)}^i, \dots, Y_{(T_{obs}+T)}^i] \\ &= [f(G, X_{(T_{obs}-\tau)}^i, \dots, X_{(T_{obs}-1)}^i), X_{(T_{obs})}^i)] \end{aligned} \quad (5)$$

3.2. Proposal Construction

To address the issue of frequent mutual and highly dynamic interactions among traffic participants in intersection scenes, we proposed a novel deep learning framework that combines a graph-based convolutional network with a gated recurrent unit network for trajectory anticipation. As Figure 2 shows, the overall architecture of the proposed proposal contains the generation of multi-agent graph module (GMG), the attention-based spatial-temporal fusion graph module (ASTFG) and the trajectory prediction module (VTP). The primary scene is first described by the GMG, in which vehicle nodes and traffic rules are implanted. With the historical sequence information as the input, the ASTFG fully utilizes the attributes to identify an interaction diagram structure and capture the spatial and temporal characteristics. In such a way, the outputs of ASTFG are fed into a VTP that embeds the decoder of a gated recurrent unit (GRU), aiming to extract the scene context and multi-agent intersection features through information transmission. Finally, the future potential positions of vehicles can be obtained by the VTP. The benefits of our united scheme lie in competently handling both spatial-temporal dependencies and resolving multi-agent mechanisms cases, where the purpose of collision-free autonomous vehicles can be achieved by referring to reliable candidate output trajectories.

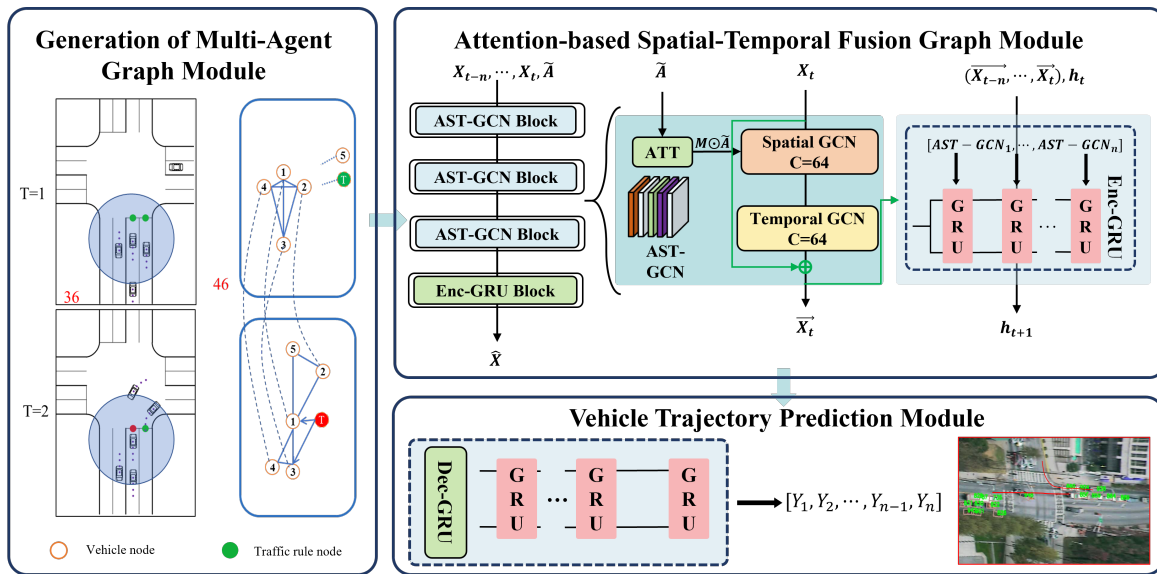


Figure 2. Framework of the proposed vehicle trajectory prediction model. The overall architecture contains three portions, namely the generation of multi-agent graph module (GMG), the attention-based spatial-temporal fusion graph module (ASTFG) and the trajectory prediction module (VTP). The attention mechanism (ATT) is introduced to the AST-GCN part, where spatial GCN and temporal GCN are elaborated.

4. Methodology

4.1. Generation of Multi-Agent Graph

The scene description in Figure 2 illustrates the dynamic structure of vehicles in the intersection at two consecutive time steps. Initially, vehicle1 senses *vehicle3*, and the phase is green with an accessible state. Next, at $T = 2$, vehicle5 enters the perception range of vehicle1 while the phase turns red, resulting in a change in the movement of vehicle1. The processes described above visually elaborate the effects of surrounding vehicles and traffic rules on the target object. Inspired by the agents' mutual topology, we depict these intersections among vehicle-vehicle and vehicle-traffic rules as a self-linkage graph $G = \{V, \mathcal{E}_S, \mathcal{E}_T\}$. Each vehicle along with traffic rules (traffic signal and stop line) are treated as the node and represented with $V = \{v_1, v_2, \dots, v_n\}$, where n is the number of nodes. The graph G edge sets E are consciously partitioned into spatial $\mathcal{E}_S$ and temporal $\mathcal{E}_T$, of which $\mathcal{E}_S = \{v_i^t v_j^t | i, j \in n\}$ and $\mathcal{E}_T = \{v_i^t v_i^{t+1} | i \in n\}$. The $\mathcal{E}_S$ indicates that at each time step, each node is connected to its surrounding nodes within a radius threshold D , which reflects the spatial intersection. The self-connection at the adjacent horizon is denoted by $\mathcal{E}_T$ to learn identity features in the temporal dimension. The graph G can be expressed mathematically as an adjacency matrix $\tilde{A} = A + I$ made up of the elements with 0 or 1, where $\tilde{A}$ can further be displayed in $\tilde{A}|G = \{V, \mathcal{E}_S, \mathcal{E}_T\} = A|\mathcal{E}_S + I|\mathcal{E}_T$. In particular, the element a of matrix A is defined as (6), representing that when the distance between nodes is less than D , the corresponding edge is connectively assigned with 1; otherwise, the value is 0.

$$a_{ij} = \begin{cases} 1, & \text{if } \|v_i^t v_j^t\| < D \\ 0, & \text{if } \|v_i^t v_j^t\| \geq D \end{cases} \quad (6)$$

It is worth noting that the traffic rules are innovatively introduced into the trajectory anticipation problem in our proposal. However, in the process of fusing trait data from diverse sources, one tricky issue arises: the fact that trajectories of vehicles are spatial-temporal data, whereas traffic-light data are time-series data. The different data characteristics would impose the generated graph transforming to heterogeneous data, arousing a high level of modeling complexity. To simplify the drawback without losing the purpose, we develop an

ingenious solving method, where the signal phase is moved to the intersection stop line and is visualized in 2D space. Only when the phase turns red does the stop line emerge. In such a way, the traffic rules become the same spatial-temporal data as vehicles trajectories, of which the stop line is applied to reflect the spatial position and the temporal attribute is described by signal phase changes. Concentrating the signal phase on the stop line of a lane depends on which direction and lane of traffic the signal light allows to pass, indicating a vehicle's right-of-way. The advantages of the introduced method lie in two respects: one is to prevent the heterogeneous scene, and the other is to integrate the different traffic rules uniformly.

4.2. Attention-Based Spatial–Temporal Fusion Graph Module (ASTFG)

After accomplishing the interaction graph construction, the well-designed networks are exploited to extract the spatial-temporal dependencies as shown in Figures 3 and 4. Each AST-GCN block learns the feature of the current layer and builds a higher-level feature map on the graph through layers stacking. The AST-GCN block is capable of capturing the connected topology between the target vehicle and its surrounding agents along with traffic rules; then, the GRU block performs, encoding the relationships as well as vehicle attributes to handle the dependencies eventually. The topology in Figure 3 changes with time but the number of nodes is constant; what changes is the influence and weight between each node. So, we use an attention mechanism on the adjacency matrix to understand the impact of different nodes on the prediction accuracy. Specifically, the connected nodes are convolved in space by a spatial convolution kernel and then between different frames of the same node in time.

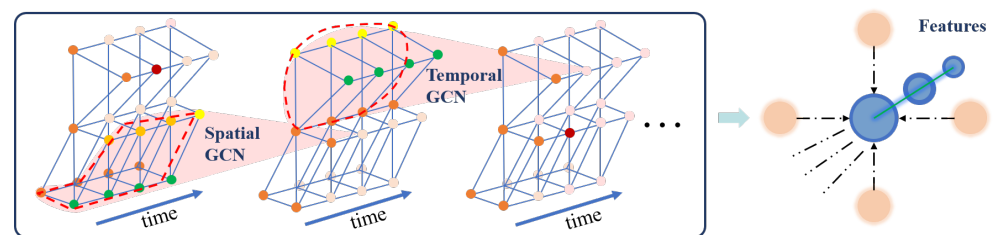


Figure 3. The process of features extraction by spatial-temporal graph convolution.

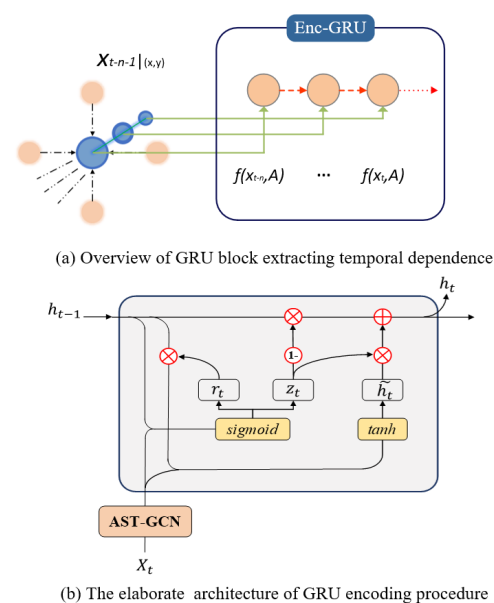


Figure 4. Schematic diagram of the GRU block in temporal dependence modeling.

4.2.1. Spatial Dependency Modeling

Figure 2 illustrates the AST-GCN block composition, where the spatial-temporal GCN and attention mechanism are concurrently designed. In our scheme, we append the temporal GCN to the end of each spatial GCN to alternately process the input data in time and space. As different nodes have different intensities of intersections, we design the attention mechanism to the adjacency matrix before inputting to the spatial GCN layer with the purpose of learning distinct nodes' effects on the prediction precision. The residual connections are leveraged to enable the model to transform features with a higher gradient and accelerate the learning rate.

To be specific, given the adjacency matrix and feature matrix, the GCN competently operates to recover attributes in the Fourier domain. The graph convolution can be stated as shown below:

$$H_{l+1} = \varpi(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^l W^l) \quad (7)$$

where H_l represents the output of the l -th layer. $\tilde{A} = A + I$, A is the adjacency matrix of G , and I is an identity matrix. $\tilde{D}$ denotes the diagonal node degree matrix, $\tilde{D} = \sum_j \tilde{A}_{ij}$, W^l is a matrix comprising the parameters of l -th and the impact of $\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}$ is to normalize the $\tilde{A}$. The $\varpi(\cdot)$ is the activation function, where the $ReLU(\cdot)$ that has had its outstanding performance in deep learning proved is accepted in our work. For the spatial GCN, its kernel size is $\Theta_G \in \mathbb{R}^{5 \times 1}$ and the layer convolves S nodes at a time. In contrast, the temporal operates in the time domain T_h , and the kernel is set as $\Theta_G \in \mathbb{R}^{T \times 1}$.

To enhance the model's robustness and speed up the model convergence rate, we exploit the batch normalization (BN) for the graph. For each batch, the input data are normalized by the following:

$$\tilde{x}^k = \frac{x^k - e[x^k]}{\sqrt{v[x^k]}} \quad (8)$$

where x^k and $\tilde{x}^k$ are the BN input and output, while $e[x^k]$ and $v[x^k]$ are the mean and variance of the input data, respectively.

As mentioned above, we introduce the attention mechanism to indicate other vehicles' heterogeneous influence on the target vehicle movements. With regard to each adjacency matrix, we design a learning representation matrix M and conduct the inner product of M and $\tilde{A}$ denoted by $(M \odot \tilde{A})_{ij}$. The M is firstly initialized as an identity matrix, which can be continuously updated to achieve the goal of distinguishing the interaction density.

Generally, the GCN is employed to mine the spatial characteristics of the trajectory prediction problem in depth. Note that both the vehicle-vehicle features and the vehicle-traffic rules relationships in space are all successfully captured in the AST-GCN block. The aim of stacking GCN layers is to extract higher-level spatial dependencies.

4.2.2. Temporal Dependency Modeling

Acquiring the temporal dependence is another crucial problem in trajectory anticipation. Although the AST-GCN block has amassed the overall spatial-temporal features, its primary focus is on other agents, and it does not emphasize inherent historical affiliated relations. In fact, prospective motions of a vehicle intensely hinge on the historical status since the driving behavior and decision-making ability remain consistent over time and thus steady for certain drivers. To this end, numerous existing research studies make an effort to model the temporal dependence issue by data-driven methods. Among them, the LSTM and other RNN-based methods are mostly developed to extract the features for trajectory prediction. However, these studies have the drawback of low training efficiency and tend to cause gradient disappearance, resulting in the prediction consequence being barely satisfactory compared to expectations.

Inspired by the outstanding performance of the GRU for sequence modeling because of its fewer parameters and cheap computation cost, we adapt the GRU to propose the Enc-GRU block to extract temporal dependency. The responsibility of Enc-GRU block is to analyze the upper layer information generated by the AST-GCN block and make a valid integration with context scenes to produce latent representations. Figure 4 depicts the process of temporal features modeling using the GRU approach.

The concrete procedure of encoding the temporal features based on GRU is formulated as follows. For each ongoing horizon, we take the spatial-temporal GCN output denoted as $f(x_t, A)$ and the previous hidden state h_{t-1} as the input for the Enc-GRU block. With iterative updating, the output serviced as current hidden state h_t at t moment can be obtained by the following:

$$\begin{cases} z_t = \sigma(W_z \cdot [f(x_t, A), h_{t-1}]) \\ r_t = \sigma(W_r \cdot [f(x_t, A), h_{t-1}]) \\ \tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, f(x_t, A)]) \\ h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \end{cases} \quad (9)$$

where W_z, W_r, W are the parameters, $\sigma(\cdot)$ is the sigmoid function, and z_t controls the degree of previous information entering the current state. The r_t is the reset gate, which is leveraged to determine the previous step information ignoring degree. $\tilde{h}_t$ represents the contents reserved at time t , and $\odot$ is the Hadamard production.

Through several rounds of recurrent updates with Equation (9), the temporal dependencies of trajectories and traffic rules are extracted from the primary state to the final state and ultimately integrated with the output Enc_{out} . In this way, the Enc_{out} simultaneously contains all features of the multi-agent interaction mechanisms and context scene, where spatial dependencies and temporal attributes are operated by the AST-GCN and Enc-GRU block separately in the fusion graph module of the proposed strategy.

4.3. Vehicle Trajectory Prediction Module (VTP)

To strive for consistency and reduce complexity without losing precision, we further employ the GRU network to perform trajectory anticipation conclusively. It is well known that the prediction task of trajectory is the sequence-generating issue, intending to approximate the real distribution as accurately as possible. Taking the Enc_{out} built by the attention-based spatial-temporal fusion graph module and previous hidden state as input, the VTP module predicts vehicle position in the next several time steps. The GRU forms a chain structure, where the former output is fed into latter object as input. Among them, for the first GRU, its input also embeds previous time coordinates of the target vehicle, as shown in Figure 5. Moreover, with the purpose of adjusting the model's prediction rate as the vehicle's moving speed changes, the residual connection is imposed between each input and output of the GRU decoder. In such a way, the outputs of the GRU decoder are combined to form a tensor $Dec_{out} \in \mathbb{R}^{B \times S \times N}$, where B is the batch size, S is the hidden size and N is the prediction horizon. Finally, by applying the full connection, the future trajectories Y_{t+n} are generated.

4.4. Training Loss and Optimizer Construction

In order to evaluate the model training performance, we design a loss function that directs the model to approach desirable outcomes. As expressed by Equation (10), we focus on minimizing the differences between prediction value $\tilde{Y}$ and ground truth Y during training.

$$loss = (1/t_f) \sum_{t=1}^{t_f} \|Y_t - \tilde{Y}_t\| \quad (10)$$

In addition, the optimizer is leveraged to adjust and update the the model parameters, which is crucial for the model convergence rate. In our work, we take the Adam method as the training optimizer considering its extrusive properties. The Adam method is momentum-based, and it is an alternative method to calculate the adaptive learning rate for each parameter in combination with the advantages of AdaGrad and RMSProp, which has been widely approved in deep learning [37].

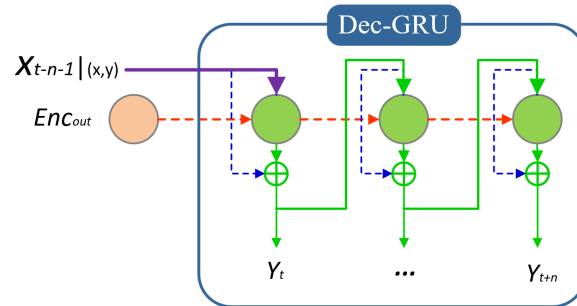


Figure 5. Future potential vehicle trajectory prediction.

5. Experiments and Results

This section firstly describes the datasets used for experiments and then introduces some state-of-the-art models to compare with the proposed scheme. The experiments are comprehensively conducted from an ablation study, quantitative as well as qualitative analysis, and interaction evaluation with the aim of providing novel insights for the vehicle trajectory prediction problem.

5.1. Dataset Description

The NGSIM dataset, comprising public vehicle trajectory data, is widely utilized for research on traffic flow and vehicle movements models [38]. As our study objective focuses on urban intersection scenes, data from Lankershim Boulevard in Los Angeles, CA, and Peachtree Street in Atlanta, GA, in the NGSIM are spontaneously chosen for training and verifying the proposal we develop, as shown in Figure 6. The size of the used dataset is 10,329 frames collected at 10 HZ with a total of 30 min of samples available. The dataset involves two time periods (12:45 P.M. to 1:00 P.M. and 4:00 P.M. to 4:15 P.M.), including scenarios with high and sparse vehicle density. Table 1 illustrates some of the data fields and explanations. In addition to the typical vehicle trajectory attribute, the dataset also contains signal phase information for each interaction, which greatly benefits our experiment.

Table 1. Description of key elements in the dataset.

Serial No.	Data Fields	Explanations
1	Vehicle _{ID}	Vehicle ID
2	Local _X	X value in regional coordinate frame
3	Local _Y	Y value in regional coordinate frame
4	v _{Length}	Vehicle length
5	v _{Width}	Vehicle width
6	v _{Vel}	Vehicle velocity
7	v _{Acc}	Vehicle acceleration
8	Lane _{ID}	The lane number of the vehicle
9	Int _{ID}	The number of intersection

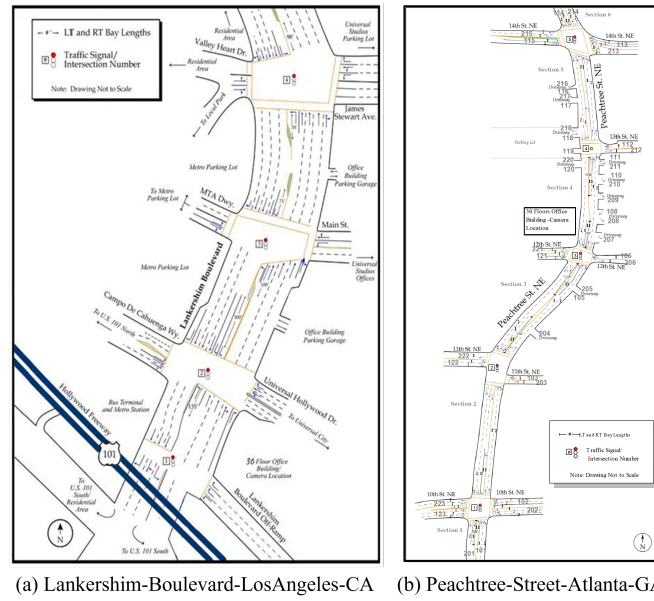


Figure 6. Views of study areas in the dataset.

5.2. Benchmark Model Comparison and Estimation Metrics

5.2.1. Baselines

We compare the performance of the proposed model with the following benchmark models:

- Constant Velocity (CV) model [39]: It assumes a constant speed for the vehicle and predicts its future trajectory based on historical states. In our work, with the 3 s historical trajectory input, the next 5 s movements are indicated through calculating the trajectory average speed of the previous time window.
- GRU model [40]: The model is a common GRU network with no aggregation mechanism, and it considers all trajectories to be independent of each other.
- Vanilla LSTM (V-LSTM) model [23]: This baseline conducts the vehicle trajectory task with an LSTM-based encoder–decoder process. Generally, the vehicle historical trajectories are taken into the encoder, and the decoder outputs the vehicle anticipation positions.
- Generative Adversarial Imitation Learning GRU (GAIL-GRU) model [41]: This baseline uses ground truth data from surrounding vehicles as input in the prediction task.

5.2.2. Estimation Metrics

To quantitatively measure the prediction capability of the proposed scheme and benchmark models, we adopt the root mean square error (RMSE) to weight the differences between ground truth and predicted results, which is calculated as follows,

$$RMSE_t = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_t[i] - \tilde{Y}_t[i])^2} \quad (11)$$

where Y_{ti} denotes the real-world position of vehicle i at moment t , and $\tilde{Y}_t[i]$ is the vehicle i prediction coordinates in time t . The RMSE is reported in every second of the prediction time window, not as mean values over the entire horizons.

5.3. Implementation Details

5.3.1. Input Description

In the implementation, we downsample the initial dataset by a factor of 2. Thus, the historical horizon T_H is set to be 3 s with 15 frames, and the prediction time step T_F is 5 s with 25 frames. We slice the raw trajectory into a sequence with a length of 40 frames

with a sliding window of stride 1; the trajectory slices shorter than 20 frames are discarded to ensure the proper input–output.

5.3.2. Training

The attention-based spatial–temporal fusion graph module consists of three AST-GCN blocks and one Enc-GRU block. Each AST-GCN block contains one spatial GCN and one temporal GCN to implement fast state propagation with 64 channels. To prevent overfitting, each GCN is normalized by BN and randomly drops information with a probability of 0.5. In the case of the Enc-GRU block, the encoder employs a one-layer GRU network with 30 hidden units. The input of the encoder is consistent with the output of the AST-GCN block, so there are 64 channels in the encoder. We execute the proposed method in PyTorch as introduced in [42]. The Adam optimizer is leveraged with a learning rate of 0.01, which is reduced by multiplying 0.1 after every 50 epochs. For each training session, we set the batch size to be 64, and the total number of epochs for training is set to 100. The data used for the experiment are split into 7:1:2 for training, validating, and testing.

5.3.3. Time Consumption

The accuracy and real-time requirements of trajectory prediction are very high. We compare the proposed TPASTGCN model with the RNN-based V-LSTM model and GRU model for data prediction, and the time consumption is compared, as shown in Table 2, which shows the average time consumption for predicting each piece of data. Although the CV baseline model is computationally efficient, it can only predict the vehicle trajectory for a short period of time, and the vehicle trajectory prediction for a long period of time is extremely poor. The proposed model can achieve a faster training process and easier convergence.

Table 2. Model calculation time consumption.

Time Consumption (s)				
CV	GRU	V-LSTM	GAIL-GRU	TPASTGCN
0.02864	0.87395	0.08742	0.99733	0.07945

5.4. Ablation Study

In this section, the effectiveness of traffic rules and neighbor agents' features are investigated. There are rarely reported works considering the traffic rules as input of trajectory prediction problem for autonomous vehicles. Therefore, it is necessary to verify the potential performance of these features to gain innovative insights. Furthermore, we will also carefully explore the role of proximate agents' characteristics of the target vehicle on our predictions.

5.4.1. Effectiveness of Traffic Rules Features

In practice, the experiment is conducted by changing the model settings in terms of input, where traffic rules are included or excluded. As shown in Figure 7, we report the RMSE values at each time step to compare the model results. It is worth noting that the prediction errors in considering traffic rules are always significantly lower than those after removing the traffic rules condition in all prediction horizons. On average, considering traffic rules improves model performance by approximately 11%. In addition, with the prediction step growing, the deviation of the two group experiments increases, indicating that the influence of traffic rules on driving decision making increases. This is because when approaching the intersection, the drivers focus not only on the surrounding agents but also the traffic light phase status, leading to continual interactions. Consequently, it can be concluded that the traffic rules features are a benefit of the proposed scheme and ought to be meticulously weighted in vehicle trajectory anticipation issues.

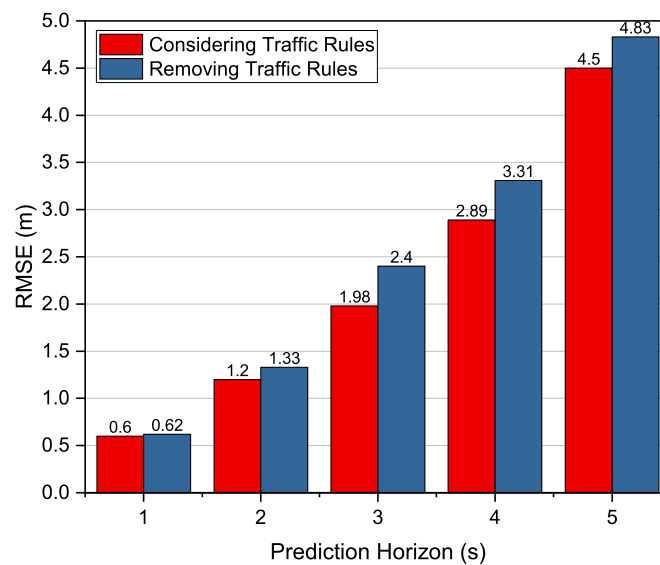


Figure 7. Effectiveness of traffic rules features investigation.

5.4.2. Neighbor Features with Different Distance Thresholds

Whether or not there are multi-agent interactions is measured by the distance threshold D as discussed in Section 4, so we should estimate the effect of changing the values and search for the optimal one from the overall perspective. In our ablation tests, the distance threshold is set to be 0 feet, 30 feet and 60 feet separately. Particularly, when $D = 0$, it means that there are no other interactions except for the target vehicle itself. As depicted in Figure 8, the scenario of $D = 0$ has the largest amount of RMSE errors compared with models that take other movements and traffic rules into account. This indicates that considering intersection mechanisms helps improve our model's prediction accuracy. However, in fact, the target vehicle is more likely to take adjacent agents' features in consideration in its decision making. Thus, if we blindly increase the distance threshold to embrace spare objects, it will result in huge computational complexity but affect few outcomes, which is obviously laborious and unnecessary. As we can see from Figure 8, the model with setting $D = 30$ achieves the lowest evaluation errors in all prediction horizons. Accordingly, we are able to deduce the fact that the more appropriate perception range for most vehicles is around 30 feet when they are driving.

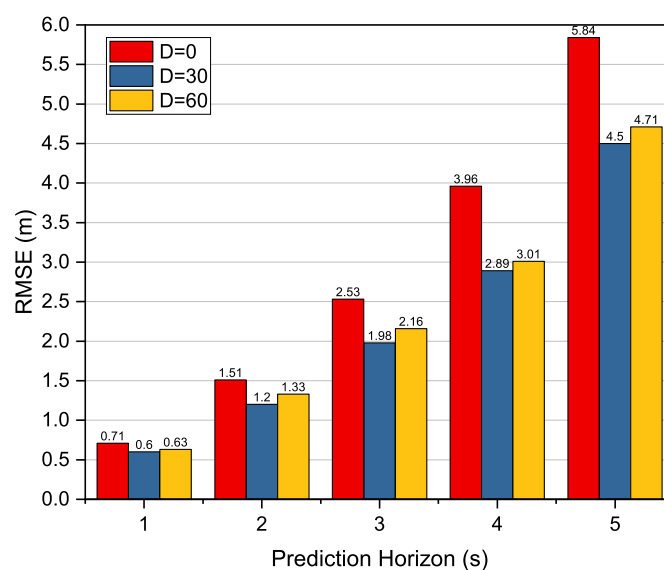


Figure 8. The evaluation of neighbor features with different distance thresholds.

5.5. Quantitative Analysis of Prediction Results

In this section, we compare the developed model and baseline models in terms of RMSE as presented in Figure 9. We can observe that in the first second of the prediction horizon, the RMSE value is approximately 0.65 m, where our proposal performs slightly better than others. Nevertheless, when the prediction time increases to 3 s, the baselines are failing to handle the prediction issue, and its evaluation errors gradually exceed 2.00 m. In contrast, the TPASTGCN model remains robust and maintains the RMSE value at a lower level. Note that the deviation grows sharply for the baseline model in 3–5 s, demonstrating that these models are not suitable for long-term anticipation tasks. However, our proposal constantly outperforms its partner models with over 20.00% accuracy by raising its RMSE value slowly. This indicates that our model acts out for generating future trajectories whether for short-term or long-term windows.

In general, the traditional CV model has the lowest prediction performance, since it does not capture the spatial-temporal features of multi-agents and is only focused on intrinsic states. Compared with C-LSTM and GRU models, the GAIL-GRU exhibits a notable improvement especially in the previous 4 seconds. This is because the GAIL-GRU method has successfully considered the impact of surrounding vehicles and incorporates them into the model. However, as shown in Figure 9, one of the main drawbacks of the GAIL-GRU model lies in its limited capacity in describing longer sequence features and ignoring traffic rules, which results in a rapid growth in RMSE value at later stages. In contrast, our scheme highlights excellent prediction performance with over 23% accuracy in the baselines, owing to the fact that we benefit from a graph topology approach to fully synthesize spatial-temporal trajectory information and multi-agent intersection mechanisms, which are critical features in extremely dynamic and variable intersection scenarios.

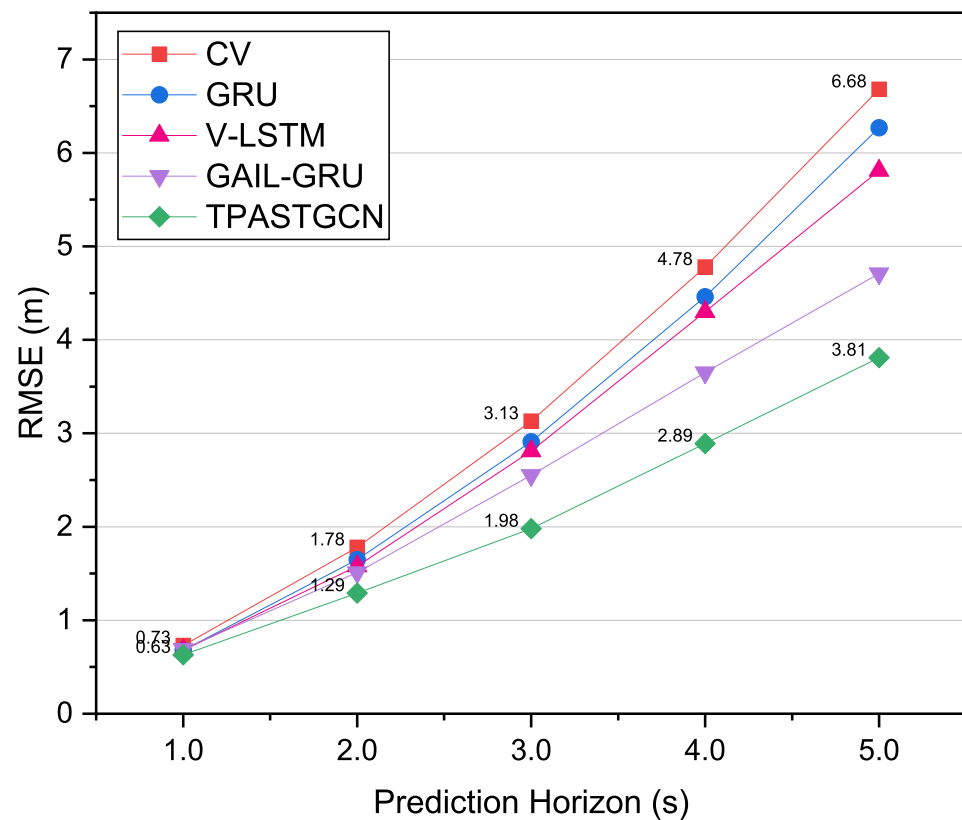


Figure 9. Comparison of RMSE for proposal model and baselines.

5.6. Qualitative Analysis of Prediction Results

To further examine the performance of TPASTGCN, several representative trajectories predicted for different scenes are visualized in Figure 10. The LocalX along with LocalY in

Figure 10 represent the traveling displacement distance of the vehicle in two directions on the ground, which is the trajectory of the vehicle on the ground. The multiple tracks the proposed model generates with 5 s are expressed in blue, while the historical paths with 3 s in yellow and red curves denote the ground truth. Figure 10a depicts the scenario of a vehicle traveling straight, which shows that the future potential trajectory determined by the proposed model is strongly consistent with the ground truth. This implies that the proposed scheme has produced the correct anticipated intention and then accurately works in trajectory construction in the next time periods. As for the vehicle's slow turn in Figure 10b, a good agreement is also found between the predictions and real-world track. The reason that our predicted future trajectory is shorter than the truth in this sample is because the vehicle travels fast when it leaves the intersection, which is beyond our concerned ranges.

Especially, Figure 10c,d present two extreme scenes, in which the vehicle suddenly and massively engages in complicated movements without any prior knowledge. As can be found in Figure 10e,f, LocalX along with LocalY in the historical track do not present obvious turning and acceleration behavior. However, our proposal successfully captures the changes and delivers the prediction trend, which is similar to the ground truth. The results displayed in Figure 10d,g together with Figure 10h demonstrate the same observation, where the predicted paths have the correct traveling intention and the velocity arrow. Combined with these intuitive analyses and previous ablation tests, we indicate the derived results are learning from multi-agent interaction mechanisms and are perfectly disposed by our developed model, as there is no prior knowledge. The above series consequence highlights that our network not only provides an adequate solution in prediction tasks with robustness but also is competent in the work of deducing vehicle traveling intentions.

5.7. Interactions and Attention Mechanism Evaluation

There is a common phenomenon in which the model usually lacks sufficient explanations for black-box features in deep-learning. Ref. [43] Since our scheme is thoroughly designed on basis of the graph, the operating mechanisms are generally interpretable in figures. As emphasized above, our proposal imposes an attention mechanism to the adjacency matrix to learn the heterogeneous intensities of interactions, which makes the graph dynamic in edges and nodes simultaneously with the purpose of capturing spatial-temporal features. To visually represent these relationships, we take the sample of *Vehicle_Id* = 856 on Peachtree Street as the target vehicle and build its interaction diagrams in Figure 11. The left side of Figure 11 is the real-world traffic scene, while the middle denotes the interaction dependency that shows the attention weights and connections. Meanwhile, the distance between nodes is balanced by the length of the connecting line.

The target vehicle movements impacted by the surrounding agents and historical features along with traffic rules can be clearly observed in Figure 11. When the vehicle travels through the intersection, its previous status has the greatest impact on the trajectory, since the drivers tend to have moderate behavior. And this is what most existing works focused on in trajectory anticipation problems. Interestingly, the traffic rule comes third among all potential factors, indicating that the validity of our scheme positively benefits from the traffic rule attributes, and these features contain a lot of information. Surrounding agents also show strong interaction with the target vehicle, especially for *Vehicle_ID* = 850, as it is close to the object and in the same lane. The above illustrations demonstrate our framework working processes and how our proposal effectively extracts multilevel features, which further verifies the performance of the developed model in producing accurate prediction results.

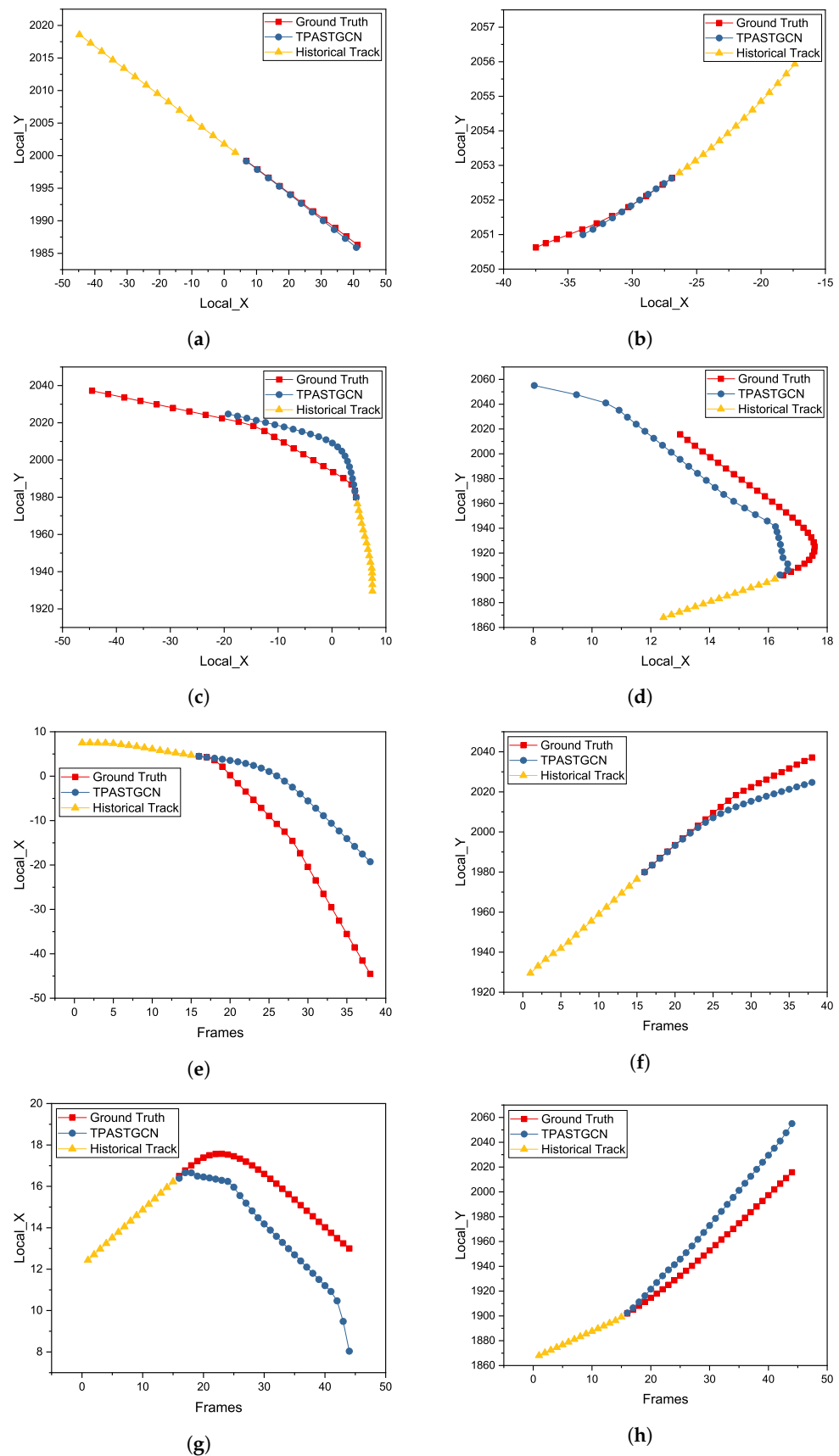


Figure 10. Results of the prediction using TPASTGCN with 3 s windows input and 5 s trajectories output, where (a–d) display different movements and (e–h) contain the description of the results index (c,d) in frames.

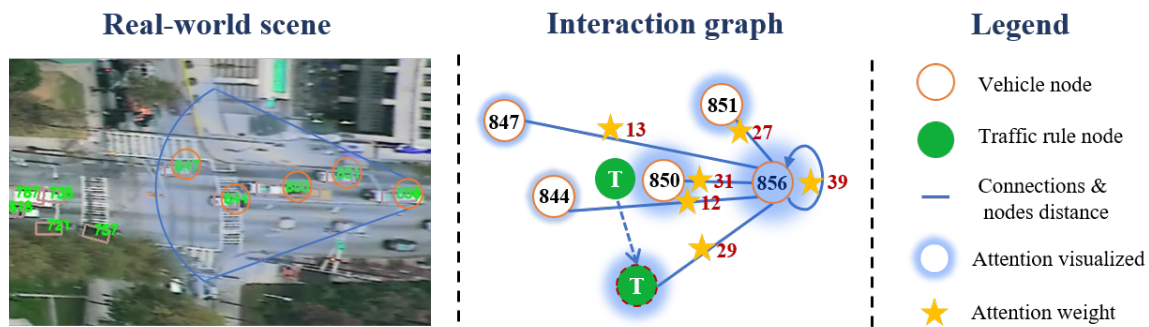


Figure 11. Explanations of the model's working mechanisms.

6. Conclusions

Predicting the trajectory of surrounding traffic is one of the essential capabilities of autonomous vehicles [44]. The vehicle senses the historical trajectories of moving objects and predicts their future behavior intentions. Taking the predicted trajectories of moving objects such as obstacles, pedestrians, and non-motorized vehicles as input, autonomous vehicles can make more reasonable driving decisions and plan more reasonable and safe vehicle movement behaviors. In order to predict vehicle trajectories more accurately, it is necessary to provide more accurate and feature-rich vehicle trajectory data and better vehicle trajectory prediction methods. In this paper, for a more accurate prediction of vehicle trajectory, more accurate and feature-rich vehicle trajectory data with a better vehicle trajectory prediction method is given. The main findings are outlined below.

- The prediction method considers the interaction between traffic agents and the influence of traffic rules on vehicle trajectories, such as the signal phase. This research proposes a novel framework, called TPASTGCN, which fully utilizes multilevel features for vehicle trajectory anticipation tasks. The effectiveness of our proposal is validated by a real-world dataset and several mainstream methods, which demonstrates the excellent performance in generating future trajectories at the highly complicated intersection scenes. The results manifest that compared with the baseline models, the proposed scheme achieves the prediction accuracy improvement with over 23% on average and 61% at most.
- Additionally, the experiment also indicates that the model is able to address the intention prediction issue to some extent, and the model and its effectiveness are superior to existing trajectory prediction methods in intersection scenarios. But the trajectory prediction model accuracy has further room for improvement. A vehicle's future trajectory is under the influence of more factors than we are able to consider here, such as vehicle interactions, driver intentions, traffic rules, etc. In this paper, only the factors of vehicle interaction and signal phasing are considered, but the remaining time of traffic signals may improve the accuracy of prediction. Thus, we encourage the authors of future research to model more influencing factors within the the model.
- The accuracy and real-time requirements for the trajectory are very high. The proposed model achieves an equilibrium of prediction speed and accuracy, but its effectiveness on real vehicles is not verified. In the future, we need to conduct extensive testing, validation, and additional research at the vehicle end to verify the validity of our models, and we intend to build spatial-temporal navigation maps and make path-planning tasks for autonomous vehicles based on this study to improve the safety and comfort of autonomous driving.

Author Contributions: Conceptualization, H.L. and Y.R.; Methodology, W.C.; Software, K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2021YFB1600504).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in [NGSIM], reference number [38].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhao, Y.; Wang, Y.; Cheng, X.; Chen, H.; Yu, H.; Ren, Y. RFAP: A Revocable Fine-Grained Access Control Mechanism for Autonomous Vehicle Platoon. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 9668–9679. [\[CrossRef\]](#)
2. Yu, H.; Liu, R.; Li, Z.; Ren, Y.; Jiang, H. An RSU Deployment Strategy based on Traffic Demand in Vehicular Ad Hoc Networks (VANETs). *IEEE Internet Things J.* **2022**, *9*, 6496–6505. [\[CrossRef\]](#)
3. Mo, X.; Huang, Z.; Xing, Y.; Lv, C. Multi-Agent Trajectory Prediction With Heterogeneous Edge-Enhanced Graph Attention Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 9554–9567. [\[CrossRef\]](#)
4. Lefèvre, S.; Vasquez, D.; Laugier, C. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH J.* **2014**, *1*, 1. [\[CrossRef\]](#)
5. Li, X.; Ying, X.; Chuah, M.C. GRIP: Graph-based Interaction-aware Trajectory Prediction. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 3960–3966. [\[CrossRef\]](#)
6. Chandra, R.; Guan, T.; Panuganti, S.; Mittal, T.; Bhattacharya, U.; Bera, A.; Manocha, D. Forecasting Trajectory and Behavior of Road-Agents Using Spectral Clustering in Graph-LSTMs. *IEEE Robot. Autom. Lett.* **2020**, *5*, 4882–4890. [\[CrossRef\]](#)
7. Mozaffari, S.; Al-Jarrah, O.Y.; Dianati, M.; Jennings, P.; Mouzakitis, A. Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 33–47. [\[CrossRef\]](#)
8. Sheng, Z.; Xu, Y.; Xue, S.; Li, D. Graph-Based Spatial-Temporal Convolutional Network for Vehicle Trajectory Prediction in Autonomous Driving. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 17654–17665. [\[CrossRef\]](#)
9. Shi, X.; Qi, H.; Shen, Y.; Wu, G.; Yin, B. A Spatial-Temporal Attention Approach for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4909–4918. [\[CrossRef\]](#)
10. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3848–3858. [\[CrossRef\]](#)
11. Ren, Y.; Jiang, H.; Ji, N.; Yu, H. TBSM: A traffic burst-sensitive model for short-term prediction under special events. *Knowl.-Based Syst.* **2022**, *240*, 108120. [\[CrossRef\]](#)
12. Park, S.H.; Kim, B.; Kang, C.M.; Chung, C.C.; Choi, J.W. Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1672–1678. [\[CrossRef\]](#)
13. Wang, L.; Zhang, L.; Yi, Z. Trajectory Predictor by Using Recurrent Neural Networks in Visual Tracking. *IEEE Trans. Cybern.* **2017**, *47*, 3172–3183. [\[CrossRef\]](#)
14. Xing, Y.; Lv, C.; Cao, D. Personalized Vehicle Trajectory Prediction Based on Joint Time-Series Modeling for Connected Vehicles. *IEEE Trans. Veh. Technol.* **2020**, *69*, 1341–1352. [\[CrossRef\]](#)
15. Su, Y.; Du, J.; Li, Y.; Li, X.; Liang, R.; Hua, Z.; Zhou, J. Trajectory Forecasting Based on Prior-Aware Directed Graph Convolutional Neural Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 16773–16785. [\[CrossRef\]](#)
16. Rainbow, B.A.; Men, Q.; Shum, H.P.H. Semantics-STGCNN: A Semantics-guided Spatial-Temporal Graph Convolutional Network for Multi-class Trajectory Prediction. In Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 17–20 October 2021; pp. 2959–2966. [\[CrossRef\]](#)
17. Zyner, A.; Worrall, S.; Ward, J.; Nebot, E. Long short term memory for driver intent prediction. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 1484–1489. [\[CrossRef\]](#)
18. Shirazi, M.S.; Morris, B. Observing behaviors at intersections: A review of recent studies & developments. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Republic of Korea, 28 June–1 July 2015; pp. 1258–1263. [\[CrossRef\]](#)
19. Lefèvre, S.; Laugier, C.; Ibañez-Guzmán, J. Exploiting map information for driver intention estimation at road intersections. In Proceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 5–9 June 2011; pp. 583–588. [\[CrossRef\]](#)
20. Streubel, T.; Hoffmann, K.H. Prediction of driver intended path at intersections. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, 8–11 June 2014; pp. 134–139. [\[CrossRef\]](#)
21. Qian, L.P.; Feng, A.; Yu, N.; Xu, W.; Wu, Y. Vehicular Networking-Enabled Vehicle State Prediction via Two-Level Quantized Adaptive Kalman Filtering. *IEEE Internet Things J.* **2020**, *7*, 7181–7193. [\[CrossRef\]](#)
22. Vemula, A.; Muelling, K.; Oh, J. Social attention: Modeling attention in human crowds. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 4601–4607.
23. Altche, F.; Fortelle, A.D. An LSTM network for highway trajectory prediction. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 353–359.

24. Kim, B.; Kang, C.M.; Kim, J.; Lee, S.H.; Chung, C.C.; Choi, J.W. Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 399–404. [\[CrossRef\]](#)
25. Phillips, D.J.; Wheeler, T.A.; Kochenderfer, M.J. Generalizable intention prediction of human drivers at intersections. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 1665–1670. [\[CrossRef\]](#)
26. Deo, N.; Trivedi, M.M. Multi-Modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1179–1184. [\[CrossRef\]](#)
27. Zhao, T.; Xu, Y.; Monfort, M.; Choi, W.; Baker, C.; Zhao, Y.; Wang, Y.; Wu, Y.N. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2019; pp. 12126–12134.
28. Mercat, J.; Gilles, T.; Zoghby, N.E.; Sandou, G.; Beauvois, D.; Gil, G.P. Multi-Head Attention for Multi-Modal Joint Vehicle Motion Forecasting. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 9638–9644. [\[CrossRef\]](#)
29. Zhang, T.; Song, W.; Fu, M.; Yang, Y.; Wang, M. Vehicle Motion Prediction at Intersections Based on the Turning Intention and Prior Trajectories Model. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 1657–1666. [\[CrossRef\]](#)
30. Zhong, Z.; Luo, Y.; Liang, W. STGM: Vehicle Trajectory Prediction Based on Generative Model for Spatial-Temporal Features. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 18785–18793. [\[CrossRef\]](#)
31. Xie, X.; Zhang, C.; Zhu, Y.; Wu, Y.N.; Zhu, S.-C. Congestion-aware Multi-agent Trajectory Prediction for Collision Avoidance. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 13693–13700. [\[CrossRef\]](#)
32. Carrasco, S.; Llorca, D.F.; Sotelo, M.A. SCOUT: Socially-Consistent and UndersTandable Graph Attention Network for Trajectory Prediction of Vehicles and VRUs. In Proceedings of the 2021 IEEE Intelligent Vehicles Symposium (IV), Nagoya, Japan, 11–17 July 2021; pp. 1501–1508. [\[CrossRef\]](#)
33. Liu, C.; Chen, Y.; Liu, M.; Shi, B.E. AVGCN: Trajectory Prediction using Graph Convolutional Networks Guided by Human Attention. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 14234–14240. [\[CrossRef\]](#)
34. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 922–929. [\[CrossRef\]](#)
35. Wang, S.; Zhang, M.; Miao, H.; Peng, Z.; Yu, P.S. Multivariate Correlation-aware Spatio-temporal Graph Convolutional Networks for Multi-scale Traffic Prediction. *ACM Trans. Intell. Syst. Technol.* **2022**, *13*, 1–12. [\[CrossRef\]](#)
36. Peng, H.; Wang, H.; Du, B.; Bhuiyan, M.Z.A.; Ma, H.; Liu, J.; Wang, L.; Yang, Z.; Du, L.; Wang, S.; et al. Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Inf. Sci.* **2022**, *521*, 277–290. [\[CrossRef\]](#)
37. Zhang, Z. Improved Adam Optimizer for Deep Neural Networks. In Proceedings of the 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 4–6 June 2018; pp. 1–2. [\[CrossRef\]](#)
38. NGSIM—Next Generation Simulation [EB/OL]. Available online: <https://ops.fhwa.dot.gov/trafficanalysis/tools/ngsim.htm> (accessed on 1 October 2022).
39. Zhao, Y.; Yu, H.; Yang, Y.; Xu, L.; Pan, S.; Ren, Y. Flexible and Secure Cross-Domain Signcrypted Data Authorization in Multi-Platoon Vehicular Networks. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–11. [\[CrossRef\]](#)
40. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.2014.
41. Ho, J.; Ermon, S. Generative adversarial imitation learning. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
42. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. In Proceedings of the Advances in Neural Information Processing Systems Workshops, Long Beach, CA, USA, 4–9 December 2017; pp. 1–4.
43. Jiang, H.; Ren, Y.; Fang, J.; Yang, Y.; Xu, L.; Yu, H. SHIP: A State-Aware Hybrid Incentive Program for Urban Crowd Sensing with For-Hire Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–13. [\[CrossRef\]](#)
44. Schöller, C.; Aravantinos, V.; Lay, F.; Knoll, A. What the constant velocity model can teach us about pedestrian motion prediction. *IEEE Robot. Autom. Lett.* **2020**, *5*, 1696–1703. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.