

Social Force Embedded Mixed Graph Convolutional Network for Multi-class Trajectory Prediction

Quancheng Du, Xiao Wang, *Senior Member, IEEE*, Shouguo Yin,
Lingxi Li, *Senior Member, IEEE*, and Huansheng Ning, *Senior Member, IEEE*

Abstract—Accurate prediction of agent motion trajectories is crucial for autonomous driving, contributing to the reduction of collision risks in human-vehicle interactions and ensuring ample response time for other traffic participants. Current research predominantly focuses on traditional deep learning methods, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These methods leverage relative distances to forecast the motion trajectories of a single class of agents. However, in complex traffic scenarios, the motion patterns of various types of traffic participants exhibit inherent randomness and uncertainty. Relying solely on relative distances may not adequately capture the nuanced interaction patterns between different classes of road users. In this paper, we propose a novel multi-class trajectory prediction method named the social force embedded mixed graph convolutional network (SFEM-GCN). The primary goal is to extract social interactions among agents more accurately. SFEM-GCN comprises three graph topologies: the semantic graph (SG), position graph (PG), and velocity graph (VG). These graphs encode various of social force relationships among different classes of agents in complex scenes. Specifically, SG utilizes one-hot encoding of agent-class information to guide the construction of graph adjacency matrices based on semantic information. PG and VG create adjacency matrices to capture motion interaction relationships between different classes agents. These graph structures are then integrated into a mixed graph, where learning is conducted using a spatio-temporal graph convolutional neural network (ST-GCNN). To further enhance prediction performance, we adopt temporal convolutional networks (TCNs) to generate the predicted trajectory with fewer parameters. Experimental results on publicly available datasets demonstrate that SFEM-GCN surpasses state-of-the-art methods in terms of accuracy and robustness.

Index Terms—Graph convolutional network, trajectory prediction, multi-class, social interactions, autonomous driving.

I. INTRODUCTION

The development of autonomous driving technology is expected to bring about profound changes in intelligent transportation systems [1]–[3]. It has the potential to improve road safety, reduce traffic congestion, and provide more convenient modes of transportation [4]. Within the domain of autonomous driving technology, trajectory prediction assumes a pivotal role. It enables autonomous vehicles to navigate intricate traffic environments by analyzing and anticipating the behaviors of surrounding traffic participants. To achieve safe and agile autonomous driving, references [5], [6] outlined requirements for trajectory prediction in terms of speed, accuracy, and generalization. However, when dealing with diverse classes of road users, such as pedestrians, cars, and cyclists, the prediction of the trajectory for autonomous driving still faces challenges due to the complexity of the scenarios and the variability of interactive behaviors [7].

In early research, pooling mechanisms grounded in recurrent neural networks (RNNs) exhibited promising performance in modeling social interactions. S-LSTM [8] is a classic application of RNN to predict the trajectory of pedestrians, which captures the movements of pedestrians using long short-term memory (LSTM) and incorporates social pooling mechanisms to model interactions among pedestrians. This approach not only surpasses traditional rule-based trajectory prediction in terms of accuracy but also demonstrates commendable robustness in complex scenarios. To further advance trajectory prediction research, SGAN [9] introduced generative adversarial networks (GANs) on top of S-LSTM to address the multimodal distribution of pedestrian motion and demonstrated advantages in robustness and accuracy. Subsequent research has incorporated attention mechanisms to more effectively model the impact of various agents on the target [10]. SoPhie [11] added social and physical attention modules to assign different weights to surrounding pedestrians and scenes. Social-BiGAT [12] introduced a graph-based GAN that models social interactions among all pedestrians in a static scene using graph attention networks. It generates socially compatible trajectories by establishing a reversible mapping between predicted trajectories and the behavioral characteristics of the target.

In recent years, graph convolution networks (GCNs) have garnered significant attention for accurately representing topological relationships among agents and achieving competitive results [13]–[15]. SSTGCNN [16] employed a GCN-based approach that utilizes relative distances to represent interac-

This work was supported by the National Natural Science Foundation of China under Grant 62173329. (*Corresponding author: Xiao Wang.*)

Quancheng Du is with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China. (E-mail: quancheng.du@xs.ustb.edu.cn)

Xiao Wang is with the School of Artificial Intelligence, Anhui University, Hefei 230031, China, and also with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. (E-mail: xiao.wang@ahu.edu.cn)

Shouguo Yin is with the School of Artificial Intelligence, Anhui University, Hefei 230031, China. (E-mail: 2638711962@qq.com)

Lingxi Li is with the School of Electrical and Computer Engineering, Indiana University-Purdue University Indianapolis, Indianapolis IN46202, USA. (E-mail: ll7@iupui.edu)

Huansheng Ning is with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China. (E-mail: ninghuansheng@ustb.edu.cn)

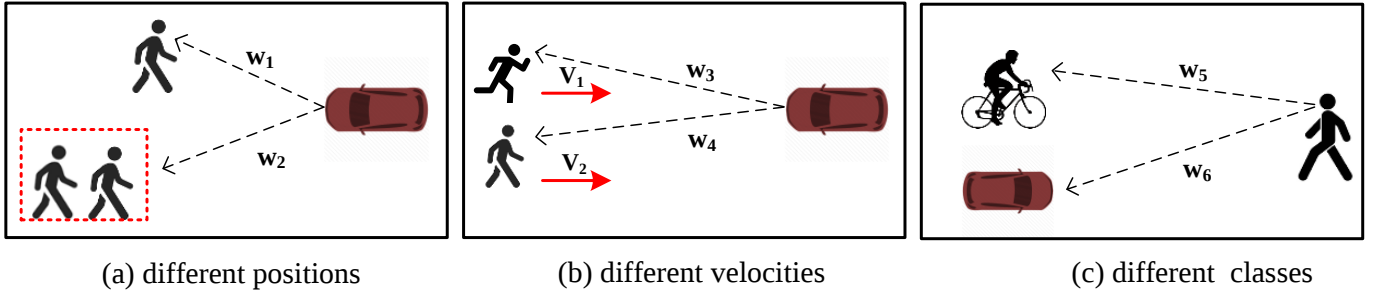


Fig. 1: Three different factors that need to be considered in the scenario, with w representing attention weights.

tions between pedestrians. This method combines GCNs with temporal convolutional networks (TCNs) to effectively handle temporal dependencies. SGCN [17] modeled the interaction between pedestrians as a directed graph, capturing not only the interactions that significantly affect the target but also the directional information between pedestrians. While GCN-based methods have demonstrated some effectiveness, real-world traffic scenarios are exceptionally complex, posing a challenge to accurately model social interactions among agents using simple graph structures [18]. Drawing inspiration from the social force relationship, it is understood that the social factors influencing agents' future trajectories encompass multiple aspects, including relative position, velocity, and agent class [19]. For example, in Figure 1(a), it depicts the interaction between agents at different positions. Since closer proximity is more likely to attract attention, a higher weight should be assigned to this interaction, i.e., $w_1 > w_2$. In Figure 1(b), assuming $V_1 > V_2$, it illustrates the interaction between agents with different velocities. Considering that higher velocity has a greater impact on the interaction, a larger weight should be assigned to the agent with a higher velocity, i.e., $w_3 > w_4$. In Figure 1(c), it represents the interaction between agents with different classes. Compared with a cyclist, a car is more likely to capture human attention, thus requiring a higher weight to be assigned to the car, i.e., $w_6 > w_5$. Some methods have overlooked the aforementioned factors and primarily focused on relative distance as the basis for modeling interactions, such as through position pooling or calculating attention weights based on distance [8], [16]. Additionally, these methods have predominantly addressed the prediction of trajectories for single class, lacking relevant research on predicting interactions among multi-class agents.

A recent study, Semantics-STGCNN [20], introduced a novel approach that incorporates agent label features into the adjacency matrix of the velocity representation. This integration effectively incorporates class information, leading to state-of-the-art performance in multi-class trajectory prediction. However, this method overlooks the influence of position information on trajectory prediction, which can significantly impact accuracy in real-world traffic scenarios [21]. In this paper, inspired by [16], [20], we present a novel approach called social force embedded mixed graph convolutional network (SFEM-GCN) to model interaction relationships among different classes agents. SFEM-GCN incorporates three types of graph topologies: semantic graph (SG), position graph (PG),

and velocity graph (VG). The goal is to achieve effective trajectory prediction by comprehensively considering the social factors that influence the movement of agents. Specifically, semantic information, position information, and velocity information of agents in the scene are encoded and processed. Social force relationships in the scene are then constructed through mixed multiple graph-guided adjacency matrices. We adopt a multi-layer spatial temporal graph convolutional neural network (ST-GCNN) with residual structures is employed to aggregate node features on the graph, and TCNs is used to predict the trajectories of multi-class agents in real-world scenarios. In summary, the main contributions of this paper are as follows.

- 1) We propose a multi-class trajectory prediction method based on social force embedded mixed graph convolutional network (SFEM-GCN) that encodes social force relationship about different social behaviors. We employ three graph topologies to delineate social interactions among agents of different classes. In contrast to existing methods, our proposed mixed graph structure captures raw information between agents and represents various social interaction features, offering a more intuitive and effective model for social interactions.
- 2) By increasing the number of convolutional layers in our model, we have successfully improved trajectory prediction within the SSTGCNN [16] architecture. This augmentation has led to noteworthy improvements in the overall performance of the model.
- 3) Our method is evaluated on a widely used trajectory prediction dataset SDD, showing that SFEM-GCN outperforms existing methods. In particular, compared to state-of-the-art multi-class trajectory prediction methods with a GCN model proposed recently [20], our model achieves approximately 3% improvement in the Minimum Average Displacement Error (mADE) and 4% improvement in the Minimum Final Displacement Error (mFDE). Furthermore, in the latest metrics, Average² Displacement Error (aADE) and Average Final Displacement Error (aFDE), our method shows approximately 8% and 13% performance improvement, respectively.

The remainder of the paper is organized into the following sections: Section II introduces a brief overview of the related work. Section III describes the problem statement and the details of our proposed method. Section IV presents the

experimental setup and results. Finally, Section V concludes the paper.

II. RELATED WORK

A. Modeling Social Interactions for Trajectory Prediction

Trajectory prediction seeks to forecast the future motion paths of agents based on provided historical trajectory data, typically encompassing a brief interval of 3 to 5 seconds [22]–[24]. Traditional trajectory prediction has primarily made use of model-driven methods, employing complex mathematical and statistical models to characterize the motion patterns of agents [25], [26]. For example, in early studies, Helbing *et al.* [27] pioneered the concept of social force, introducing an energy potential field (repulsive and attractive forces) to model social interactions among pedestrians. Kooij *et al.* [28] integrated Bayesian filters and kinematic models to formulate a context-sensitive dynamic network for predicting the trajectories of agents. Furthermore, Gaussian process [29] and Markov decision process [30] have been employed to model social interactions in trajectory prediction. In recent years, with the continuous advancement of deep learning techniques, data-driven artificial intelligence methods have emerged as a popular research direction for trajectory prediction [5], [31], [32]. The S-LSTM model, introduced by Alahi *et al.* [25], stands as a classic application of RNN in predicting pedestrian trajectories. Leveraging LSTM, the model adeptly captures the intricate motion patterns of pedestrians while integrating a social pooling mechanism to model interaction relationships. SGAN [9] employed GAN to predict multimodal trajectories and introduced a novel pooling mechanism to compute social interactions based on relative distances between pedestrians. Beyond the previously mentioned networks, GCNs have also emerged as a popular approach for accurately predicting trajectories of agents within various scenes. SSTGCNN [16] embedded social interactions among pedestrians in an adjacency matrix, modeling them as a graph network structure. It utilized kernel functions to process the adjacency matrix and captured spatial and temporal interaction information. STGAT [33] learned the influence weights between pedestrians through graph attention networks, resulting in outstanding pedestrian trajectory prediction.

Although previous methods have successfully modeled social interactions among agents and achieved satisfactory prediction results, they often do not fully consider the various social factors that influence agents. For instance, the influence of agents with different positions, velocities, and classes is crucial for trajectory prediction [34]. Furthermore, existing methods have primarily focused on trajectory prediction for single class agents, overlooking the influence of interaction dynamics among different classes. In this paper, we propose a mixed graph convolutional network that comprehensively considers the influence of the factors mentioned above, with the goal of achieving more accurate multi-class trajectory prediction. To capture complex social interactions, we construct a multi-type graph-guided adjacency matrix that represents the interaction relationships within the scene. This approach enables precise calculation of interaction degrees between agent nodes within

the graph structure, resulting in a comprehensive representation of social force relations among agents and enhanced accuracy in capturing social interaction features.

B. Graph Convolutional Networks in Trajectory Prediction

GCN is a neural network capable of handling graph-structured data, extending the concept of convolutional neural networks (CNNs) to graphs [35]. GCN applies discrete convolutions to weigh and aggregate node and edge features, assigning different weights based on neighboring nodes to effectively capture the interaction relationships [36]. Due to its powerful representation capabilities in non-Euclidean spaces, GCN has been widely applied in trajectory prediction tasks in recent years [16], [37]. For example, Huang *et al.* [33] introduced the STGAT model, which treats each pedestrian as a node on the graph and shares information among different pedestrians. Mohamed *et al.* [16] proposed a social spatial-temporal graph convolutional neural network (SSTGCNN). This model embeds social interaction relationships between pedestrians into the adjacency matrix, treating them as a graph structure. It utilizes kernel functions to process the adjacency matrix, capturing both spatial and temporal interactions. Considering the directional nature of pedestrian interactions, Shi *et al.* [17] introduced the sparse graph convolution network (SGCN) for pedestrian trajectory prediction, with the aim of addressing the modeling redundancy issue in existing approaches when dealing with dense undirected interactions among pedestrians. However, it's worth noting that the use of directed graphs can lead to computationally complex models that consume more computational resources [38]. In addition, the emergence of graph attention networks (GATs) [39] has yielded promising predictive performance. STGAT [33] enhances the interaction between pedestrians by introducing a flexible spatio-temporal graph attention mechanism. Social-BiGAT [12] learns reliable feature representations for agent interactions using GAT, resulting in accurate multimodal trajectory prediction. GAT enhances trajectory prediction by processing encoded trajectory information and provides the model with interpretability.

Despite the significant progress of GCN-based methods in the field of trajectory prediction, they exhibit limitations in practical applications. For instance, they demand a substantial amount of data to discern differences between nodes in the graph and consume more computational resources [18]. In light of this, drawing inspiration from reference [20], we propose a mixed graph convolutional network based on social force relation to address these limitations. In contrast to existing GCN methods, our approach emphasizes manual design by incorporating social force relationships and domain expertise to guide the model construction process. Consequently, our method requires fewer computational parameters and achieves faster execution speed. Additionally, by considering more social factors, our method can better differentiate the influences between different nodes, leading to more accurate prediction.

III. METHODOLOGY

A. Problem Statement

The trajectory prediction can be fundamentally transformed into a decision problem of time series [40]. Specifically, given

observation of M agents in the scene over a fixed duration in the past, their historical trajectories are denoted as $Y^i = \{Y^1, Y^2, \dots, Y^M\}$ where $i \in \{1, 2, \dots, M\}$. The observed trajectory of the i -th agent at time step t can be expressed as

$$Y_t^i = \{(x_t^i, y_t^i) \in \mathcal{R}^2 | t = 1, 2, \dots, T_{\text{obs}}\} \quad (1)$$

where (x_t^i, y_t^i) denotes the true position coordinates. Over observation time-steps $1 \leq t \leq T_{\text{obs}}$, we aim to predict the likely trajectories for the i -th agent in the future time-steps $T_{\text{obs}} + 1 \leq t \leq T_{\text{pred}}$, denoted as

$$\hat{Y}_t^i = \{(\hat{x}_t^i, \hat{y}_t^i) \in \mathcal{R}^2 | t = T_{\text{obs}} + 1, \dots, T_{\text{pred}}\} \quad (2)$$

Similarly to [8], [16], our method employs a bivariate Gaussian distribution to characterize the coordinates (x_t^i, y_t^i) of agent trajectories as random variables within the probability distribution of agent positions. We assume that the predicted agent trajectory coordinates $\hat{Y}_t^i = (\hat{x}_t^i, \hat{y}_t^i)$ also follow a bivariate Gaussian distribution. The model parameters are learned by minimizing the negative log-likelihood loss, as shown in the following:

$$\hat{Y}_t^i \sim N(\tilde{\mu}_t^n, \tilde{\sigma}_t^n, \tilde{\rho}_t^n) \quad (3)$$

$$L(W) = - \sum_{t=T_{\text{obs}}+1}^{T_{\text{pred}}} \sum_{i=1}^N \log \left(\hat{Y}_t^i \sim N(\tilde{\mu}_t^n, \tilde{\sigma}_t^n, \tilde{\rho}_t^n) \right) \quad (4)$$

where W represents the learned network parameters. Our goal is to minimize the loss value, allowing us to obtain optimal weights for the network. $\tilde{\mu}_t^n$ represents the mean, $\tilde{\sigma}_t^n$ represents the variances, and $\tilde{\rho}_t^n$ represents the correlation of the bivariate Gaussian distribution.

B. Overall Model

This section present SFEM-GCN, a novel trajectory prediction model that constructs a mixed graph convolutional neural network by effectively embedding the social force relationships among multi-class agent. The proposed method is depicted in Figure 2. The model takes both the semantic information of the scene and the historical trajectory information of the agents, which includes their position coordinates, as input. The coordinate information is then transformed into a graph representation, where each agent is depicted as a node in the scenario. Furthermore, the semantic information of the scene is encoded using one-hot encoding to generate corresponding class labels. Subsequently, we introduce a mixed graph neural network that integrates semantic, position, and velocity information to comprehensively characterize the social interaction behavior of agents within the scene. Specifically, we have constructed three types of graph topologies and incorporated social force relationship into the adjacency matrix of the GCN, enabling accurate capture of the strength of interactions between agents. We employ ST-GCNN to extract spatio-temporal interaction features among agents in the scene. Finally, the processed graph sequence is fed into TCNs to estimate the future trajectory of the multi-class agents. More details of the implementation and methodology will be explained later.

C. Social Force Model

The social force model (SFM) simulates collective behavior by describing the interactions between different agents within the same environment [41]. This model considers both the internal motivations driving individual movements and the influence of surrounding agents on their motion. In the context of an SFM, agents are treated as point mass particles, and their movement is governed by the principles of Newton's equations of motion [27], which are expressed as

$$\frac{d^2 \vec{p}}{dt^2} = \frac{d\vec{r}}{dt} = \frac{\vec{f}_{\text{SFM}}}{m} \quad (5)$$

where $\vec{p}$ and $\vec{r}$ are represent the agent position and velocity respectively. The mass of the agent, denoted as m , is a scalar value that represents the agent's inertia and responsiveness to external forces. $\vec{f}_{\text{SFM}}$ represents the social force between agents, which influences their accelerations. It is computed by considering the interactions between an agent and its surrounding environment as well as other agents. This modeling approach aims to simulate motion behaviors similar to pedestrians, encompassing path selection, obstacle avoidance, and interactions with other agents. By embedding the social force relationships into the graph neural network, we have developed SFEM-GCN. This integration enables a more accurate simulation of agent behavior, leading to more realistic and detailed trajectory predictions.

D. Position Graph Topology Representation

Building upon SFM, it is established that the shorter the relative distance between neighboring agents, the greater their mutual influence [18], [42]. Similarly to [16], we use the relative positions of agents to construct a position graph topology. First, we construct a set of spatial graphs G_t^p that represent the relative positions of agents in the scene at each time step t . G_t^p is defined as $G_t^p = (V_t, E_t^p)$, where $V_t = \{p_t^i | \forall i \in \{1, \dots, M\}\}$ represents the set of vertices on the graph G_t^p . The observed positions (x_t^i, y_t^i) are the attributes of the corresponding vertices p_t^i . E_t^p is the set of edges in the graph G_t^p that is represented as $E_t^p = \{e_t^{ij} | \forall i, j \in \{1, \dots, M\}\}$. If v_i and v_j are connected, $e_t^{ij} = 1$; otherwise, $e_t^{ij} = 0$. To accurately model the degree of mutual influence between neighboring agents at the time step t , we introduce the adjacency matrix A_t^p and adopt the kernel function for the calculation as follows:

$$A_t^p(i, j) = \begin{cases} \frac{1}{\|p_t^i - p_t^j\|_2}, & \text{if } p_t^i \neq p_t^j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where i and j are the indices of two agents in the observed scene; $p_t^i = (x_t^i, y_t^i)$ is the position of agent i at time step t ; $p_t^j = (x_t^j, y_t^j)$ is the position of agent j at time step t ; $\|\cdot\|_2$ is the l_2 norm. Considering that agents at longer distances have a diminished influence on trajectories, we employ the inverse Euclidean distance to define the edge set of the position graph.

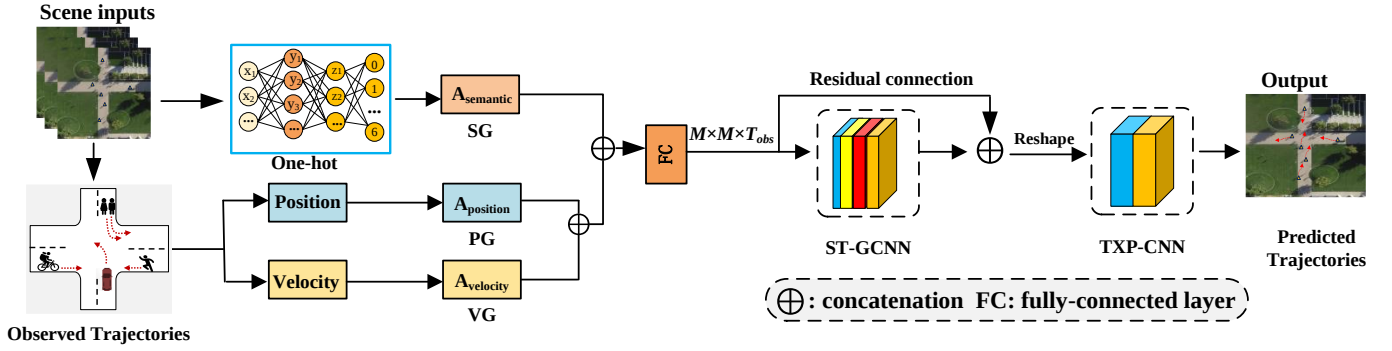


Fig. 2: Overall architecture of SFEM-GCN. Our approach constructs SG, PG, and VG by embedding social force relationships. Then, we utilize ST-GCNN to aggregate the temporal and spatial feature information from the graph nodes. Finally, we employ the TXP-CNN to obtain the predicted trajectories.

E. Velocity Graph Topology Representation

Velocity information plays a pivotal role in modeling interactions among agents [43]. While position-based interaction accurately characterizes social interactions through relative distances, the velocity of each agent is a significant factor influencing the future trajectories of others in real-world scenarios [44]. In practical settings, faster-moving agents often attract more attention from their surroundings due to rapid movement, leading to quicker approaches and an increased risk of collisions [45], [46]. To comprehensively consider agents' velocity information and enhance the accuracy of predicting future interaction behaviors, we construct a velocity-based graph network as the second sub-graph to model relationships between agents.

Similarly to the construction of the graph network based on position in III-D, we have also built a velocity graph, denoted by $G_t^v = (V_t, E_t^v)$ to model interactions relationship between agents. Here, $V_t = \{v_t^i | \forall i \in \{1, \dots, M\}\}$ is the set of vertices with M nodes, and E_t^v represents the set of edges modeling the geometric correlations between nodes in the velocity graph G_t^v , defined by the $M \times M$ adjacency matrix A_t^v , which is expressed as

$$A_t^v(i, j) = \begin{cases} \frac{1}{\|v_t^i - v_t^j\|_2}, & \text{if } v_t^i \neq v_t^j \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where $v_t^i = (x_t^i - x_{t-1}^i, y_t^i - y_{t-1}^i)$ represents the moving velocity of agent i at step t ; $v_t^j = (x_t^j - x_{t-1}^j, y_t^j - y_{t-1}^j)$ represents the moving velocity of agent j at step t .

F. Semantic Graph Topology Representation

In this section, we introduce the construction of the third sub-graph, namely the semantic graph. It involves creating an adjacency matrix guided by the label information to connect different classes of agents, such as cars, pedestrians, and cyclists. Considering that different classes of agents have varying impacts on surrounding agents, in complex traffic scenarios, we may prioritize our attention towards larger vehicles approaching us rather than cyclists [47]. Therefore, prediction models based on a single class trajectory cannot fully capture real-life scenarios involving multi-class agents interacting with

each other over long distances. However, incorporating label information from agents of different classes can significantly influence the relationships between trajectories. By assigning different weights to distinct agents, we can effectively address this issue.

Inspired by [20], we initially employ one-hot encoding to convert different classes of agents as vectors, where each vector has length L and depends on the number of classes of agents in the scene. From the one-hot encoding, the corresponding classes of agents are encoded as 1, while the remaining positions are set to 0 [6]. Through M iterations (assuming M agents in the scene), we create a tensor T of size $L \times M$. Subsequently, we construct graph relationships based on label information and establish a label-based adjacency matrix between objects. Specifically, we replicate the rows and columns of both T and the transpose matrix T^T in the list using one-hot encoding, resulting in two tensors of size $M \times M$, each of dimension L . These tensors are then reshaped and concatenated along the label dimension L , producing a tensor C with the size of $M \times M \times 2L$, which represents the interactive relationships between agents. Finally, a fully connected layer is employed to extract pertinent characteristics from the label information in C , generating a semantic adjacency matrix A^s with trainable parameters. A^s automatically encodes valuable features from label information, acquiring information crucial for trajectory prediction.

We fuse the position graph, the velocity graph, and the semantic graph to construct a mixed graph network, then connect A_t^p , A_t^v , and A^s to the characteristic dimension, and the resulting adjacency matrix is passed through a fully connected layer to produce an adjacency matrix A^f with a size of $M \times M \times T_{\text{obs}}$. The fusion process can be formalized as

$$A^f = f(A_t^p, A_t^v, A^s) \quad (8)$$

where f denotes the fully connected layer.

G. The Spatial-Temporal Graph Convolution Neural Network

Following the construction of the graph topology, we employ ST-GCNN to integrate the graph representation. This integration is achieved through the application of a spatio-temporal convolution process, as detailed in [20]. By applying

a kernel, this operation extends the array-to-graph convolution process to aggregate features within the local neighborhood of a graph across both spatio-temporal dimensions. Furthermore, we normalize the adjacency matrix A^f at each time step t , following the methodology of [16], ensuring a precise extraction of the features. The expression is as follows:

$$\hat{A}_t^f = \Lambda_t^{-\frac{1}{2}} \left(A_t^f + I \right) \Lambda_t^{-\frac{1}{2}} \quad (9)$$

where I is the identity matrix and Λ_t represents the diagonal node degree matrix with each element corresponding to the row summation of the matrix $A_t^f + I$. A_t refers to the adjacency matrix at time t , which has been normalized to incorporate weights. $\hat{A}_t^f$ is the normalized weighted adjacency matrix at time t , ensuring positive semi-definite for spectral decomposition in the GCN. The ST-GCNN is further defined by convoluting the position graph $P(l)$, the velocity graph $V(l)$ and the semantic graph $S(l)$, with the kernel $W(l)$ in layer l under the graph laplacian $\hat{A}_t^f$. The computational process is formulated as

$$\varphi(V(l), A) = \delta \left(\Lambda_t^{-\frac{1}{2}} \hat{A}_t^f \Lambda_t^{-\frac{1}{2}} V(l) W(l) \right) \quad (10)$$

H. Trajectory Generation

Beyond considering the spatial information of agents, incorporating the temporal information of trajectories is crucial for accurate trajectory prediction. Drawing inspiration from [16], we employ a novel approach that combines TCN and CNN techniques to process multi-class agent trajectories, named the time extrapolator convolutional neural network (TXP-CNN). By leveraging convolutional operations in the temporal domain, we effectively capture intricate temporal relationships across multiple time scales. This enables us to make sequential predictions while addressing the issue of error accumulation commonly observed in RNN-based methods. Specifically, we employ a temporal stacking approach in which multiple convolutional layers are stacked along the temporal dimension of the output feature map $\hat{V}$ originating from the last ST-GCNN layer. Each output from the temporal layer is then residual connected to the previous layer.

IV. EXPERIMENT AND EVALUATION

We conduct evaluation experiments on a widely used benchmark dataset and compare the results with other state-of-the-art methods. Specifically, we evaluate SFEM-GCN model on the stanford drone dataset (SDD) [48], which offers trajectory data and class labels for various object classes. The SDD comprises six different classes: biker, pedestrian, car, cart, bus, and skater. The scenes in the SDD are captured from a top-down view using drones, and the tracking coordinates are measured in pixels. The structure of GCN used in SFEM-GCN is similar to those [8], [18], [20], we also input the trajectory of eight time steps (3.2 s) and predict the next 12 time steps (4.8 s). To assess the effectiveness of our method, we randomly sampled 20 trajectories ($K = 20$) from the predicted multinomial distribution.

A. Experiments

- 1) **Evaluation Criteria:** Following the approach employed by other baseline, the performance of the model is evaluated using the Average Displacement Error (ADE) and the Final Displacement Error (FDE) metrics introduced in [16]. ADE measures the average prediction error across all time steps, while FDE focuses on the error at the final time step. Recognizing the multimodal characteristics of agent motion, we leverage the bivariate Gaussian distribution generated by the proposed model and adopt the Minimum Average Displacement Error (mADE) and Minimum Final Displacement Error (mFDE) metrics [49] to evaluate the performance of our model. To do so, we sample K instances from the predicted distribution and identify the instances that produce the minimum errors. mADE and mFDE are used as the standard metrics, which are defined as

$$mADE_k = \frac{\sum_{i=1}^N \min_{k=1}^K \sum_{t=T_{\text{obs}}+1}^{T_{\text{pred}}} \|Y_t^i - \hat{Y}_t^{ik}\|_2}{N \times (T_{\text{pred}} - T_{\text{obs}})} \quad (11)$$

$$mFDE_k = \frac{\sum_{i=1}^N \min_{k=1}^K \|Y_t^i - \hat{Y}_t^{ik}\|_2}{N}, t = T_{\text{pred}} \quad (12)$$

where the ground truth of the future trajectories is denoted as $Y_t^i = (x_t^i, y_t^i)$, and $\hat{Y}_t^{ik}$ represents the predicted position of the k -th ($1 \leq k \leq K$) sampled trajectory at time t . Recognizing that relying solely on minimum errors may not adequately demonstrate the robustness of the generated bivariate Gaussian distributions, we introduce new metrics referred to as the Average² Displacement Error (aADE) and the Average Final Displacement Error (aFDE) [20]. These metrics offer a more comprehensive comparison of models and are defined as follows:

$$aADE = \frac{\sum_{i=1}^N \sum_{t=T_{\text{obs}}+1}^{T_{\text{pred}}} \sum_{s=1}^S \|Y_t^i - \hat{Y}_t^{is}\|_2}{N \times T_{\text{pred}} \times S} \quad (13)$$

$$aFDE = \frac{\sum_{i=1}^N \sum_{s=1}^S \|Y_t^i - \hat{Y}_t^{is}\|_2}{N \times S}, t = T_{\text{pred}} \quad (14)$$

where S denotes a set of samples generated by sampling from a bivariate Gaussian distribution. The metrics aADE and aFDE are more accurate than mADE and mFDE, as they take into account S samples from the predicted bivariate Gaussian distribution and average the errors across these samples.

- 2) **Implementation Details:** The proposed solution was implemented using the PyTorch deep learning framework [50]. The NVIDIA GV100 GPU was used for training and evaluation. The model was trained for 250 epochs with a batch size of 128. Stochastic gradient descent (SGD) was employed as the neural network optimizer. The initial learning rate was set to 0.01, and the decay was set to 0.002 after 150 epochs. The PReLU activation function [51] was utilized in the model architecture.

TABLE I: QUANTITATIVE COMPARISONS WITH MADE AND MFDE (MINIMUM ERROR).

Methods \ Metrics	Linear	SF	S-LSTM	SGAN	CAR-Net	DESIRE	Trajectron++	SSTGCNN	Semantics-STGCNN	Ours
mADE ↓	37.11	36.48	31.19	27.25	25.72	19.25	19.30	26.46	18.12	15.31
mFDE ↓	63.51	58.14	56.97	41.44	51.80	34.05	32.70	42.71	29.70	25.72

TABLE II: QUANTITATIVE ANALYSIS WITH AADE AND AFDE (AVERAGE ERROR).

Methods \ Metrics	SSTGCNN	Semantics-STGCNN	Ours
aADE ↓	49.96	33.14	25.45
aFDE ↓	90.10	61.14	48.28

TABLE III: NETWORK COMPARISONS WITH THE NUMBER OF TRAINABLE PARAMETERS AND INFERENCE TIME (S).

Methods \ Metrics	SSTGCNN	Semantics-STGCNN	Ours
Parameters Number ↓	7563	7852	4420
Inference Time ↓	0.302	0.309	0.276

B. Baselines

We compared the proposed method with state-of-the-art baselines as follows:

- 1) Linear [8]: This method utilizes a linear regression to estimate the linear parameters by minimizing the least squares error.
- 2) SF [25]: This method incorporates the implementation of the social force model, which includes modeling several factors such as group affinity and predicted destinations.
- 3) S-LSTM [8]: The proposed method is an LSTM-based trajectory prediction model that incorporates an innovative pooling mechanism.
- 4) DESIRE [52]: This method introduces a conditional variational auto-encoder system, which is based on the concept of inverse optimal control.
- 5) SGAN [9]: This method introduces a GAN into pedestrian trajectory prediction and incorporates global pooling for interaction modeling.
- 6) CAR-Net [53]: This method introduces an attentive recurrent model called CAR-Net, which integrates the information of saliency regions within the scenes.
- 7) Trajectron++ [54]: This method has the ability to integrate diverse data sources beyond historical trajectory information and is capable of generating predictions for future conditions that adhere to dynamical constraints. It can be classified as a generative multi-agent trajectory prediction method.
- 8) SSTGCNN [16]: This method models pedestrian trajectories as graphs and utilizes STGCNN to capture and analyze social interactions.
- 9) Semantics-STGCNN [20]: This method introduces a state-of-the-art model for multi-class trajectory prediction based on an undirected graph architecture. Leveraging undirected edges in the graph, this model effectively captures complex relationships and dependencies between different classes of objects, leading to accurate trajectory predictions.

C. Quantitative Analysis

In this subsection, we present a comparison between our proposed method and nine recently proposed methods. Table

I shows the results achieved by each method in various scenarios. Based on these findings, our method has demonstrated outstanding performance in multi-class trajectory prediction. Specifically, the Linear method [8], which assumes a linear correlation between the motion attributes of agents, tends to deviate significantly from real-world social patterns, resulting in comparatively less accurate results. SF [25] provides a simplistic approach to trajectory prediction but has limitations in capturing the complexity and individuality of human behavior in real-world scenarios. Due to the inherent errors introduced during trajectory prediction by RNN, LSTM-based methods [8], [9] have failed to achieve competitive results. This also highlights the limitations of the pooling structure based on LSTM hidden states in accurately extracting interaction features in complex scenarios. Although CAR-Net [53] can incorporate salient information from the scene, its performance remains suboptimal due to the uncertainty of agent movements and the complexity of the environment. DESIRE [52], which uses a conditional variational auto-encoder model for multimodal trajectory prediction, achieves promising results. Trajectron++ [54], with its advanced prediction framework and unique modeling approach, consistently remains at the forefront of trajectory prediction methods, outperforming contemporaneous approaches in terms of evaluation criteria. Our method builds on the foundations of SSTGCNN [16] and Semantics-STGCNN [20], which fully incorporates prior knowledge of social interactions between different classes of agents, enabling the model to better balance social force relationships among agents. Taking into account the significant contributions mentioned above, our model achieves a 11% reduction in mADE and a 17% reduction in mFDE compared to SSTGCNN. Compared to Semantics-STGCNN, our method reduces mADE by 3% and mFDE by 4%.

In Table II, we conducted a comprehensive comparison of the experimental results for the evaluation metrics aADE and aFDE, as introduced in a recent method [16], [20]. Our method takes into account the influence of position, velocity, and semantic information on social interactions among agents, leading to state-of-the-art performance in comparative experiments. Specifically, we observed significant improvement by

8% and 13% in the aADE and aFDE metrics, respectively.

Furthermore, we place significant emphasis on the inference time of our method, recognizing its crucial role in ensuring the safety of autonomous driving systems. Table III provides a detailed comparison of the model's parameter count and inference time. Encouragingly, our method, designed for agent interaction modeling, not only reduces the parameter count but also achieves faster inference speed, making it the optimal choice. This also highlights the advantages of social force-based interaction models. Importantly, even with the utilization of multiple graphs to capture social force relationships among different classes agents, the inference performance of our method remains unaffected.

D. Qualitative Analysis

In Figure. 3, we showcase the experimental prediction results of two methods, Semantics-STGCNN and SFEM-GCN, for one frame from three different scenarios of the SDD. Overall, our model outperforms Semantics-STGCNN by generating more realistic predicted trajectories. Specifically, in the leftmost image in Figure. 3, where there is minimal interaction among the agents, even when the motion trajectories between agents exhibit a linear relationship, Semantics-STGCNN fails to handle the behaviors of agents at different positions effectively, resulting in deviations between the predicted and ground truth. As for the middle image in Figure. 3, where most agents move in different directions and significant interaction occurs, Semantics-STGCNN shows noticeable directional biases, leading to significant deviations between the predicted and ground truth trajectories, particularly in destination predictions. For example, in the result of Semantics-STGCNN from the second subfigure on the middle-hand side, the cyclist's predicted destination as the lawn is evidently unreasonable. In contrast, our predictions align closely with the ground truth. Additionally, in the rightmost image in Figure. 3, we present the motion of multiple agent classes in a crowded scene, which resembles real-world autonomous driving scenarios and poses more complex challenges in handling. From the visual results, both Semantics-STGCNN and SFEM-GCN exhibit some incorrect predictions. For example, in the circular road scenario depicted in the image, both methods make erroneous predictions regarding the agents' turning intentions. Considering the higher uncertainty in predicting the intentions of different agents in crowded and highly dynamic scenes, such erroneous predictions are understandable. However, compared to Semantics-STGCNN, SFEM-GCN achieves better trajectory prediction results in most cases by incorporating position and velocity graph to balance the social force relationships among agents and utilizing semantic graph to assign different label as guidance for various agents, enabling more effective learning of the differences between trajectories of multi-class agents. In summary, the visual results from these three different scenarios demonstrate that the three constructed graph topologies effectively simulate the social interactions between multi-class agents, leading to more realistic trajectory prediction.

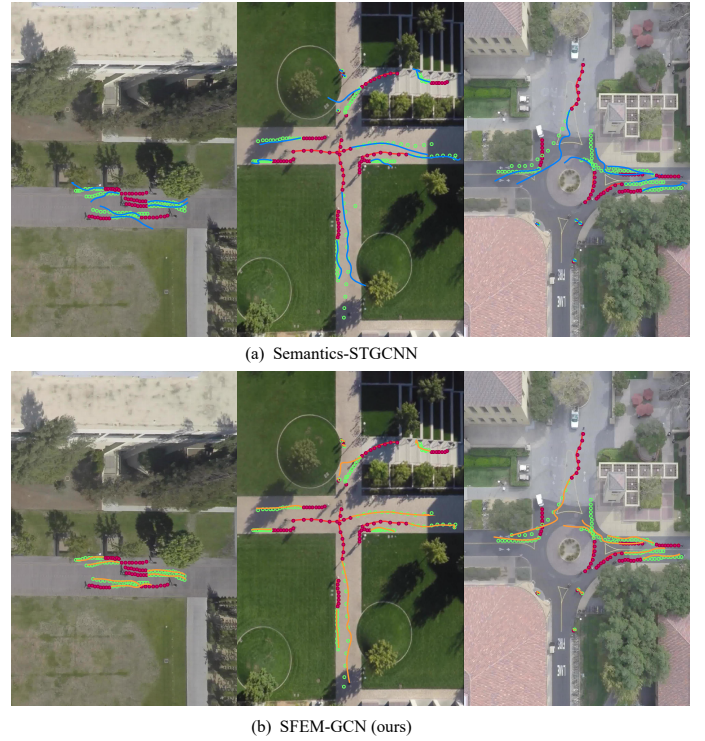


Fig. 3: Visualization of the trajectory distribution predicted by compare models of multi-class agent motion on SDD. Red filled circle is observed trajectories (8 frames), green hollow circles are ground-truth (12 frames), blue lines in (a) are predicted results by Semantics-STGCNN, orange lines in (b) are predicted results by the proposed SFEM-GCN.

E. Ablation study

- 1) **Ablation study on model architecture:** Our method is constructed by stacking multiple layers of ST-GCNN and TXP-CNN. The selection of appropriate parameters for the network structure is crucial to optimizing the results of trajectory prediction. Therefore, we employ a similar approach as presented in [55] to test and select the optimal parameters for the network structure. In each experiment, we maintained the same hyperparameter settings for both the testing and training stages. The maximum number of layers for ST-GCNN and TXP-CNN was set to 5. The experimental results can be found in Table IV. Through testing with different numbers of network layers, we found that setting ST-GCNN and TXP-CNN to 4 layers and 2 layers, respectively, yielded optimal values for the metrics.
- 2) **Ablation study on model components:** To demonstrate the performance of each component in the proposed model, we tested three different variants on the SDD. The test results are presented in Table V. Specifically, “w/o SG” indicates the removal of the semantic information sub-graph component, “w/o PG” represents the removal of the position information sub-graph component, and “w/o VG” represents the exclusion of the velocity information sub-graph component. The results indicate that each component of the proposed model significantly

TABLE IV: ABLATION STUDY OF OUR MODEL. THE FIRST ROW CORRESPONDS TO THE NUMBER OF TXP-CNN LAYERS, AND THE FIRST LIST CORRESPONDS TO THE NUMBER OF ST-GCNN LAYERS.

Methods	1	2	3	4	5
1	18.35/31.16	19.27/34.61	18.89/32.98	20.72/37.34	20.25/32.76
2	16.85/29.11	20.61/30.33	17.29/28.65	18.88/33.28	21.25/34.71
3	20.67/36.94	15.95/27.88	21.15/37.38	20.84/29.71	18.16/31.50
4	17.68/31.22	15.31/25.72	21.75/38.64	17.48/27.81	17.50/31.72
5	24.51/34.91	20.28/33.95	25.84/45.93	20.11/31.06	19.99/33.61

TABLE V: ABLATION STUDY OF OUR MODEL IN TERMS OF MADE/MFDE/ADE/AFDE METRICS.

Methods	mADE	mFDE	aADE	aFDE
w/o SG	17.46	28.68	33.04	61.36
w/o PG	26.65	52.09	39.11	78.16
w/o VG	19.88	28.25	29.36	56.58
Ours	15.31	25.72	25.45	48.28

contributes to the accuracy of trajectory prediction. Among these, the positional information between agents has a more pronounced impact on trajectory prediction. Specifically, the adoption of multi-graph fusion enables effective modeling of social interaction relationships between different classes of agents. Additionally, guidance based on social force relationship is crucial for improving the performance of the model.

V. CONCLUSION

In this paper, we propose SFEM-GCN, a multi-class trajectory prediction method that incorporates various factors of social and scene interaction to construct a mixed graph topology based on social interaction. When handling social interaction information between different classes of agents, we use coordinate data to calculate the positions and velocities of the agents, constructing corresponding adjacency matrices to measure their influences. Simultaneously, we consider scene context information by constructing a graph adjacency matrix guided by semantic information, highlighting the varying influences of different classes in the scene. Attention weights are computed by incorporating information from the label. The constructed mixed graph topology is then input to ST-GCNN with residual structures, and predicted trajectories are generated using TCNs. We conducted experiments on the publicly available dataset SSD and compared our method with state-of-the-art approaches. The quantitative analysis of the experimental results has shown that our method outperforms existing methods in terms of accuracy and generalization ability of trajectory prediction. Furthermore, qualitative analysis confirms that our method accurately predicts the motion trajectories of multi-class agents in the scene, yielding more plausible trajectory predictions.

While our method has demonstrated promising performance on the open dataset SSD, accurately predicting the future motion trajectories of agents in high-dynamic and crowded

scenes remains a challenging task. For instance, issues such as inaccurate inference of agent turning intentions leading to prediction failures need to be addressed. Therefore, further research is necessary in addressing these challenging scenarios. Additionally, future work can explore the use of attention mechanisms [56]–[58] in deep learning to prioritize key agents, capture their motion patterns more accurately, and achieve precise trajectory predictions.

REFERENCES

- [1] L. Chen, X. Hu, W. Tian, H. Wang, D. Cao, and F.-Y. Wang, “Parallel planning: A new motion planning framework for autonomous driving,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 1, pp. 236–246, 2018. I
- [2] T. Liu, B. Tian, Y. Ai, and F.-Y. Wang, “Parallel reinforcement learning-based energy efficiency improvement for a cyber-physical system,” *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 2, pp. 617–626, 2020. I
- [3] P. Ghorai, A. Eskandarian, Y.-K. Kim, and G. Mehr, “State estimation and motion prediction of vehicles and vulnerable road users for cooperative autonomous driving: A survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 16 983–17 002, 2022. I
- [4] Y. Tian, X. Li, H. Zhang, C. Zhao, B. Li, X. Wang, and F.-Y. Wang, “Vistagpt: Generative parallel transformers for vehicles with intelligent systems for transport automation,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 9, pp. 4198–4207, 2023. I
- [5] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, “A survey on trajectory-prediction methods for autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 652–674, 2022. I, II-A
- [6] X. Wang, K. Tang, X. Dai, J. Xu, J. Xi, R. Ai, Y. Wang, W. Gu, and C. Sun, “Safety-balanced driving-style aware trajectory planning in intersection scenarios with uncertain environment,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2888–2898, 2023. I, III-F
- [7] L. Shi, L. Wang, S. Zhou, and G. Hua, “Trajectory unified transformer for pedestrian trajectory prediction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9675–9684. I
- [8] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, F.-F. Li, and S. Savarese, “Social LSTM: Human trajectory prediction in crowded spaces,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 961–971. I, III-A, IV, 1, 3, IV-C, IV-C
- [9] A. Gupta, J. Johnson, F.-F. Li, S. Savarese, and A. Alahi, “Social GAN: Socially acceptable trajectories with generative adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264. I, II-A, 5, IV-C
- [10] K. Zhang, L. Zhao, C. Dong, L. Wu, and L. Zheng, “Ai-tp: Attention-based interaction-aware trajectory prediction for autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 73–83, 2022. I
- [11] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, “Sophie: An attentive GAN for predicting paths compliant to social and physical constraints,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1349–1358. I
- [12] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, H. Rezatofighi, and S. Savarese, “Social-bigat: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks,” in *Proceedings of the Advances in Neural Information Processing Systems*, 2019, pp. 137–146. I, II-B
- [13] X. Mo and C. Lv, “Predictive neural motion planner for autonomous driving using graph networks,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1983–1993, 2023. I
- [14] Z. Sheng, Y. Xu, S. Xue, and D. Li, “Graph-based spatial-temporal convolutional network for vehicle trajectory prediction in autonomous driving,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17 654–17 665, 2022. I
- [15] M. Golchoubian, M. Ghafurian, K. Dautenhahn, and N. L. Azad, “Pedestrian trajectory prediction in pedestrian-vehicle mixed environments: A systematic review,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 11 544–11 567, 2023. I

- [16] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 424–14 432. I, 2, II-A, II-B, II-B, III-A, III-D, III-G, III-H, I, 8, IV-C, IV-C
- [17] L. Shi, L. Wang, C. Long, S. Zhou, M. Zhou, Z. Niu, and G. Hua, "Sgcnn: Sparse graph convolution network for pedestrian trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8994–9003. I, II-B
- [18] P. Lv, W. Wang, Y. Wang, Y. Zhang, M. Xu, and C. Xu, "Ssagcn: social soft attention graph convolutional network for pedestrian trajectory prediction," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023. I, II-B, III-D, IV
- [19] Shen, Yu and Wang, Xiao, and Han, Shuangshuang, and Chen, Long and Wang, Feiyue, "Agent-based technology in intelligent vehicles and driving: state-of-the-art and prospect," *Journal of Command and Control*, vol. 5, no. 2, pp. 87–98, 2019. I
- [20] B. A. Rainbow, Q. Men, and H. P. Shum, "Semantics-stgcnn: A semantics-guided spatial-temporal graph convolutional network for multi-class trajectory prediction," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2021, pp. 2959–2966. I, 3, II-B, III-F, III-G, IV, 1, 9, IV-C, IV-C
- [21] H. Xue, D. Q. Huynh, and M. Reynolds, "SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction," in *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2018, pp. 1186–1194. I
- [22] Y. Xu, L. Wang, Y. Wang, and Y. Fu, "Adaptive trajectory prediction via transferable gnn," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6520–6531. II-A
- [23] Z. Zuo, X. Wang, S. Guo, Z. Liu, Z. Li, and Y. Wang, "Trajectory prediction network of autonomous vehicles with fusion of historical interactive features," *IEEE Transactions on Intelligent Vehicles*, pp. 1–13, 2023. II-A
- [24] D. Xu, X. Shang, Y. Liu, H. Peng, and H. Li, "Group vehicle trajectory prediction with global spatio-temporal graph," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1219–1229, 2022. II-A
- [25] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg, "Who are you with and where are you going?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 1345–1352. II-A, II-A, 2, IV-C
- [26] P. Ye, X. Wang, C. Chen, Y. Lin, and F. Wang, "Hybrid agent modeling in population simulation: Current approaches and future directions," *Journal of Artificial Societies and Social Simulation*, vol. 19, no. 1, p. 12, 2016. II-A
- [27] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, p. 4282, 1995. II-A, III-C
- [28] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrilu, "Context-based pedestrian path prediction," in *Proceedings of the European Conference on Computer Vision*. Springer, 2014, pp. 618–633. II-A
- [29] D. Ellis, E. Sommerlade, and I. Reid, "Modelling pedestrian trajectory patterns with gaussian processes," in *Proceedings of the International Conference on Computer Vision*. IEEE, 2009, pp. 1229–1234. II-A
- [30] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, "Activity forecasting," in *Proceedings of the European Conference on Computer Vision*. Springer, 2012, pp. 201–214. II-A
- [31] Z. Ding and H. Zhao, "Incorporating driving knowledge in deep learning based vehicle trajectory prediction: A survey," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 8, pp. 3996–4015, 2023. II-A
- [32] R. Korbmaier and A. Tordeux, "Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24 126–24 144, 2022. II-A
- [33] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "Stgat: Modeling spatial-temporal interactions for human trajectory prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6272–6281. II-A, II-B, II-B
- [34] T. Westny, J. Oskarsson, B. Olofsson, and E. Frisk, "Mtp-go: Graph-based probabilistic multi-agent trajectory prediction with neural odes," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 9, pp. 4223–4236, 2023. II-A
- [35] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2020. II-B
- [36] S. Teng, X. Hu, P. Deng, B. Li, Y. Li, Y. Ai, D. Yang, L. Li, Z. Xuanyuan, F. Zhu *et al.*, "Motion planning for autonomous driving: The state of the art and future perspectives," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 6, pp. 3692–3711, 2023. II-B
- [37] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-RNN: Deep learning on spatio-temporal graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5308–5317. II-B
- [38] R. W. Liu, M. Liang, J. Nie, Y. Yuan, Z. Xiong, H. Yu, and N. Guizani, "Stmgcn: Mobile edge computing-empowered vessel trajectory prediction using spatio-temporal multigraph convolutional network," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7977–7987, 2022. II-B
- [39] X. Mo, Z. Huang, Y. Xing, and C. Lv, "Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9554–9567, 2022. II-B
- [40] C. Zhang, Z. Ni, and C. Berger, "Spatial-temporal-spectral LSTM: A transferable model for pedestrian trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, pp. 1–14, 2023. III-A
- [41] L. Crosato, H. P. Shum, E. S. Ho, and C. Wei, "Interaction-aware decision-making for automated vehicles using social value orientation," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1339–1349, 2022. III-C
- [42] Y. Li, X. Lu, J. Wang, and K. Li, "Pedestrian trajectory prediction combining probabilistic reasoning and sequence learning," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 3, pp. 461–474, 2020. III-D
- [43] L. Yang, X. Wang, J. J. Zhang, M. Zhou, and F. Wang, "Pedestrian choice modeling and simulation of staged evacuation strategies in daya bay nuclear power plant," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 3, pp. 686–695, 2020. III-E
- [44] Y. Su, J. Du, Y. Li, X. Li, R. Liang, Z. Hua, and J. Zhou, "Trajectory forecasting based on prior-aware directed graph convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16 773–16 785, 2022. III-E
- [45] H. Minoura, T. Hirakawa, Y. Sugano, T. Yamashita, and H. Fujiyoshi, "Utilizing human social norms for multimodal trajectory forecasting via group-based forecasting module," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 836–850, 2022. III-E
- [46] M. Zhou, W. Xu, X. Liu, Z. Zhang, H. Dong, and D. Wen, "Acp-based parallel railway traffic management for high-speed trains in case of emergencies," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 11, pp. 4588–4598, 2023. III-E
- [47] Wang, Xiao and Zhang, Xiangyu, and Zhou, Rui, and Tian, Yonglin, and Wang, Jiangong, and Chen, Long, and Sun, Changyin, "An intelligent architecture for cognitive autonomous driving based on parallel testing," *Acta Automatica Sinica*, vol. 49, p. 1, 2023. III-F
- [48] L. F. Chiara, P. Coscia, S. Das, S. Calderara, R. Cucchiara, and L. Ballan, "Goal-driven self-attentive recurrent networks for trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2518–2527. IV
- [49] C. Xu, M. Li, Z. Ni, Y. Zhang, and S. Chen, "Groupnet: Multiscale hypergraph neural networks for trajectory prediction with relational reasoning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6498–6507. I
- [50] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proceedings of the Advances in Neural Information Processing Systems*, 2019, pp. 8026–8037. 2
- [51] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034. 2
- [52] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 336–345. 4, IV-C
- [53] A. Sadeghian, F. Legros, M. Voisin, R. Vesel, A. Alahi, and S. Savarese, "Car-net: Clairvoyant attentive recurrent network," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 151–167. 6, IV-C
- [54] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 683–700. 7, IV-C
- [55] R. Chai, A. Tsourdos, A. Savvaris, S. Chai, Y. Xia, and C. P. Chen, "Multiobjective overtaking maneuver planning for autonomous ground vehicles," *IEEE Transactions on Cybernetics*, vol. 51, no. 8, pp. 4035–4049, 2020. I

- [56] K. Zhang, X. Feng, L. Wu, and Z. He, "Trajectory prediction for autonomous driving using spatial-temporal graph attention transformer," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22 343–22 353, 2022. V
- [57] D. Yang, H. Zhang, E. Yurtsever, K. A. Redmill, and Ü. Özgüner, "Predicting pedestrian crossing intention with feature fusion and spatio-temporal attention," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 221–230, 2022. V
- [58] X. Wang, K. Tang, X. Dai, J. Xu, Q. Du, R. Ai, Y. Wang, and W. Gu, "S4tp: Social-suitable and safety-sensitive trajectory planning for autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, pp. 1–12, 2023. V



Huansheng Ning (Senior Member, IEEE) received his B.S. degree from Anhui University in 1996 and his Ph.D. degree from Beihang University in 2001. Now, he is a professor of the School of Computer and Communication Engineering, University of Science and Technology Beijing, China. His current research focuses on the Internet of Things and general cyberspace. He is the founder of the cyberspace and cybermatics international science and technology cooperation base.



Quancheng Du received his B.S. degree from Henan Normal University in 2018. He currently working toward his Ph.D. degree in the School of Computer and Communication Engineering, University of Science and Technology Beijing, China. His main research interests include autonomous driving and trajectory prediction.



nizations, and multi-agent modeling.

Xiao Wang (Senior Member, IEEE) received the Bachelor's degree in network engineering from Dalian University of Technology, Dalian, China, in 2011, and the Ph.D. degree in social computing from the University of Chinese Academy of Sciences, Beijing, China, in 2016. She is currently a professor with the School of Artificial Intelligence, Anhui University, Hefei, and the President of the Qingdao Academy of Intelligent Industries, Qingdao, China. Her research interests include social network analysis, social transportation, cybermovement organizations, and multi-agent modeling.



Shouguo Yin received a bachelor's degree from Sichuan Normal University in 2023. He is currently pursuing a master's degree at the School of Artificial Intelligence, Anhui University, China. His current research focuses on intelligent transportation systems and trajectory prediction.



Lingxi Li (Senior Member, IEEE) received the B.E. degree in automation from Tsinghua University, Beijing, China, in 2000, the M.S. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2003, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2008. Since August 2008, he has been with Indiana University-Purdue University Indianapolis (IUPUI), Indianapolis, IN, USA, where he is currently a

Professor of electrical and computer engineering. He has authored/coauthored one book and more than 100 research articles in refereed journals and conferences. His current research focuses on the modeling, analysis, control, and optimization of complex systems, intelligent transportation systems, digital twins and parallel intelligence, and human-machine interaction.