Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/OJITS.xxxx

Risk-Aware Stochastic Vehicle Trajectory Prediction with Spatial-Temporal Interaction Modelling

Yuxiang Feng[1] (Member, IEEE), Qiming Ye[2], Eduardo Candela[1], Jose Escribano-Macias[1], Bo Hu[3], Yiannis Demiris[4] (Senior Member, IEEE), Panagiotis Angeloudis[1]

[1]Centre for Transport Engineering and Modelling, Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK
[2]Future Cities Laboratory Global, Singapore Hub, Singapore-ETH Centre at CREATE, 1 Create Way, 138602, Singapore
[3]School of Vehicle Engineering, Chongqing University of Technology, 401135, Chongqing
[4]Personal Robotics Laboratory, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK

CORRESPONDING AUTHOR: Yuxiang Feng (e-mail: y.feng19@imperial.ac.uk).

ABSTRACT Autonomous vehicles need to continuously analyse the driving context and establish a comprehensive understanding of the dynamic traffic environment. To ensure the safety and efficiency of their operations, it would be beneficial to have accurate predictions of surrounding vehicles' future trajectories. AVs can adjust their motions proactively to improve road safety and comfort with such information. This paper proposes a novel approach to predict the future trajectories of interacting vehicles, through a model of potential spatial-temporal interactions. A unique kernel function that emphasises risk-awareness was developed to extract spatial dependencies. The established model was trained and evaluated with the publicly available Highway Drone Dataset and Intersection Drone Dataset. The performance of the developed model was assessed with eight state-of-the-art methods. An ablation study and safety analysis were also conducted to evaluate the proposed risk-awareness kernel function. Results show that the proposed model's inference speed is over eight times faster than the commonly used LSTM-based models. It also achieves an improvement of over 8% in prediction accuracy when compared with the state-of-the-art model.

INDEX TERMS Stochastic trajectory prediction, Risk awareness, Spatial-temporal modelling, Autonomous Vehicles

I. INTRODUCTION

Recent studies have shown that human factors, such as speeding, drunk driving, and distracted driving, have the strongest influence and are responsible for 80-90% of road traffic accidents [1]. Thus, replacing human drivers with more reliable solutions would potentially mitigate the majority of catastrophic traffic events by reducing human error. This potential in improving travel safety is one of the primary anticipations that motivates research and investments into autonomous vehicles (AVs) and advanced driver-assistance systems (ADAS) [2].

Most existing AV controllers are designed in a reactive manner, therefore focusing on the current state of other road users [3]. While this reactive design can suffice the goal of navigating AVs without collision under most circumstances, their inability to infer driver intent can lead to conservative driving strategy and affect traffic efficiency. Instead, if AVs were to predict surrounding traffic conditions based on past information [4], they would have more opportunities to proactively plan and execute safety manoeuvres, improving road safety and driving comfort while minimising conflicts with other road users.

Predicting trajectories accurately requires considering a multitude of factors that influence vehicle motion behaviour. These factors can be broadly categorised into two types: map-based and map-free. Both categories use historical trajectories of vehicles to predict their future paths, with the former also integrating HD maps. However, HD maps, while providing valuable contextual information, have started to reveal some issues in their application [5]. Similar to other AV systems, researchers have begun to investigate how to conduct trajectory predictions without depending on

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

HD maps. Recent studies [6], [7] have shown that map-free methods can surpass map-based approaches in terms of both prediction accuracy and processing speed. Consequently, this paper exclusively concentrates on map-free methods.

Meanwhile, it should be noted that while human drivers' behaviours tend to preserve some tendencies, but they are not deterministic and their reactions to the same driving scenario may differ at each time [8]. Factors affecting this behaviour include the driver's eagerness to finish the trip, time of the day, weather conditions, among others [9]. More importantly, even if the driver maintains the same intention, the execution could differ in speed and pattern, resulting in different manoeuvres [10]. Thus, this study generates a stochastic multimodal prediction of vehicles' future trajectory to ensure AVs' safety operation.

While many studies have been proposed to predict vehicle trajectories [3], [4], [11]–[26], there is a lack of a thorough comparison of these methods, especially for deep learning studies [3], [4], [12], [15]–[26]. Meanwhile, although collision risks were estimated in some studies [11], [13], they were mainly used for the selection of candidate manoeuvres, which would restrict the influence of driver's risk awareness on their driving decision.

Furthermore, in many existing research studies, spatial relations are commonly treated as fully connected, without distinguishing their individual influences. This approach significantly limits the interpretability of the model, making it challenging or even impossible to comprehend the results it produces. This lack of interpretability poses a significant obstacle to the safety certification and widespread adoption of AVs. Addressing this issue, Explainable Artificial Intelligence (XAI) provides a solution by enabling human users to better understand, interpret, and trust the outputs generated by AI-powered systems. By incorporating explainability, we can ensure that only meaningful variables contribute to the output, thereby guaranteeing that the models capture accurate causality. In order to enhance explainability, our research introduces risk-awareness in spatial relations, thus improving the overall explainability of the model.

The primary contributions of this study are outlined as follows:

1) We propose a novel approach to model the spatial-temporal interactions for vehicle trajectory prediction, which achieves an improvement of over 15% in prediction accuracy when comparing with the state-of-the-art model.
2) A unique kernel function that emphasises risk-awareness is developed to dynamically extract spatial dependencies between vehicles in the scene. The influence of this kernel function is evaluated in an ablation study.
3) We create an enhanced safety protection layer based on the stochastic vehicle trajectory predictions, which evolves with the prediction horizon to incorporate the increasing uncertainties in predicted positions.
4) A unified comparison analysis is conducted to evaluate the proposed method's prediction accuracy, inference speed and distributional performance with state-of-the-art models.

This paper is structured into five sections. Section 2 briefly reviews some relevant literature, while Section 3 explains the adopted methodology. Afterwards, the obtained results and discussion of analysis are presented in Section 4. Finally, Section 5 discusses the findings and limitations of this study and suggests potential future works.

A. Related Work

Extensive research efforts have taken place over the last decade in the area of AV safety and trajectory prediction, underpinned by rapid advances in sensing technologies and the intense pace of AV deployment efforts. For the purposes of this study, we conducted a literature review for approaches based on deep-learning. An overview of reviewed literature is presented in Table 1.

B. Deep learning Studies

Trajectory prediction is fundamentally a time-series classification or generation problem, which makes it particularly suitable for the application of deep learning techniques. Long Short-Term Memory (LSTM) methods, based on Recurrent Neural Network (RNN) architectures, have been especially prominent amongst the reviewed literature, as they are capable of extracting long-term relations amongst the various actors in their models. Existing LSTM-based models can consider fixed number of surrounding vehicles [3], [15], [16] or dynamically capture them over an occupancy grid [17].

However, although these models can implicitly infer the dependencies between vehicles, they can lose information of their relative positions. To compensate this deficiency, Deo and Trivedi [12] enhanced LSTM to improve the extraction of spatial relations. They added convolutional social pooling layers, which generate a context vector consisting of a compact representation of vehicle interactions. Instead of focusing only on vehicles, He et al. [19] used Multiple Layer Perceptron (MLP) and LSTM to develop a unifying framework that can predict trajectories of different road agents, such as vehicles, pedestrians and cyclists.

To better extract spatial relations, graph-based methods have been increasingly adopted for trajectory prediction. They denote vehicles as nodes, with their interactions represented using edges. Spatial information is then captured using Graph Convolutional Networks (GCNs) or Graph Attention Networks (GATs). For example, Zhao et al. [20] assumed a full connection between all vehicles in the scene and applied a set of two-layer GCNs to capture their spatial relations. However, these fully connected edges lead to equivalent interactions among vehicles, regardless of their respective positions, which is not realistic in the real-world driving context. To better address the different influences of interactions, Jeon et al. [21] used Edge-enhanced Graph

TABLE 1. Summary of Literature

Author	Approach	Features
Altché and Fortelle [3]	LSTM	nine surrounding vehicles
Woo et al. [15]	LSTM	lane-changing manoeuvre, four surrounding vehicles
Deo and Trivedi [16]	LSTM	stochastic prediction, six surrounding vehicles
Kim et al. [17]	LSTM+occupancy grid	region of interest
Deo and Trivedi [12]	Enhanced LSTM	convolutional social pooling
Zhao et al. [18]	LSTM+convolutional network	multi-agents prediction
He et al. [19]	LSTM+MLP	different road agents
Messaoud et al. [4]	LSTM+attention	multi-head attention
Zhao et al. [20]	LSTM+graph	GCN
Jeon et al. [21]	LSTM+graph	Edge-enhanced GCN
Li et al. [22]	LSTM+ graph	temporal convolutional layer
Li et al. [23]	GRU+graph	additional trainable graph
Dong et al. [24]	LSTM+graph	GAT, different road agents
Zhang et al. [25]	TCN	lane-changing manoeuvre
Strohbeck et al. [26]	TCN	rasterised image and vehicle state information as input
Li et al. [27]	TCN, pooling, attention	local and global interaction
Chen et al. [28]	graph transformer	non-autoregressive approach with low prediction latency
Liao et al. [29]	LSTM, MLP, attention, DGG	incorporate human behaviour and traffic psychology
Liao et al. [30]	LSTM, GCN, Linformer, pooling, DGG	adaptive structure-aware interactive graph convolutional network
Chen et al. [31]	LSTM, attention, GLU	intention-specific feature fusion

Convolutional Networks, a variant of conventional GCNs, for the prediction of vehicle trajectories. This approach obtains a weighted adjacency matrix through the edge-enhanced attention mechanism, which calculates the edge feature matrix using two paired vehicles' relative position and velocity. Another GCN-based study was implemented by Li et al. [23]. Unlike the previous two studies, they incorporated a 2D Temporal Convolutional layer after the graph operation. They also added a trainable graph to the fixed graph representation of the scene to mitigate performance degradation in urban traffic scenarios [22]. The temporal correlations in these studies were captured using LSTM [20], [21], [23] and Gated Recurrent Unit (GRU) [22].

While GATs have been used in several studies for pedestrian trajectory prediction, its application to vehicle trajectory prediction is limited, with only one publication identified [24]. Similar to [19], this model can also predict trajectories of other road agents by defining an attention circle and limiting potential conflicts only among the road users within this circle, the radius of which is determined by agent speeds, prediction times, and agent lengths. Additionally, the semantic map was used as an input that also included traffic rule information. Other attention mechanisms are also increasingly used in recent studies. For example, Chen et al. [28] introduce a novel non-autoregressive graph transformer that incorporates a self-attention module to address dynamic variations in social behaviour and a graph attention module to depict interactions between vehicles. This non-autoregressive approach enables the model to achieve both diverse trajectories and low prediction latency. Liao et al. developed two models, BAT [29] and MFTraj [30]. Both models integrate human features into trajectory prediction, incorporating a behaviour-aware module based on dynamic geometric graphs (DGGs) to capture the behavioural features of road users. In a recent study, the Human-Like Trajectory Prediction (HLTP) model integrates human cognition and decision-making processes through an advanced teacher-student knowledge distillation framework [32]. The "teacher" model, featuring an adaptive visual sector, simulates the visual processing capabilities of the human brain, specifically the occipital and temporal lobes. Meanwhile, the "student" model emphasises real-time interaction and decision-making, akin to the roles of the prefrontal and parietal cortices.

Another trend in existing deep learning studies is to use Temporal Convolutional Networks (TCNs) as an alternative for time-series prediction. TCNs employ casual convolutions and dilations to handle sequential data with their temporality and large receptive fields. As a variant of the convolutional neural networks, TCNs can alleviate RNN's accumulated error problem while achieving comparable or even better performance when predicting sequential data [33]. Two TCNs-based studies were identified. The first study [25] focused on predicting vehicle's long-term lane-changing behaviours and trajectories. The prediction accuracy was benchmarked with two traditional neural networks, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). The TCN-based method was found to achieve better performance when considering the prediction accuracy and computational cost. Another TCN-based model was developed by Strohbeck et al. [26]. This model was evaluated using the Argoverse Motion Forecasting Dataset

TABLE 2. Prediction Accuracy Comparison on NGSIM (RMSE in Metre)

Model	Prediction Horizon (s)				
	1	2	3	4	5
CS-LSTM [12]	0.61	1.27	2.09	3.10	4.37
M-LSTM [16]	0.58	1.26	2.12	3.24	4.66
MATF [18]	0.66	1.34	2.08	2.97	4.13
UST [19]	0.58	1.20	1.96	2.92	4.12
UST-180 [19]	0.56	1.15	1.82	2.58	3.45
GISNet [20]	0.33	0.83	**1.42**	**2.14**	3.23
GRIP++ [22]	0.38	0.89	1.45	**2.14**	**2.94**
TCN-SA [27]	0.4	1.1	2.14	2.87	4.08
BAT [29]	**0.23**	**0.81**	1.54	2.52	3.62
MFTraj [30]	0.38	0.87	1.52	2.23	2.95
STDAN [31]	0.42	1.01	1.69	2.56	3.67

and validated with the Argoverse Baseline and a CNN-based Multiple-Trajectory Prediction (MTP) model. Li et al. [27] consider both local and global features in their study. They utilise a social convolutional pooling layer to capture local interaction features between vehicles and a multi-head self-attention layer to capture global interaction features.

C. Performance Evaluation

While some studies were evaluated with different public datasets [24] or collected their private training data [17], Next Generation Simulation (NGSIM) is found to be the most commonly used dataset for trajectory prediction. Among the identified 20 deep learning-based studies, 11 of them published their prediction accuracy with the NGSIM dataset, with a prediction horizon ranging from 1s to 5s. The prediction accuracy of these models evaluated with NGSIM is summarised in Table 2.

Table 2 reveals that no single model achieves the best performance across all prediction horizons. In general, BAT, GISNet, and GRIP++ are the top-performing models. BAT excels in shorter prediction horizons ($\leq$2s), GISNet in the medium range (3-4s), and GRIP++ in longer terms ($\geq$4s). Specifically, BAT improved prediction accuracy by 30.30% and 2.41% compared to the second-best model, GISNet, in short prediction horizons. This improvement highlights the benefit of incorporating driving behaviour in short-term predictions. Conversely, GISNet demonstrates promising performance across all prediction steps, ranking first in the medium range (3-4s) and second in other horizons. It is noteworthy that both GISNet and GRIP++ utilise GCNs to capture spatial relations between vehicles, resulting in significant accuracy enhancements.

Overall, graph-based approaches employing GCNs or GATs demonstrate superior ability to capture spatial relations among vehicles. The surrounding vehicles in most existing studies are either selected within a predefined area of interest [3], [12], [15], [16] or by simulating the effect of sensor detection [17]. While this selection process increases the similarity to the current traffic environment, insufficient information about the entire driving context can cause prediction errors. From human drivers' perspective, their anticipation of driving is highly concerned with the risk of collisions. They tend to pay more attention to road agents posing higher level of risks.

Considering the temporal domain, among the three TCN-based studies [25]–[27], only TCN-SA [27] provides results on the NGSIM dataset, where it achieves a middle level of performance. It is noteworthy that another TCN-based model [26] achieves better prediction accuracy than the UST model [19] on the Argoverse dataset. However, as shown in Table 2, the prediction accuracy of UST is not among the top tier. Therefore, it remains challenging to determine whether LSTM-based or TCN-based models can achieve better prediction accuracy overall. Although TCNs theoretically offer faster inference speeds due to their parallelism [33], there has been limited analysis on this hypothesis within vehicle trajectory prediction studies.

This paper aims to fill these gaps by proposing a novel approach that combines TCNs for capturing temporal correlations and risk-enhanced GCNs for extracting social interactions among vehicles. By evaluating the risk of collision, surrounding vehicles are automatically selected, and their differentiated influence on other vehicles is accounted for, enhancing the accuracy and explainability of the model.

II. Methodology

This section consists of the dataset preparation, the formulation of the research problem, the model's specification and the prediction of future trajectory.

A. Problem Description

Given a set of N vehicles in the scene with their past track information $s_t^i = (x_t^i, y_t^i), i \in \{1, 2, \ldots, N\}$, observed over a period $t \in \{1, 2, \ldots, T_{obs}\}$, this study aims to predict the possible future trajectories of these vehicles $\hat{s}_t^i$ over a time horizon $t \in \{T_{obs+1}, T_{obs+2}, \ldots, T_{pred}\}$ concurrently. The proposed model predicts all vehicle trajectories for ten future frames (2s) based on observations over the past twenty frames (4s).

B. Model Architecture

The architecture of the proposed model is illustrated in Fig. 1. It consists of three paimary components: (i) a GCNs module for extracting spatial features, (ii) a Temporal Convolutional Networks (TCNs) module for capturing temporal correlations, and (iii) a decoder module for predicting future trajectories.

1) Spatial Feature Extraction

To capture spatial interactions, an undirected graph representation of vehicle trajectories is constructed at each time step t, represented as $G_t = (V_t, E_t)$. Each node $v_t^i \in V_t, i \in \{1, 2, \ldots, N\}$ corresponds to a vehicle within the scene, and

This article has been accepted for publication in IEEE Open Journal of Intelligent Transportation Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJITS.2025.3530268

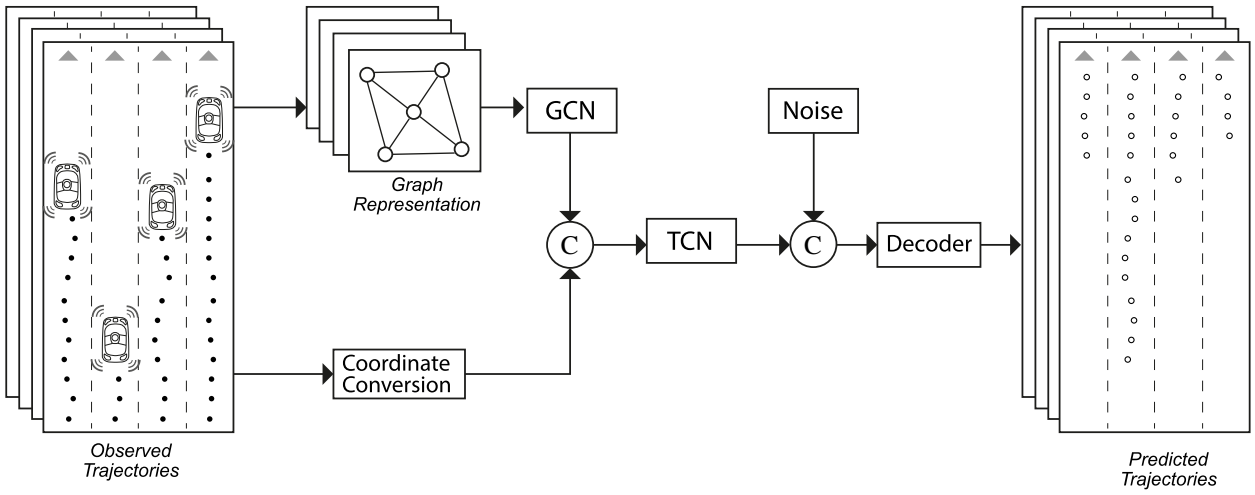

FIGURE 1. **The architecture of the proposed model**

each pair of vehicles $(i, j \in \{1, 2, \ldots, N\})$ is connected via edge $e_t^{ij} \in E_t$. The spatial relations between vehicles within each graph G_t is represented by a weighted adjacency matrix A_t, where the edge weight e_t^{ij} is determined by a risk-awareness kernel function [34]. This kernel incorporates both longitudinal and lateral risks between vehicles, formulated as:

$$e_t^{ij} = \begin{cases} 0, & if\ i = j \\ r_t^{lon_{ij}} * r_t^{lat_{ij}}, & otherwise \end{cases} \tag{1}$$

where $r_t^{lon_{ij}}$ and $r_t^{lat_{ij}}$ quantify the longitudinal and lateral risks between vehicle i and vehicle j at time t, respectively. The computation of these risk values can be described as,

$$r = \begin{cases} 0, & if d \geq d_{min} \geq 0 \\ 1 - \frac{d - d_{min,b}}{d_{min} - d_{min,b}}, & if d_{min} \geq d\, 0 \\ 1, & otherwise \end{cases} \tag{2}$$

$$d_{min} = \left[v_j \tau + \tfrac{1}{2} \tau^2 a^j_{max\_a} + \frac{(v_j + \tau a^j_{max\_a})^2}{2 a^j_{min\_b}} - \frac{v_i^2}{2 a^i_{max\_b}} \right]_+ \tag{3}$$

$$d_{min\_b} = \left[v_j \tau + \tfrac{1}{2} \tau^2 a^j_{max\_a} + \frac{(v_j + \tau a^j_{max\_a})^2}{2 a^j_{max\_b}} - \frac{v_i^2}{2 a^i_{max\_b}} \right]_+ \tag{4}$$

where $[x]_+ = max\{x, 0\}$ for either direction, d is the distance between vehicle i and j, d_{min} is the guaranteed safe distance, d_{min_b} is the minimum safe distance, v_i and v_j are the current velocity of vehicle i and j, τ is the response time, $a^j_{max\_a}$ is the maximum acceleration of vehicle j, $a^i_{max\_b}$ and $a^j_{max\_b}$ are the maximum deceleration of vehicle i and vehicle j, and $a^j_{min\_b}$ is the minimum deceleration of vehicle j. Following recommendation from existing literature, τ is set to $1.5s$ [35]. Following the findings of Bokare and Maurya [36], the recommended accelerating and braking parameters of each agent category are summarised in Table 3.

As the graph representation is undirected, it can be noted that $e_t^{ij} = e_t^{ji}$, which indicates that the influence

TABLE 3. **Acceleration/Braking Parameters for Risk Computation** (m/s^2)

Parameter	car	truck	bus	cyclist	pedestrian
a_{max_a}	2.9	1.0	1.0	2.0	0.5
a_{max_b}	3.9	4.0	4.5	6.0	0.8
a_{min_b}	1.0	0.8	1.0	1.5	0.2

of vehicle interaction is assumed equal between vehicle i and vehicle j. Although this assumption contravenes human driving patterns, as drivers tend to spend more attention to leading vehicles than those at their rears, it should be noted that whatever the relative positions of the vehicles are, the risk posed on them to be involved in a collision should be the same. Moreover, as the proposed prediction model shares past kinematic information of all vehicles within the scene, they have same anticipation of traffic movements and intentions to avoid collisions. Therefore, the undirected graph representation is adopted in this study.

As suggested by Kipf and Welling [57], the adjacency matrix can then be normalised as,

$$\hat{A}_t = \Lambda_t^{-\frac{1}{2}} \widetilde{A}_t \Lambda_t^{-\frac{1}{2}} \tag{5}$$

where $\widetilde{A}_t = A_t + I$ and $\Lambda_t = \sum_j \widetilde{A}_t^{ij}$, I denotes and identity matrix.

The output of the lth layer can be denoted as,

$$H^l = \sigma\left(\hat{A}_t H^{l-1} W^l\right) \tag{6}$$

where W^l is the trainable weights at layer l, H^{l-1} is the output of the $(l-1)$th layer, σ is the activation function.

2) Temporal Feature Extraction

Temporal dependencies are captured using TCNs, which offer a robust alternative to Recurrent Neural Networks (RNNs). Unlike RNNs, TCNs avoid issues like accumulated error propagation and are computationally efficient [33]. A

customised 3-layer TCN is employed to process historical trajectory data across varying temporal scales effectively. The output of the lth layer can be denoted as,

$$H^l = \sigma\left(\sum_{i=0}^{k-1} W^l \cdot H^{l-1}[t - d \cdot i] + B^l\right) \tag{7}$$

where $k = 3$ is the kernel size, W^l is the trainable weights at layer l, t is the time index, $d = [1, 2, 4]$ is the dilation rate, and B^l is the bias of the lth layer.

At each time step, the absolute positions of all vehicles are transformed into a localised coordinate frame and embedded into fixed-length vectors. The spatial features derived from GCNs are concatenated with these positional embeddings, creating combined feature vectors. These vectors are then processed by the TCN module to capture both spatial and temporal relations among vehicles.

3) Future Trajectory Prediction

The variety loss proposed by Gupta et al. [37] is adopted to generate multiple reasonable and realistic future trajectories. A shared random noise z is concatenated with the extracted embeddings before being sent to the decoder as input. The noise is randomly sampled from $\mathcal{N}(0, 1)$. The hidden state vector can be presented as,

$$h = \overline{m} \mid\mid z \tag{8}$$

where $\overline{m}$ represents the combined spatial-temporal features and z denotes the added random noise.

These concatenated vectors are sent to the MLP-based decoder (3 layers) and twenty possible future trajectories are computed at each step.

The model is trained to minimise the variety loss, computed by choosing the trajectory with the minimum ADE to ground truth. The loss function can be denoted as,

$$L_{variety} = \min_k \| \hat{s}_i^k - s_i \|_2, \ k \in \{1, 2, ..., 20\} \tag{9}$$

where s_i is the ground-truth trajectory of vehicle i during the prediction horizon, $\hat{s}_i^k$ is the kth possible trajectories of vehicle i.

C. Evaluation Metrics

To evaluate the performance of the model, ADE, FDE and $Negative\,Log\,Likelihoood\,(NLL)$ are implemented as evaluation metrics. ADE measures the root mean squared error average between the ground truth and the predicted trajectory, while FDE calculates the offset at the endpoints. Mean NLL is used to evaluate the variance and multi-modality between the ground truth trajectory and stochastic predictions [38]. At each timestep, Gaussian Kernel Density Estimate (KDE) is performed to obtain the probability density function of the sampled trajectories, which is then used to compute the average NLL of the ground truth trajectory.

$$ADE = \frac{1}{NT_{pred}} \sum_{i=1}^{N} \sum_{t=1}^{T_{pred}} \| \hat{s}_t^i - s_t^i \|_2 \tag{10}$$

$$FDE = \frac{1}{N} \sum_{i=1}^{N} \| \hat{s}_{T_{pred}}^i - s_{T_{pred}}^i \|_2 \tag{11}$$

$$KDE = \frac{1}{N\sigma\sqrt{2\pi}} \sum_{i=0}^{N} e^{-\frac{1}{2}(\frac{\hat{s}_i - s_i}{\sigma})^2} \tag{12}$$

$$NLL = -\frac{1}{N} \sum_{i=1}^{N} \log P(\hat{s_i}|s_i) \tag{13}$$

III. Results and Discussion

The proposed model was developed with PyTorch, an open-source machine learning Python library. The training and evaluation were implemented on a Desktop PC (CPU: Intel(R) Core(TM) i9-9940X CPU @ 3.30GHz, GPU: 2 x NVIDIA GeForce RTX 2070 Super). The model was trained using Adam optimiser for 200 epochs with a learning rate of 0.001 and a batch size of 128. ReLU was used as the activation function of all networks, and a dropout rate of 0.2 was adopted.

This section has been divided into four parts to present the obtained results. First, an example of the constructed graph is illustrated to demonstrate the proposed risk-awareness kernel function. Second, a sample of the prediction results is presented. Afterwards, a benchmark analysis is conducted to compare the performance of this model with existing studies. Finally, an ablation study is undertaken to evaluate the proposed kernel function further.

A. Data Preparation

This study used two publicly available datasets for training and evaluation, the Highway Drone Dataset (highD) [19] and Intersection Drone Dataset [39]. Both datasets were collected by the same research group at RWTH Aachen University following the same methodology.

The highD dataset was established by recording traffic flow at six different locations on German motorways using drones. Each drone covers a road segment of approximately $420m$ and includes trajectories of 110,000 vehicles, recording 5,600 completed lane changes within the observation area. In total, the dataset comprises 447 driven hours and covers a total driven distance of approximately 45,000km. The inD dataset was recorded at four unsignalised junctions in Aachen, covering areas from 80x40m to 140x70m. In total, the dataset consists of 10 driven hours. It records 11,500 trajectories, comprising pedestrians, bicycles, cars, trucks, and buses.

Compared with the most widely used NGSIM dataset reviewed in the previous section, the highD and inD dataset prevails in several aspects: the data length, vehicle variety, and accuracy of the extracted trajectories. More importantly, it should be noted that the typical positioning error in

This article has been accepted for publication in IEEE Open Journal of Intelligent Transportation Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJITS.2025.3530268

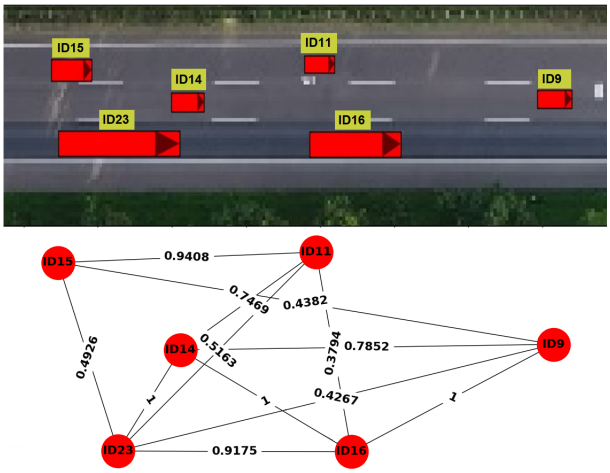

FIGURE 2. **Sample of constructed graph**

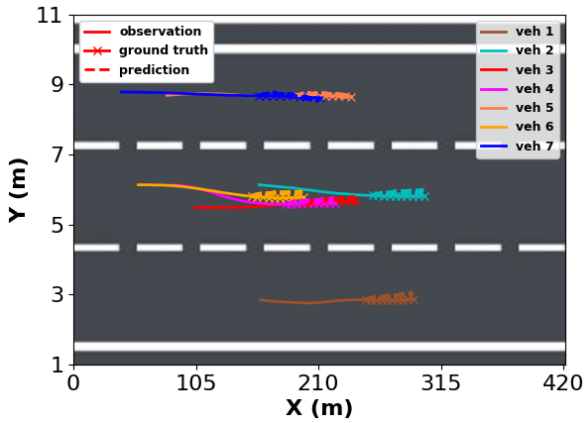

FIGURE 3. **Sample of prediction results**

highD and inD is less than $10cm$, which provides much better accuracy of vehicle trajectories and minimises the requirements for additional post-processing.

During preprocessing, a reconstruction of the data was carried out into the preferred structure to facilitate the proposed study, as well as a downsampling to 5 Hz to reduce the computation. This restructured dataset was divided into three subsets for training, validation and test purposes, with a ratio of 7:1.5:1.5.

B. Risk-Awareness Graph Construction

A sample scene was selected to demonstrate the proposed risk-awareness graph. As shown in Fig. 2, the plot above shows a frame of the vehicle sequence data. While there are more vehicles in this frame, this segment was selected for demonstration. There are six vehicles along this 100m road segment, and they are split into three car lanes. The average speed of these six vehicles is 96.53 km/h, with the maximum speed to be 111.42 km/h of vehicle ID9 and the minimum speed to be 77.12 km/h of vehicle 23.

The graph constructed using the risk-awareness kernel function is shown in the subplot below. The weight of edges ranges from 0 to 1, and a larger weight value indicates a more substantial social influence between vehicles. Noticeably, if the weight of an edge is zero, this edge is then removed for a clear illustration. As shown in the graph, large risk indices occurred between vehicles in the same lane. This is because the vehicles travel fast, and there are no lane-changing events in this selected sample.

Meanwhile, vehicle ID16 and ID23 are associated with larger risk values. This is because the vehicular physic parameters when computing risk index are differentiated to reflect different vehicle types. As both vehicles are trucks, they have severer restrictions on braking capability and are more prone to collisions. Thus, their associated risks are higher.

Moreover, as risk index is defined according to the limit of vehicle dynamics, a positive value can be acquired between vehicles within a more reasonable scope. While this leads to a larger graph at each frame, it ensures a more comprehensive spatial relation is extracted. Thus, AVs can have more thorough anticipations of the scene. Even if vehicles perform unusual risky manoeuvres at a distance, host AVs can still have enough time and space to avoid collisions.

C. Trajectory Prediction Example

A sample is illustrated in Fig. 3 to present the prediction results of the proposed method. The solid lines represent the 4 seconds past trajectories, the dashed lines are the stochastic predictions for the future 2 seconds, and the marked lines are the ground truth for the same period. Although the prediction errors vary among the six vehicles presented in this sample, the average ADE of these vehicles is 0.40m with a standard deviation of 0.38m. Meanwhile, the average FDE and its standard deviations are 1.02m and 0.66m, respectively. As the average travelling speed of these six vehicles is about 104.4 km/h, the prediction accuracy is very promising in this context.

D. Performance Evaluation

Benchmark Models. With the sample result presented, a benchmark analysis with some state-of-the-art approaches was conducted to evaluate the performance of the proposed model. It should be noted that some of the selected methods were commonly used in assessing the performance of newly proposed trajectory prediction approaches, and the others represent the key milestones and more recent techniques in this field of research. Therefore, these baseline models were selected to evaluate the performance of the model proposed in this study.

- V-LSTM: Vanilla LSTM is one of the classical methods for time-series prediction. Its application in trajectory prediction uses a single LSTM to encode the motion history of the ego vehicle without considering its spatial interactions with surrounding vehicles.

- S-LSTM [40]: Social LSTM is initially developed for pedestrian trajectory prediction. Each pedestrian is modelled using an LSTM. The hidden states of pedestrians within a specific area are pooled using fully connected social pooling to preserve spatial interactions.
- CS-LSTM [12]: Convolutional Social LSTM is designed for vehicle trajectory prediction. It uses convolutional layers with social pooling and generates a multimodal prediction based on six manoeuvres.
- S-GAN [37]: Social GAN models each pedestrian motion with an LSTM, and the hidden states of pedestrians are pooled using global pooling. GAN is used to generate multimodal prediction results.
- STGAT [41]: Originally developed for pedestrian trajectory prediction. It is a seq2seq model which uses one LSTM for each pedestrian's motion state. Meanwhile, the spatial interactions are extracted with Graph Attention Network (GAT), and an extra LSTM models the temporal correlations between interactions.
- GRIP++ [22]: GRIP++ ranked first in the 2019 ApolloScape trajectory competition and achieved top accuracy with the NGSIM dataset, as listed in Table 2. It uses fixed and dynamic graphs for spatial relations and a two-layer GRU for trajectory prediction, applicable to various traffic agents like vehicles, pedestrians, and cyclists.
- Social-STGCNN [42]: Originally developed for pedestrian trajectory prediction, Social-STGCNN uses Spatio-Temporal Graph Convolutional Neural Networks (ST-GCNN) to extract spatial and temporal relations from spatio-temporal graphs. It introduces a weighted adjacency matrix to model social influence between pedestrians, achieving notable improvements in prediction accuracy and speed.
- STDAN [31]: STDAN introduces a novel spatial-temporal dynamic attention network for vehicle trajectory prediction. It incorporates a driving intention-specific feature fusion mechanism, allowing the adaptive integration of temporal and social features for maneuver-based, multi-modal trajectory prediction.

Since these baseline models were initially proposed for different purposes and were trained with various datasets, they were retrained using the same dataset in this study to facilitate the evaluation on a unified basis. Same training parameters were adopted, and the comparison was implemented with an observation horizon of 4s and a prediction horizon of 2s. The comparison mainly focuses on the accuracy of the prediction results and inference speed.

Prediction Accuracy Performance. The comparison results between the proposed method and the other eight models were summarised in Table 4. For the highD dataset, the proposed model outperforms all others, achieving the best accuracy with an ADE of 0.43m and an FDE of 0.79m. In comparison, STDAN, which previously represented the state-of-the-art, achieves an ADE of 0.47m and an FDE of 0.81m. Despite STDAN's strong performance, the proposed model achieves an additional improvement of 8.51% in ADE and 2.47% in FDE. This improvement underscores the superior prediction accuracy of the proposed method. Among the existing models, V-LSTM exhibits the most significant prediction errors, likely due to its reliance solely on motion history and lack of spatial interaction considerations. This highlights the importance of incorporating spatial features into vehicle trajectory prediction to account for the interactions between vehicles, which can lead to more accurate results.

In the inD dataset, CS-LSTM achieves the best ADE of 0.66m, indicating its strong performance in average displacement prediction. However, Social-STGCNN records the lowest FDE of 1.28m, outperforming other models in terms of final displacement accuracy. The proposed model performs well, with an ADE of 1.07m and an FDE of 1.29m, placing it among the top models for final displacement but slightly behind in average displacement.

Social-STGCNN's performance, while notable, requires further discussion. Although it ranks third in prediction accuracy, this model was initially designed to predict pedestrian trajectories, where it has achieved state-of-the-art performance. The fundamental principles and scales involved in pedestrian trajectory prediction differ significantly from those required for vehicle trajectory prediction, potentially limiting Social-STGCNN's ability to generalise to vehicle scenarios. In this study, the model was used as a benchmark, and only limited adaptations were made to ensure vehicle data were appropriately scaled and transmitted. No further modifications were applied to the model's core specifications. Thus, with more targeted adjustments and calibrations, its accuracy in vehicle trajectory prediction could potentially be improved.

A general trend observed when comparing the two datasets is the degradation in performance for most models when tested on the inD dataset. This is despite the fact that vehicles in the inD dataset generally move slower. The likely explanation lies in the more complex interactions among vehicles in the intersection scenarios found in the inD dataset. Social pooling and graph-based models, which are designed to capture pairwise interactions, struggle to extract higher-order relationships in these more complex environments. This may explain why CS-LSTM achieves better performance in the inD dataset, as its convolutional social pooling mechanism aggregates information from multiple neighboring vehicles simultaneously and captures the collective influence of all vehicles within a specific spatial area. The ability to handle these high-order interactions more effectively allows CS-LSTM to perform better in scenarios involving complex vehicle interactions, such as intersections.

Inference Speed Performance. As shown in Table 4, the advantage of extracting spatial relations with Temporal Convolutional layers over LSTM is revealed. All three

TABLE 4. Prediction Performance Comparison with highD and inD (RMSE in Metre)

Model	highD		inD		Inference speed (s)
	ADE (m)	FDE (m)	ADE (m)	FDE (m)	
V-LSTM	1.68	2.23	4.75	4.86	0.102
S-LSTM	1.01	1.87	1.67	3.60	0.204
CS-LSTM	0.96	1.84	**0.66**	1.35	0.641
S-GAN	1.27	1.93	2.22	4.25	0.955
STGAT	0.69	1.34	1.14	2.08	0.943
GRIP++	0.51	0.91	0.72	1.65	0.022
Social-STGCNN	0.94	1.57	1.24	**1.28**	**0.013**
STDAN	0.47	0.81	0.81	1.51	0.016
Proposed model	**0.43**	**0.79**	1.07	1.29	0.023

models (GRIP++, Social-STGCNN and the proposed model) obtain significant improvements in inference speed than LSTM-based models. Social-STGCNN achieves the best performance among all models in comparison. It is over 73 times faster than the slowest method (S-GAN) and about 1.23 times faster than the STDAN, the fastest LSTM-based model. As the proposed model needs to construct the risk graph to determine the spatial relations at each timestep, it has not achieved the best performance in inference speed. However, it still obtains a notable improvement than most existing models and is about four times faster than V-LSTM and over 41 times faster than S-GAN.

Distributional Performance. As ADE and FDE can not compare the distributions produced by generative models, NLL is adopted to evaluate the variance and multimodality, without assumptions about the output's distribution [38]. The state-of-the-art stochastic model (STGAT) is used for evaluation. The proposed model was evaluated over time to investigate the performance changes along the prediction horizon. Results of 1000 sampled trajectories are shown in Fig. 4, and error bars are bootstrapped at 95% confidence intervals. As the average NLL of the proposed model is smaller than STGAT at every prediction step, it indicates our model's consistent multimodal modelling capacity.

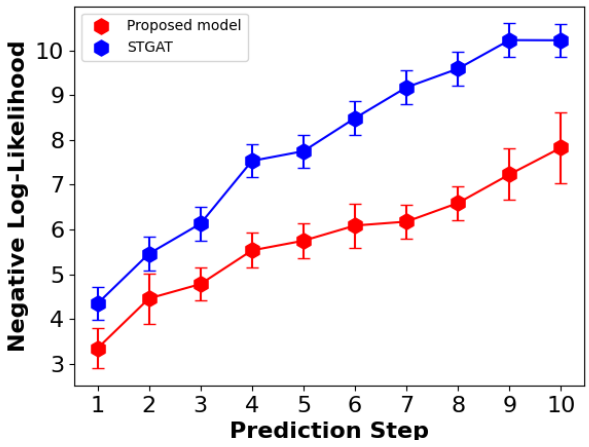

FIGURE 4. Mean NLL across prediction horizon

E. Ablation Study of Risk-awareness Kernel Function

Since the kernel function for computing the adjacency matrix represents the social influence between vehicles, it would be beneficial to evaluate if the proposed approach can effectively capture the essence in spatial relations while remaining computationally efficient. Thus, another two commonly used kernel functions were adopted to benchmark the performance of the proposed approach using the highD dataset.

The first kernel function is to treat all vehicles within the neighbourhood area equally. The spatial relation captured with this kernel function is similar to some reviewed studies [3], [15]–[17]. This kernel function can be represented as,

$$e_t^{ij} = \begin{cases} 1, & if \parallel s_t^i - s_t^j \parallel_2 < threshold \\ 0, & otherwise \end{cases} \tag{14}$$

Meanwhile, the second kernel function is based on the relative distance between two vehicles to model their social impacts [42]. It can be easily interpretable, despite the vehicles' dynamic relations, spatially closer vehicles tend to have a more significant influence on each other. It also correlates that human drivers tend to pay more attention to vehicles in closer proximity. This kernel function can be denoted as,

$$e_t^{ij} = 1 - \frac{\parallel s_t^i - s_t^j \parallel_2}{max\_length} \tag{15}$$

where max_length is the length of the road segment.

Both kernel functions were implemented for comparison. As listed in Table 5, the first kernel function has the worst prediction accuracy. This is because this kernel function only considers vehicles within the neighbourhood and does not differentiate the levels of their social impacts. The neighbourhood area selected with fixed thresholds can lose the essential information that determines the driver's motivations. Moreover, assuming vehicles within the neighbourhood have the same social impacts on the ego vehicle can also lead to severe issues. Different vehicle positions and dynamic states can pose distinct levels of danger and stress on the ego driver.

Meanwhile, the second kernel function performs better than the first, improving 43.3% in ADE and 21.0% in FDE. While it confirms the benefits of differentiating the social impacts between vehicles, solely depending on distance is insufficient to determine the spatial relations. This is because surrounding vehicles at the same distance can pose distinct social impacts on the ego vehicle. Thus, although the second kernel function can differentiate the social relations between most vehicles, they can lead to a false adjacency matrix in some cases. The proposed risk-awareness kernel function compensates for these drawbacks and improves 42.8% in ADE and 51.2% in FDE.

TABLE 5. Ablation Study (RMSE in Metre)

Model	ADE (m)	FDE (m)	Inference time (s)
Equation 12	1.36	2.05	0.016
Equation 13	0.77	1.62	0.012
Proposed model	**0.43**	**0.79**	**0.023**

From the perspective of inference speed, while all three kernel functions achieve fast inference speed, the proposed kernel function is the slowest among them. It is mainly caused by additional variables and calculations required to construct the graph representation of the scene at each time step.

With the comparison of prediction accuracy and inference speed, it can be noted that the proposed risk-awareness kernel function can better extract the essential spatial relations between vehicles and yield improved performance in predicting the future trajectory of vehicles. Although its inference speed is slower than the other two kernel functions, it is still much faster than LSTM-based models and should be sufficient for most applications.

IV. Conclusion

This paper presented a novel approach for vehicle trajectory prediction by modelling the spatial-temporal interactions among vehicles. The proposed model was trained and evaluated using the publicly available Highway Drone Dataset and Intersection Drone Dataset, demonstrating promising improvements in both prediction accuracy and inference speed compared to eight existing methods. An ablation study further validated the effectiveness of the risk-awareness kernel function, highlighting its contribution to enhancing the model's explainability.

The primary contribution of this research lies in the development of a stochastic vehicle trajectory prediction method that models vehicular spatial-temporal interactions. A novel risk-awareness kernel function was introduced to construct a weighted adjacency matrix, effectively capturing the spatial relationships between vehicles. The model employs GCNs to extract spatial features and TCNs to model temporal dependencies. The combination of these features in a decoder produces a stochastic, multimodal prediction of future vehicle trajectories.

Despite the promising results, the map-free nature of the proposed approach presents limitations, particularly in more complex road geometries, such as urban environments. The use of relatively simple road structures in the datasets may have contributed to the observed high accuracy, with the model's performance deteriorating when applied to the inD dataset, which features more intricate road layouts. Future work will focus on evaluating the model with data from more complex road segments to further test its robustness. Moreover, to address the limitations observed in complex datasets such as inD, future research will focus on integrating hypergraph representation learning into the current model. By leveraging hyperedges to capture higher-order interactions among road agents, this approach will enable the model to extract and represent spatial relations and interaction dynamics more effectively. Such advancements are expected to significantly improve the model's ability to handle scenarios with intricate traffic dynamics, enhancing risk-awareness and prediction accuracy. Additionally, the potential trade-offs between incorporating map information and maintaining model scalability will be explored to enhance the model's applicability in diverse traffic scenarios.

REFERENCES

[1] S. Herrero, J. Febres, W. Boulagouas, J. Gutiérrez, and M. Saldaña, "Assessment of the influence of technology-based distracted driving on drivers' infractions and their subsequent impact on traffic accidents severity," *International Journal of Environmental Research and Public Health*, vol. 18, p. 7155, 2021.

[2] M. Morando, Q. Tian, L. Truong, and H. Vu, "Studying the safety impact of autonomous vehicles using simulation-based surrogate safety measures," *Journal of advanced transportation*, vol. 2018, 2018.

[3] F. Altché and A. de La Fortelle, "An lstm network for highway trajectory prediction," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 353–359.

[4] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Attention based vehicle trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 175–185, 2021.

[5] Z. Wang, J. Zhang, J. Chen, and H. Zhang, "Spatio-temporal context graph transformer design for map-free multi-agent trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 1369–1381, 2024.

[6] J. Lian, S. Li, D. Yang, J. Zhang, and L. Li, "Encoding the intrinsic interaction information for vehicle trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 2600–2611, 2024.

[7] J. Schmidt, J. Jordan, F. Gritschneder, and K. Dietmayer, "Cratpred: Vehicle trajectory prediction with crystal graph convolutional neural networks and multi-head self-attention," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 7799–7805.

[8] Y. Feng, S. Pickering, E. Chappell, P. Iravani, and C. Brace, "Driving style analysis by classifying real-world data with support vector clustering," in *2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, 2018, pp. 264–268.

[9] Y. Feng, P. Iravani, and C. Brace, "A fuzzy logic-based approach for humanized driver modelling," *Journal of Advanced Transportation*, vol. 2021, pp. 1–13, 2021.

[10] Y. Feng, S. Pickering, E. Chappell, P. Iravani, and C. Brace, "Driving style modelling with adaptive neuro-fuzzy inference system and real driving data," in *Advances in Human Aspects of Transportation*, N. Stanton, Ed., 2019, pp. 481–490.

[11] S. Ammoun and F. Nashashibi, "Real time trajectory prediction for collision risk estimation between vehicles," in *2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing*, 2009, pp. 417–422.

[12] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *2018 IEEE/CVF Conference on Computer*

Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 1549–15 498.

[13] A. Eidehall and L. Petersson, "Statistical threat assessment for general road scenes using monte carlo sampling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 137–147, 2008.

[14] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1450–1464, 2006.

[15] H. Woo, M. Sugimoto, J. Wu, Y. Tamura, A. Yamashita, and H. Asama, "Trajectory prediction of surrounding vehicles using lstm network," 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:196589931.

[16] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based lstms," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1179–1184.

[17] B. Kim, C. M. Kang, S. H. Lee, H. Chae, J. Kim, C. C. Chung, and J. W. Choi, "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network," 2017. [Online]. Available: https://arxiv.org/abs/1704.07049.

[18] T. Zhao, Y. Xu, M. Monfort, W. Choi, C. Baker, Y. Zhao, Y. Wang, and Y. N. Wu, "Multi-agent tensor fusion for contextual trajectory prediction," 2019. [Online]. Available: https://arxiv.org/abs/1904.04776.

[19] H. He, H. Dai, and N. Wang, "UST: Unifying spatio-temporal context for trajectory prediction in autonomous driving," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5962–5969.

[20] Z. Zhao, H. Fang, Z. Jin, and Q. Qiu, "Gisnet:graph-based information sharing network for vehicle trajectory prediction," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–7.

[21] H. Jeon, J. Choi, and D. Kum, "Scale-net: Scalable vehicle trajectory prediction network under random number of interacting vehicles via edge-enhanced graph convolutional neural network," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 2095–2102.

[22] X. Li, X. Ying, and M. C. Chuah, "Grip++: Enhanced graph-based interaction-aware trajectory prediction for autonomous driving," 2019. [Online]. Available: https://arxiv.org/abs/1907.07792.

[23] X. Li, X. Ying, and M. C. Chuah, "Grip: Graph-based interaction-aware trajectory prediction," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 3960–3966.

[24] B. Dong, H. Liu, Y. Bai, J. Lin, Z. Xu, X. Xu, and Q. Kong, "Multi-modal trajectory prediction for autonomous driving with semantic map and dynamic graph attention network," 2021. [Online]. Available: https://arxiv.org/abs/2103.16273.

[25] Y. Zhang, Y. Zou, J. Tang, and J. Liang, "Long-term prediction for high-resolution lane-changing data using temporal convolution network," *Transportmetrica B: Transport Dynamics*, vol. 10, no. 1, pp. 849–863, 2021. [Online]. Available: https://doi.org/10.1080%2F21680566.2021.1950072.

[26] J. Strohbeck, V. Belagiannis, J. Müller, M. Schreiber, M. Herrmann, D. Wolf, and M. Buchholz, "Multiple trajectory prediction with deep temporal and spatial convolutional neural networks," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 1992–1998.

[27] Q. Li, B. Ou, Y. Liang, Y. Wang, X. Yang, and L. Li, "Tcn-sa: A social attention network based on temporal convolutional network for vehicle trajectory prediction," *Journal of Advanced Transportation*, vol. 2023, pp. 1–12, 2023.

[28] X. Chen, H. Zhang, Y. Hu, J. Liang, and H. Wang, "Vnagt: Variational non-autoregressive graph transformer network for multi-agent trajectory prediction," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 10, pp. 12 540–12 552, 2023.

[29] H. Liao, Z. Li, H. Shen, W. Zeng, D. Liao, G. Li, and C. Xu, "BAT: Behavior-aware human-like trajectory prediction for autonomous driving," in *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, vol. 38, no. 9, 2024, pp. 10 332–10 340.

[30] H. Liao, Z. Li, C. Wang, H. Shen, B. Wang, D. Liao, G. Li, and C. Xu, 2024.

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

[32] H. Liao, Y. Li, Z. Li, C. Wang, Z. Cui, S. E. Li, and C. Xu, "A cognitive-based trajectory prediction approach for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 4, pp. 4632–4643, 2024.

[33] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018. [Online]. Available: https://arxiv.org/abs/1803.01271.

[34] E. Candela, Y. Feng, P. Angeloudis, and Y. Demiris, "Quantitative risk indices for autonomous vehicle training systems," 2021. [Online]. Available: https://arxiv.org/abs/2104.12945.

[35] H. Han, S. Kim, J. Choi, H. Park, J. H. Yang, and J. Kim, "Driver's avoidance characteristics to hazardous situations: A driving simulator study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 81, pp. 522–539, 2021.

[36] P. Bokare and A. Maurya, "Acceleration-deceleration behaviour of various vehicle types," *Transportation Research Procedia*, vol. 25, pp. 4733–4749, 2017.

[37] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264.

[38] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2375–2384.

[39] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 1929–1934.

[40] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 961–971.

[41] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "Stgat: Modeling spatial-temporal interactions for human trajectory prediction," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6271–6280.

[42] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 14 412–14 420.

V. Biography Section

Yuxiang Feng is a Research Associate and Lab Manager at the Transport Systems and Logistics Laboratory (TSL), Imperial College London. He received a BEng in Mechanical Engineering from Tongji University and an MSc in Mechatronics and PhD in Automotive Engineering from the University of Bath. His main research interests include environment perception, sensor fusion and artificial intelligence for robotics and autonomous vehicles.

Qiming Ye is a Postdoctoral Research Fellow at the Future Cities Laboratory Global of ETH Zurich. He received PhD in transportation engineering from Imperial College London, and BEng and MEng Degrees from Tongji University. His research focuses on developing AI-based tools to for intelligent allocation of road space usage in the era of autonomous vehicles transport, and adaptive planning of charging infrastructure for electric vehicles.

Eduardo Candela was a PhD student at the Transport Systems and Logistics Laboratory and Personal Robotics Laboratory at Imperial College London. He holds a MSc in Operations Research from the Massachusetts Institute of Technology, BSc in Industrial Engineering and Mechatronics from Instituto Tecnologico Autonomo de Mexico. His research focuses on safety modelling of vehicle systems, and strategic decision making for Autonomous Vehicles using reinforcement learning and optimisation.

Jose Javier Escribano Macias is a Lecturer at the Centre of Transport Engineering and Modelling, Imperial College London. He joined CTS in October 2015 as part of the EPSRC Centre for Doctoral Training (CDT) in Sustainable Civil Engineering and was awarded his Ph.D. in March 2021. His research focuses on collaborative vehicle control, optimisation of last-mile logistics, urban air mobility, and machine learning and game theoretical models.

Bo Hu was born in Hefei, China, in 1989. He received the B.S. degree in automotive engineering from Chongqing University of Technology (CQUT), Chongqing, China, in 2011, and the M.S. and Ph.D. degrees in automotive engineering from the University of Bath, Bath, U.K., in 2012 and 2016, respectively.,He is currently an Associate Professor with the Key Laboratory of Advanced Manufacturing Technology for Automobile Parts, Ministry of Education, CQUT. His current research interests include machine learning based control of intelligent and connected vehicles and modeling and control of advanced boosted engine systems.

Yiannis Demiris (SM'03) received the B.Sc. (Hons.) degree in artificial intelligence and computer science and the Ph.D. degree in intelligent robotics from the Department of Artificial Intelligence, University of Edinburgh, Edinburgh, U.K., in 1994 and 1999, respectively. He is a Professor with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., where he is the Royal Academy of Engineering Chair in Emerging Technologies, and the Head of the Personal Robotics Laboratory. His current research interests include human-robot interaction, machine learning, user modeling, and assistive robotics. Prof. Demiris is a Fellow of the Institution of Engineering and Technology (IET), and the British Computer Society (BCS).

Panagiotis Angeloudis is Professor of Transport Systems & Logistics and Head of the Transport Systems and Logistics Laboratory (TSL), based in the Centre for Transport Engineering and Modelling (CTEM) at Imperial College London. Before establishing TSL, Panagiotis held a JSPS Research Fellowship at Kyoto University. He previously obtained a PhD in Transportation at Imperial College London and spent periods as a research analyst at DP World and the United Nations in Geneva. His research focuses on the study of networks, optimisation methods and multi-agent systems, as well as their applications in autonomous transport systems, urban infrastructure and logistics.