This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Trajectory Forecasting Based on Prior-Aware Directed Graph Convolutional Neural Network

Yuchao Su, *Student Member, IEEE*, Jie Du, *Member, IEEE*, Yuanman Li, *Member, IEEE*, Xia Li, *Member, IEEE*, Rongqin Liang, *Student Member, IEEE*, Zhongyun Hua, *Member, IEEE*, and Jiantao Zhou, *Senior Member, IEEE*

***Abstract*— Predicting the motion trajectories of moving agents in complex traffic scenes, such as crossroads and roundabouts, plays an important role in cooperative intelligent transportation systems. Nevertheless, accurately forecasting the motion behavior in a dynamic scenario is challenging due to the complex cooperative interactions between moving agents. Graph Convolutional Neural Network has recently been employed to deal with the cooperative interactions between agents. Despite the promising performance of resulting trajectory prediction algorithms, many existing graph-based approaches model interactions with an undirected graph, where the strength of influence between agents is assumed to be symmetric. However, such an assumption often does not hold in reality. For example, in pedestrian or vehicle interaction modeling, the moving behavior of a pedestrian or vehicle is highly affected by the ones ahead, while the ones ahead usually pay less attention to the ones behind. To fully exploit the asymmetric attributes of the cooperative interactions in intelligent transportation systems, in this work, we present a directed graph convolutional neural network for multiple agents trajectory prediction. First, we propose three directed graph topologies, i.e., view graph, direction graph, and rate graph, by encoding different prior knowledge of a cooperative scenario, which endows the capability of our framework to effectively characterize the asymmetric influence between agents. Then, a fusion mechanism is devised to jointly exploit the asymmetric mutual relationships embedded in constructed graphs. Furthermore, a loss function based on Cauchy distribution is designed to generate multimodal trajectories. Experimental results on complex traffic scenes demonstrate the superior performance of our proposed model when compared with existing approaches.**

***Index Terms*— Cooperative intelligent transportation systems, trajectory prediction, directed graph convolutional neural network, asymmetric interactions.**

Manuscript received June 20, 2021; revised October 8, 2021 and November 30, 2021; accepted December 31, 2021. This work was supported in part by the Natural Science Foundation of China under Grant 62001304, Grant 61871273, Grant 61971476, and Grant 62006160; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515110410; in part by the Educational Commission of Guangdong Province under Grant 2020KQNCX062; in part by the Shenzhen Fundamental Research Program under Grant 20200813102946001; and in part by the Macau Science and Technology Development Fund under Grant SKL-IOTSC-2021-2023, Grant 0072/2020/AMJ, Grant 077/2018/A2, and Grant 0060/2019/A1. The Associate Editor for this article was S. H. Shah. *(Yuchao Su and Jie Du contributed equally to this work.) (Corresponding author: Yuanman Li.)*

Yuchao Su, Yuanman Li, Xia Li, and Rongqin Liang are with the Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: 1950432007@email.szu.edu.cn; yuanmanli@szu.edu.cn; lixia@szu.edu.cn).

Jie Du is with the Health Science Center, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: dujie@szu.edu.cn).

Zhongyun Hua is with the School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China (e-mail: huazhongyun@hit.edu.cn).

Jiantao Zhou is with the State Key Laboratory of Internet of Things for Smart City and the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: jtzhou@um.edu.mo).

Digital Object Identifier 10.1109/TITS.2022.3142248

I. Introduction

COOPERATIVE intelligent transportation systems are advanced traffic management for improving social and economic efficiency [1]–[3]. Moving agents as main traffic participants are the most important varieties in traffic systems. Thus, forecasting the motion behaviors of multiple moving agents in complex traffic scenes, such as crossroads and roundabouts, is essential for a number of intelligent traffic applications, such as traffic planning, traffic management, and traffic control, etc. [4], [5]. Despite its importance, accurately predicting the trajectories of agents is very challenging in cooperative intelligent transportation systems due to the following facts. First, the social interactions between agents can significantly impact their moving trajectories, such as moving with a group behavior, avoiding a collision, and catching up or following others. Modeling these complex interactions is rather difficult in reality. Second, the motion patterns of agents are highly multimodal. Namely, given the same historical trajectory, an agent can select multiple feasible future paths to a destination.

Pioneering works for trajectory prediction mainly adopted handcrafted features [6], but these methods cannot sufficiently characterize the complex social interactions between agents and fail to model the multimodal feature due to a single predicted trajectory generated for a given historical path. Recently, a number of machine learning models [2], [7]–[11] have attracted the attention of researchers and designed to forecast trajectories for cooperative intelligent transportation systems, including Recurrent Neural Networks [12], Generative Adversarial Networks (GANs) [13], and graph-based models [14].

The widely known approach, called Social-LSTM, was proposed in [15] as the pioneer of the deep learning model for pedestrian trajectory prediction. In [15], Long Short-Tern Memory (LSTM) was intuitively adopted to learn the trajectories and a pooling layer based on a grid method was proposed to extract the hidden states of LSTM for different agents as the social interactions. However, the multimodal trajectories of agents are not considered. As a result, an LSTM method was reported in [16] by using a Gaussian mixture model for the multimodal trajectories of agents. While in the recent works [17]–[19], the authors proposed to improve the performance of [15] by using new pooling mechanisms to utilize the scene context as side information. In [20], an LSTM method was leveraged to model the moving behaviors of all agents by considering the different importance of agents for a target agent. Recently, CF-LSTM [21] was proposed to utilize the vital velocity features of agents for improving the performance of LSTM. Very recently, an LSTM model [22] was also designed for cyclists' movement prediction, in which an LSTM encoder was used to interact with cyclists and roads and a focal attention mechanism was used to improve the performance of LSTM. For improving recurrent neural networks, Gate Recurrent Unit (GRU) was also adopted for trajectory prediction. In [23], a convolutional reward model was introduced to generate multimodal plans on a 2D grid, and then GRU was used to decode and output valued multimodal trajectories. For generating diverse multimodal paths, conditional variational autoencoders were also employed in [24], [25] for trajectory prediction. However, these models have two major issues. One of which is the abstract of agents' relationship. Though recurrent models were simple ways for sequence prediction problems but is difficult to describe the spatial relationship of agents. Most of them adopt the inside features of recurrent models as the quantitative social relationship instead of using some intuitive metrics, such as the Euclidean distance between agents. It causes the weak interpretation for social relationships between agents. The other issue is that the multimodal trajectory generated by a preset probability distribution may be inappropriate in some cases because of the uncertainty of human social behavior.

To deal with the uncertainty of human social behavior, GANs have been employed to generate the multimodal trajectory by learning implicit probability distributions [13], [26]. In [13], a GAN model was firstly reported to learn implicit probability distributions for generating the multimodal trajectories of pedestrians. A similar work adopting GRU was reported in [27]. While in [28], an attention pooling was proposed to extract the temporal-spatial features of each agent to improve the performance of GAN. In [29], a GAN model with a reinforcement learning was proposed for generating real and safe trajectories. MATF [30] encoded observed trajectories of agents and the scene contexts were used to capture the social interactions through an adversarial loss. The syntax tree features were adopted into GAN [31] to integrate temporal logic rules for traffic agent trajectory prediction. Very recently, a GAN model [32] based on a temporal pyramid network was also reported to intensify the extraction of the temporal features. However, these methods also face two issues for the agent trajectory prediction. First, though the GAN models can utilize implicit probability distributions to generate multimodal trajectories, the outcomes are uncontrollable and some prior knowledge of the probability distributions are difficult to be utilized. Second, similar to those LSTM-based approaches, GANs cannot also effectively describe the social interactions between agents, and need additional modules to handle the social interactions between agents.

To more directly model the social interactions of agents, methods based on Graph Attentions (GATs) have recently been designed [33]–[35]. A model formulated the agent trajectory prediction by a spatio-temporal graph in [33], while in [34] a social graph network was reported to construct the position and direction of agents. In [35], a GAT model was embedded into GAN for the social attention and multimodal trajectory. An LSTM model with a GAT module [14] was also introducted to interact with the hidden features of LSTM. Though these GAT methods can logically describe the spatial attention of each agent by constructing graph information, they still lose some real social interactions. In fact, the GATs focus on connecting the hidden state of each agent extracted by LSTM, GAN, or other modules, instead of using the raw spatial information from agents' observed paths.

Recently, Graph Convolution Neural Networks (GCNs) have been received increasing attention and many works have been reported for cooperative intelligent transportation systems [1], [36]. For pedestrian trajectory prediction, a method based on a GCN model, called Social-STGCNN [37], have been introduced to directly construct an undirected graph topology by pedestrians' observed trajectories, and the undirected graph topologies and observed trajectories of pedestrians are inputted into a GCN model for fusing spatial-temporal features. The experimental results have verified that Social-STGCNN is superior to those models based on LSTMs, GANs, and GATs.

Despite the advance of Social-STGCNN, it still fails to correctly construct the interaction relationship between pedestrians due to the undirected graph topology, in which the social interactions are naively assumed as the equipotent impacts for each two pedestrians. However, social interactions can be highly asymmetric in reality. In this work, we propose a directed graph method to describe the asymmetric interactions among agents, called VDRGCN, which includes three directed graph topologies, i.e., View Graph (VG), Direction Graph (DG), and Rate Graph (RG). We illustrate a case in Fig. 1 to show the differences of the undirected graph and the three proposed directed graphs. In Fig. 1(a), the biker C cannot generally see the other agents, the person A shouldn't be impacted by the agents E and B, and the pedestrian B should consider the influence by the person D only. While in Fig. 1(b), the edges of the undirected graph cannot correctly express the relationship of these agents. For VG in Fig. 1(c), the view information of all agents is extracted by their positions and moving directions. Please note that agents of different types should be with divergent view ranges. While in DG in Fig. 1(d), when agents in the view of an agent, the directed edges will be constructed according to their moving directions. Moreover, agents' moving rates are also considered to set the weights of the directed edges in RG, as shown in Fig. 1(e). By resorting to the three directed graph topologies, the social interactions among agents can be well described by the raw information.

Based on the above directed graphs, we construct the social interactions of multi-type agents including pedestrians, bikes, motorcycles, and various vehicles. Further, we propose a novel

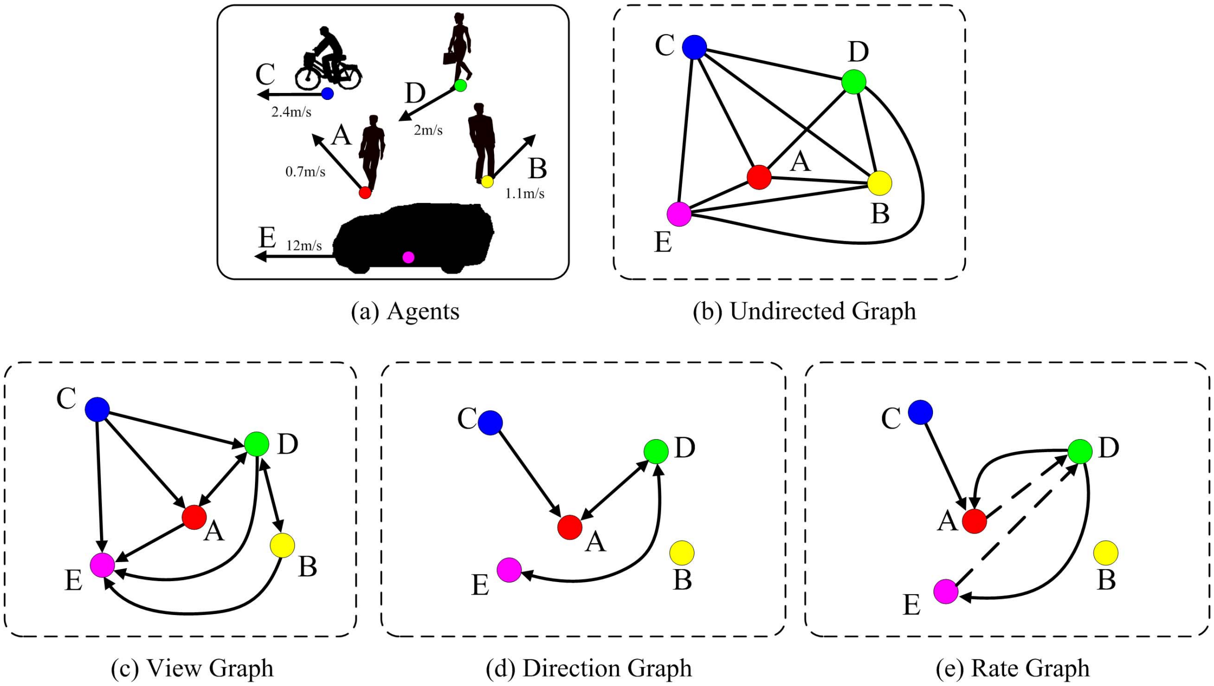


Fig. 1. The differences between the undirected graph and the three novel directed graphs for agents' relationship, where the points and lines with arrows indicate the positions and moving directions of the agents in (a), respectively, and the undirected or directed edges represent the impacts of the agents. In (c)–(e), a directed edge indicates an influence from a start node to a finish node; a line with two arrows indicates that the two nodes have influence of the same intensity, while two different lines (dash and solid) for a couple of nodes represent divergent influence intensities.

loss function to generate diverse multimodal trajectories. Our main contributions are summarized below.

- We design three directed graph topologies VG, DG, and RG, as shown in Fig. 1, to encode the asymmetric social interactions between agents. Different from existing methods, our proposed directed graphs extract the raw information among agents and describe three diverse features of the social interactions, which are more intuitive and effective for characterizing the asymmetric properties of trajectories.
- We propose a novel loss function based on Cauchy distribution for generating diverse positions by utilizing the fat tail property. By comparing with the control group using Gaussian distribution, the efficiency of the proposed loss function has been verified.
- Based on two widely used datasets of the trajectory prediction, the superiority of our proposed method has been experimentally demonstrated, and our proposed method beats all the compared methods and achieves over 8% and 31% performance improvement at least on predicting pedestrian trajectories, for the average and final predicted errors, respectively. Moreover, to predict trajectories of multi-type agents, our proposed model also improves about 56% and 57% performance, according to the average and final predicted errors, when compared with a GCN model proposed recently, which adopts an undirected graph topology.

The remainder of this paper is organized as follows. Section II provides the details of the proposed model, while Section III reports the experimental results and analyses. Finally, Section IV presents the conclusions.

II. Directed Graph Convolutional Neural Network for Trajectory Prediction

In this section, we firstly introduce the problem definitions of trajectory prediction, and then we devise three directed graph topologies for characterizing the social interactions between agents. In the end, the network framework and the loss function based on Cauchy distribution are proposed.

A. Definitions of Trajectory Prediction

Given observed positions of N agents over a fixed duration in the past, the aim of the trajectory prediction is to learn the potential moving rules of the agents, and predict their the future trajectories for the q upcoming time steps. Define

$$[u_1^i, \ldots, u_o^i, u_{o+1}^i, \ldots, u_{o+q}^i], \tag{1}$$

as a sequence of true positions of the i-th agent, where $u_t^i = (x_t^i, y_t^i)$ denotes the true position at the t-th time step.

Models for trajectory prediction aim to predict a future trajectory for the i-th agent ($i = 1, \ldots, N$), by inputting the observed positions, where the mapping can be defined as follows:

$$[u_1^i, \ldots, u_o^i] \Rightarrow [v_{o+1}^i, \ldots, v_{o+q}^i], \tag{2}$$

where v_t^i ($t = o + 1, \ldots, o + q$) indicates the t-th predicted position.

B. Directed Graphs

As an effective representation technique for data with irregular structures, graph representation is a powerful and direct model for modeling the complex social behaviors between agents. Recently, Social-STGCNN [37] has pioneered an undirected GCN model for pedestrian trajectory prediction, in which the symmetric influences of the undirected graphs are adopted to describe the relationship between each two pedestrians. However, the assumption of symmetric relationship does not hold in reality. A persuasive example is shown in Fig. 1, where the connections between agents should be highly asymmetric, and the undirected graph fails to construct such asymmetric impacts. As a consequence, we propose a directed GCN tailored for trajectory prediction. Different from existing graph-based approaches, we model the interactions in a scene as directed graphs by encoding the information of views, walking directions, and walking rates of agents. As will be shown later that our framework is more intuitive and effective to reveal the asymmetric mutual influences between agents. We first introduce the general definitions of our directed graphs.

Denote $G_t = (U_t, E_t)$ as a directed graph topology to describe the relationships among agents in the t-th time step of a scene, where $U_t = \{u_t^i | i = 1, 2, \ldots, N\}$ indicates a set of N nodes defined in (1), and $E_t = \{e_t^{ij} | i, j = 1, 2, \ldots, N\}$ includes a series of the weighted edges, where $e_t^{ij} = [0, 1]$. In our graph, there is a direction from the node u_t^j to the node u_t^i, only if $e_t^{ij} = (0, 1]$. A larger value of e_t^{ij} indicates the impact strength is stronger. In the remaining, we simplifies the above graph and its elements as $G = (U, E)$, u^i, and e^{ij}, by omitting the notation t for convenience.

1) View Graph (VG): Intuitively, the motion behavior of an agent is significantly impacted by his/her view. For example, we always pay attention to the agents in our views. As shown in Fig. 1(a) and (c), the biker C cannot see any agents, and thus he would not be impacted by the other agents. However, since he appears in the view field of the agents A, D, and E, his behavior could impact their future motions. Inspired by this prior, we construct VG according to the horizontal views of agents. It is worth noting that we assume that agents are walking forward and agents' view fields are equally divided by their motion directions. For the sake of simplicity, we set the view range for a pedestrian/non-motor vehicle as π and for a motor vehicle as 2π. As a result, in Fig. 1(c), we can find that the pedestrian A is not influenced by B because B is out of the view range of A, and the vehicle E is always impacted by all other agents.

Define VG as $G_v = (U, E_v)$, the connection will be built from u_j (start node) to u_i (finish node) if the j-th agent is in the view filed of the i-th agent, otherwise, the impact will be ignored. According to this rule, the elements in $E_v = \{e_v^{ij} | i, j = 1, 2, \ldots, N\}$ can be defined as follows:

$$e_v^{ij} = \begin{cases} \dfrac{1}{||u^i - u^j|| + 1}, & if \ (acos(\alpha) < \dfrac{\pi}{2} \wedge i \in n) \vee (i \in m) \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where

$$\alpha = \frac{d^i \cdot d^{ij}}{||d^i|| \times ||d^{ij}||}, \tag{4}$$

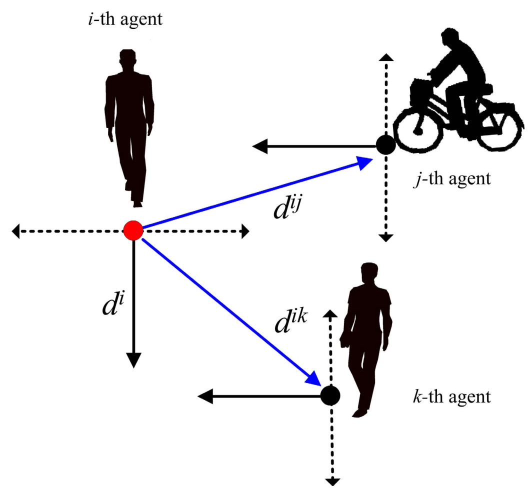


Fig. 2. Illustration of the view impacts, where the circles, dashed lines, and solid lines indicate the positions, the boundaries of the views, and moving directions, respectively, and the lines painted by blue are the direction vectors from the i-th agent to other agents.

$d^i = u_t^i - u_{t-1}^i$ denotes the moving direction of the i-th agent, and d^{ij} is a direction vector from the i-th agent to the j-th agent, $acos$ is the inverse function of cos, and m and n collect the indexes of all motor vehicles and other agents, respectively.

We further illustrate the construction of VG in Fig. 2, where we can see that the i-th agent should pay attention to the k-th agent due to the angle between d^i and d^{ik} is smaller than $\pi/2$. However, the i-th agent would be not impacted by the j-th agent, considering that the j-th agent is out of the i-th agent's view.

2) Direction Graph (DG): VG only exploits the position information of agents. In a crowded scenario, we also need to pay some attention to those agents who may have a potential collision with us. In this paper, we propose DG to characterize such influence by resorting to the moving directions of agents. If two agents' moving directions have a crossover, we could think that they have a potential collision risk. As illustrated in Fig. 1(d), there is a possibility of collision between the pedestrians A and D. According to this rule, denote DG as $G_d = (U, E_d)$, and we can define the elements in $E_d = \{e_d^{ij} | i, j = 1, 2, \ldots, N\}$ as follows:

$$e_d^{ij} = \begin{cases} \dfrac{1}{||u^i - u^j|| + 1}, & \text{if } \eta \text{ exists} \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where $\eta = (x, y)$ is an intersection of moving directions of two agents and obeys the following constraints

$$\frac{y_t^i - y_{t-1}^i}{x_t^i - x_{t-1}^i} \neq \frac{y_t^j - y_{t-1}^j}{x_t^j - x_{t-1}^j}, \tag{6}$$

$$||\eta - u_{t-1}^i|| > ||\eta - u_t^i|| \text{ and } ||\eta - u_{t-1}^j|| > ||\eta - u_t^j||. \tag{7}$$

Please note that (6) indicates the two lines are not parallel and the significance is unchanged even if a slope does not exist. If η exists, it will appear in front of the positions of two agents at the t time step. As illustrated in Fig. 3, we can observe that the moving directions of the i-th and k-th agents

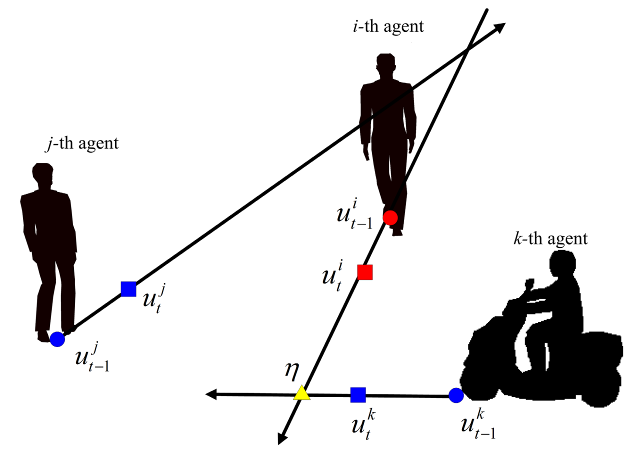


Fig. 3. Illustration of the direction influences, where the lines represent the moving directions, the circles and squares indicate the positions, and the triangle denotes an intersection η subjected to (6) and (7).

could cause a collision, since the intersection of their moving directions occurs in front of them. On the contrary, the i-th and j-th agents will not encounter a collision, because the intersection is in the back of the i-th agent and doesn't meet the constraint (7).

3) Rate Graph (RG): VG and RG model the view and moving direction features of agents, respectively. However, the moving rate of each agent is also an important factor to influence others' future trajectories. Generally, one agent with a faster motion rate will attract more attention from surrounding agents. In this paper, we propose RG to exploit this prior of agents. Specifically, one agent would be impacted by the other's rates, only when they are connected in the VG and DG. According to this rule, Define RG as $G_r = (U, E_r)$ and the elements in $E_r = \{e_r^{ij}|i, j = 1, 2, \ldots, N\}$ is defined as follows:

$$e_r^{ij} = \begin{cases} tanh(||r^j||), & \text{if } e_d^{ij} \neq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

where $r^j = u_t^j - u_{t-1}^j$ denotes the moving rate of the j-th agent.

As demonstrated in Fig. 1(e), though the pedestrian A would pay attention to the moving rate of the biker C, C would not be impacted due to the view field. The persons B and D have not the impact of the moving rate each other because of a low collision risk. Note also the influence strengths of the moving rates are often not equivalent, as shown in Fig. 4, as an agent with a higher moving rate should have a greater impact on others generally.

C. Directed Graph Convolutional Neural Network

Based on the above three directed graph topologies, in the subsection, we present the framework of the directed graph convolutional neural network for trajectory prediction. The framework of our proposed method is shown in Fig. 5, which mainly includes three components, i.e., Graph Fusion, Temporal Convolution, and Directed Graph Convolution.

Graph Fusion: Though VG, DR, and RG can characterize the different priors of social interactions, it is challenging to

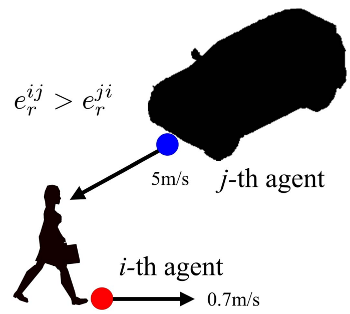


Fig. 4. Illustration of the rate impacts, where the lines and circles represent the moving directions and positions, respectively.

simultaneously exploit the information embedded on different graphs. To address this issue, we propose to fuse VG, DR, and RG into a unified directed graph through an Multilayer perceptron (MLP). Assume that there are N agents in the t-th time step, we can construct three weighted adjacent matrices containing all the edges of VG, DR, and RG, according to (3), (5), and (8), respectively. We then stack them to a tensor of $N \times N \times 3$, and use an MLP to adaptively fuse the weighted edges. We finally obtain a unified directed graph with the weighted adjacent matrix $E_f \in \mathcal{R}^{N \times N \times 1}$. The fusion process can be formalized as follows:

$$E_f = MLP(E_v, E_d, E_r), \tag{9}$$

where (9) is formed by three fully connection layers and $Tanh$ is employed as the activation function. The fused directed graph FG can be defined as $G_f = (\hat{U}, E_f)$, where $\hat{U}$ will be detailed in (10). For the efficiency, we design our method to simultaneously predict q future positions of all agents in a scenario, by inputting their o observed positions. As a result, the input size and the output size of (9) are $N \times N \times 3o$ and $N \times N \times o$, respectively.

Temporal Convolution: In addition to the spatial information of agents, the temporal information of trajectories is also important for the future trajectory prediction. In this work, we adopt a 2D convolution as reported in [38], to extract the temporal information of the observed trajectories. Assume that there are N agents with o observed time steps. The temporal information of all the agents can be extracted as follow:

$$\hat{U} = Conv_t(U), \tag{10}$$

where $U \in \mathcal{R}^{2 \times N \times o}$ represents the trajectories of N agents over the observed o time steps. In our experiment, we use one 1×3 convolutional kernel to extract the temporal features of trajectories.

Directed Graph Convolution: In this work, we propose a directed graph convolutional network to exploit both the spatial and temporal information of agents' trajectories. According to [39], a directed graph topology can be directly addressed by

$$H^{\ell+1} = \delta(E_f H^{\ell} W^{\ell}), \tag{11}$$

where $PRelu$ is adopted as the activation function δ; W^{ℓ} is a matrix containing all the trainable parameters at the ℓ-th graph convolution layer, and $H^0 = \hat{U}$. Similar to the traditional

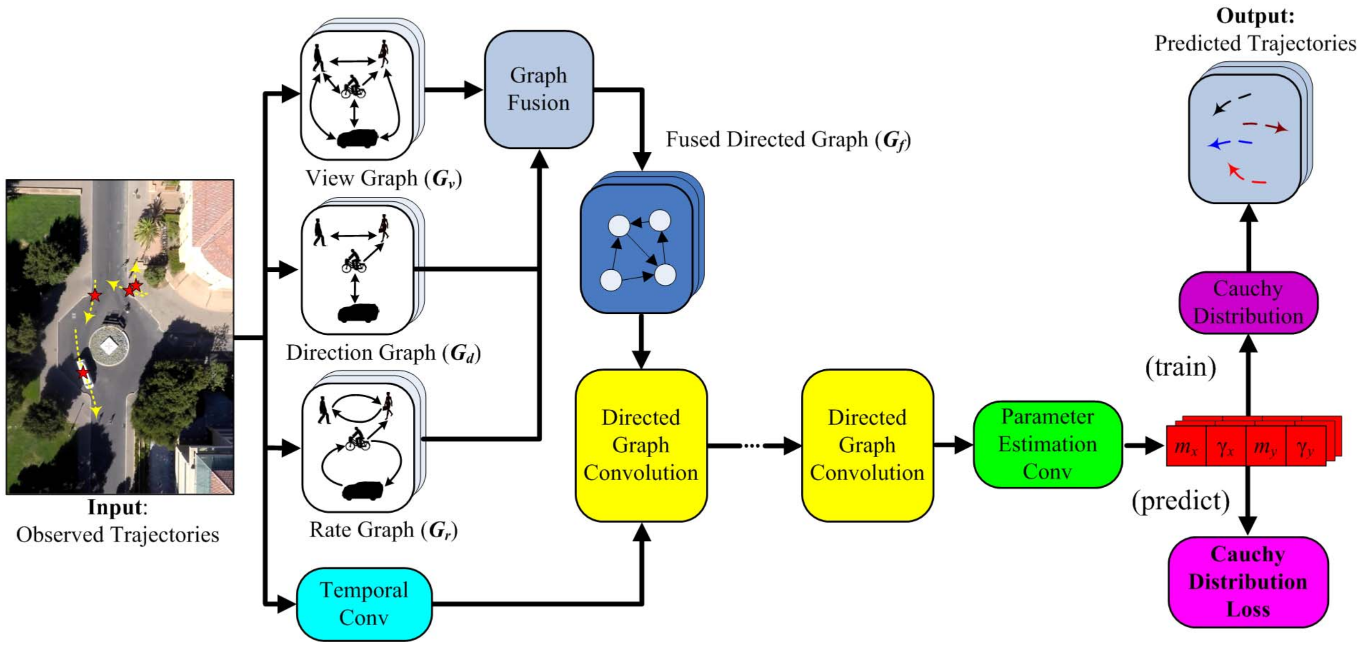


Fig. 5. The framework of VDRGCN, where m_x, γ_x, m_y, and γ_y indicate the position and scale parameters of Cauchy distributions, respectively, for the x and y element of a position.

undirected graph convolutional network, we also use a row normalization for the directed graph, and then (11) can be written as follows:

$$H^{\ell+1} = \delta(norm(E_f)H^{\ell}W^{\ell}), \tag{12}$$

where $norm$ is executed on each sub-set of E_f at each time step to normalize the social impacts for each agent, as follows:

$$e_f^{ij} = \frac{e_f^{ij}}{\sum_{j=1}^{N} e_f^{ij}}, \tag{13}$$

where $\{e_f^{ij}|i, j = 1, 2, \ldots, N\}$ is a sub-set of E_f at a certain time step.

D. Multimodal Trajectory Based on Cauchy Distribution

In order to generate multiple socially-acceptable trajectories with more diverse modes [25], we assume that the future multimodal trajectories obey to Cauchy distributions. Different from the previous work based on Gaussian distributions [37], we utilize the fat tail property of Cauchy distributions to reduce the redundancy of predicted positions. The related experimental analysis is presented in Section III.B.4. The probability density function f of a Cauchy distribution is defined as follows:

$$f(z; m, \gamma) = \frac{1}{\pi} \times \frac{\gamma}{(z-m)^2 + \gamma^2}, \tag{14}$$

where z is a random variable, and m and γ are the location and scale parameters of a Cauchy distribution, respectively. Denote that the x and y elements of a true position of an agent at the t-th time step respectively obey Cauchy distributions, namely

$$f_t(x; m_x, \gamma_x) = \frac{1}{\pi} \times \frac{\gamma_x}{(x-m_x)^2 + \gamma_x^2}, \tag{15}$$

$$f_t(y; m_y, \gamma_y) = \frac{1}{\pi} \times \frac{\gamma_y}{(y-m_y)^2 + \gamma_y^2}, \tag{16}$$

where $\{m_x, \gamma_x, m_y, \gamma_y\}$ is an estimated parameter set of Cauchy distributions. Based on the maximum likelihood estimation, the parameters can be evaluated by maximizing (17) as follows:

$$loss = \frac{\sum_{i=1}^{N} \sum_{t=o+1}^{o+q} log(f_t(x; m_x, \gamma_x) f_t(y; m_y, \gamma_y))}{N \times q}, \tag{17}$$

where N indicates the number of agents.

Thus, there are $4 \times N \times q$ estimated parameters and two consecutive convolution operations are adopted to learn them as follows:

$$\Theta = Conv_{c_2}^{k}(Conv_{c_1}(H)), \tag{18}$$

where H is the output of the last layer in Eq. (11) or Eq. (12), and $Conv_{c_1}$ is a 2D convolution operator, which takes input $H \in \mathcal{R}^{2\times N\times o}$, and outputs a tensor with the size $4 \times N \times q$, by channel reshaping and changing. $Conv_{c_2}^{k}$ is with k layer convolution operators applied to generate the estimated parameters for all the agents in q predicted time steps. The above convolution operators are with the kernel size 1×3.

To predict the future multimodal paths for N agents at q time steps, Θ achieved by (18) is adopted for $2 \times N \times q$ Cauchy distributions to generate multiple socially-acceptable trajectories.

III. Experimental Results

We implement the proposed model using the PyTorch framework with an NVIDIA RTX 3090 GPU.

A. Performance Metrics

Following prior works, we use two error metrics widely adopted to evaluate the performance of the proposed method and compared models.

B. Average Displacement Error (ADE)

It is calculated by the average Euclidean distance between the predicted trajectory and the true one for each agent over all predicted time steps, and a smaller value indicates better performance. ADE is defined as follows:

$$ADE = \frac{\sum_{i=1}^{N}\sum_{t=o+1}^{o+q}||v_i^t - u_i^t||}{N \times q}, \tag{19}$$

C. Final Displacement Error (FDE)

It is computed by the average Euclidean distance between the predicted trajectory and the true one at the final position of each agent, and a smaller value indicates a better result. The definition of FDE is given as follows:

$$FDE = \frac{\sum_{i=1}^{N}||v_i^{o+q} - u_i^{o+q}||}{N}, \tag{20}$$

D. Pedestrian Trajectory Prediction

In this subsection, we evaluate our model in pedestrian trajectory prediction with complex interactions.

1) Dataset and Experimental Setting: Our proposed model is verified on two widely used pedestrian trajectory datasets, i.e., ETH [40] and UCY [41]. The two datasets contain five scenarios: ETH, HOTEL, UNIV, ZARA1, and ZARA2. There are totally 1536 pedestrians containing challenging social behaviors.

We compare our proposed model with two well-known baselines and nine models proposed recently:

1) Social-LSTM [15]: A neural network based pedestrian trajectory prediction algorithm, which adopts an LSTM model and a social pooling, for learning the features of the sequences and social behaviors of pedestrians, respectively.
2) Social-GAN [13]: A GAN based method for generating the multimodal pedestrian trajectories.
3) PIF [42]: A multi-task LSTM model to utilize the visual and interactive features.
4) SoPhie [43]: A model adopts an attentive GAN to consider the physical constraints and social attentions.
5) SR-LSTM [17]: A state refinement method for extracting the social features of pedestrian trajectories.
6) STSGN [34]: A graph attention based an LSTM method for modeling the social interactions of pedestrians.
7) CGNS [27]: A conditional generative network based a GRU model for the multimodal pedestrian trajectories.
8) Social-BiGAT [35]: A bicycle-GAN with a GAT module for the multimodal paths and social interactions of pedestrians.
9) STGAT [14]: A spatial-temporal graph attention network for trajectory prediction.
10) TPNSTA [32]: A temporal pyramid network with a spatial-temporal attention for pedestrian trajectory prediction.
11) Social-STGCNN [37]: An undirected graph convolutional network for trajectory prediction.

Following previous works [13], [37], the time step numbers of the observed trajectories and predicted trajectories are set to 8 (3.2 seconds) and 12 (4.8 seconds), respectively. In training, we set the batch size to 64 for each scene, and the training epoch is set to 100 for ETH and 1000 for other scenes, as it is easy to over-fitting for ETH. The Adam optimizer with $1e-3$ learning rate is used, which is reduced by multiplying 0.9 after every 50 epochs. The layer numbers of $Conv_{c_2}$ in (18) and graph convolution are set to 10 and 1, respectively. Especially, the minimum ADE and FDE from 20 generated trajectories are selected as the evaluation metrics ADE and FDE for multimodal trajectory prediction models.

2) Model Comparison: In this subsection, we compare the proposed model VDRGCN with the above 11 recently proposed methods. Table I shows the results obtained by each model on the five scenarios, and we also report the average results for each method in the last column, where the values highlighted by red and blue represent the best and second best results, respectively. According to the results, we draw the following conclusions:

- It can be easily observed that our proposed model achieves the best or second best rank for each dataset. Moreover, the proposed method VDRGCN achieves the best performance in terms of the average ADE and FDE results, which gains over 8% / 31% performance improvement at least.
- When compared with Social-STGCNN based on GCN with an undirected graph topology, our algorithm outperforms Social-STGCNN on all the datasets and achieves about 20% and 33% performance improvement in terms of the average ADE and FDE results, respectively. This verifies that the proposed directed graphs characterizing the asymmetric influences between pedestrians are indeed helpful for modeling the social interactions.
- Even without employing the scene image as the side information, our method VDRGCN is still superior to those methods utilizing the scene features, such as [35], [42], [43]. This represents that the performance of VDRGCN could potentially be further improved by considering the scene context.

We conduct two cases to show how VDRGCN successfully models the social behaviors of pedestrians and predicts the trajectories, and we qualitatively compare the prediction results obtained by our method and the undirected graph based method Social-STGCNN as shown in Figs. 6-7.

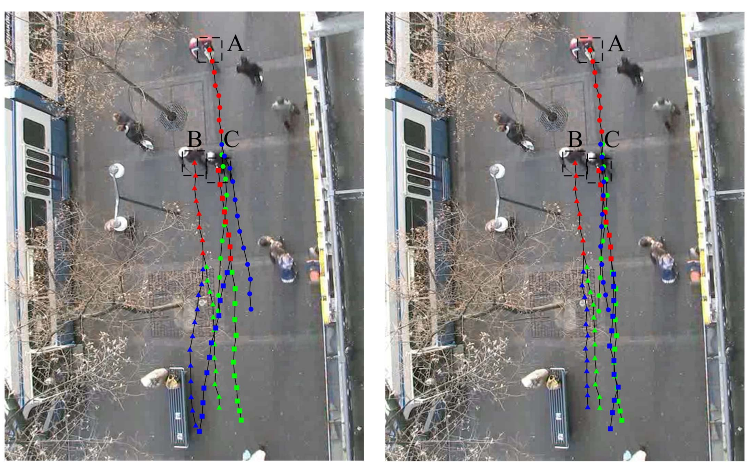
- Group walking: When two or more persons are walking with a group, they often keep a close distance, similar

TABLE I
THE RESULTS OF ADE / FDE OBTAINED BY DIFFERENT ALGORITHMS

	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
Social-LSTM	1.09 / 2.35	0.79 / 1.76	0.67 / 1.40	0.47 / 1.00	0.56 / 1.17	0.72 / 1.54
Social-GAN	0.68 / 1.26	0.47 / 1.01	0.56 / 1.18	0.34 / 0.69	0.31 / 0.64	0.47 / 0.97
SoPhie	0.70 / 1.43	0.76 / 1.67	0.54 / 1.24	0.30 / 0.63	0.38 / 0.78	0.54 / 1.15
CGNS	0.62 / 1.40	0.70 / 0.93	0.48 / 1.22	0.32 / 0.59	0.35 / 0.71	0.49 / 0.97
PIF	0.73 / 1.65	0.30 / 0.59	0.60 / 1.27	0.38 / 0.81	0.31 / 0.68	0.46 / 1.00
STSGN	0.75 / 1.63	0.63 / 1.01	0.48 / 1.08	0.30 / 0.65	0.26 / 0.57	0.48 / 0.99
STGAN	0.65 / 1.12	0.35 / 0.66	0.52 / 1.10	0.34 / 0.69	0.29 / 0.60	0.43 / 0.83
SR-LSTM	0.63 / 1.25	0.37 / 0.74	0.51 / 1.10	0.41 / 0.90	0.32 / 0.70	0.45 / 0.94
Social-BiGAT	0.69 / 1.29	0.49 / 1.01	0.55 / 1.32	0.30 / 0.62	0.36 / 0.75	0.48 / 1.00
TPNSTA	0.55 / 0.91	0.23 / 0.40	0.52 / 1.10	0.34 / 0.70	0.26 / 0.55	0.38 / 0.73
Social-STGCNN	0.64 / 1.11	0.49 / 0.85	0.44 / 0.79	0.34 / 0.53	0.30 / 0.48	0.44 / 0.75
VDRGCN	0.62 / 0.81	0.27 / 0.37	0.38 / 0.58	0.29 / 0.42	0.21 / 0.32	0.35 / 0.50

(a) Social-STGCNN (b) VDRGCN

Fig. 6. The predicted trajectories in ZARA2, where the shapes painted by red, green, and blue indicate the observed positions, future positions, and predicted positions based on ADE, respectively.



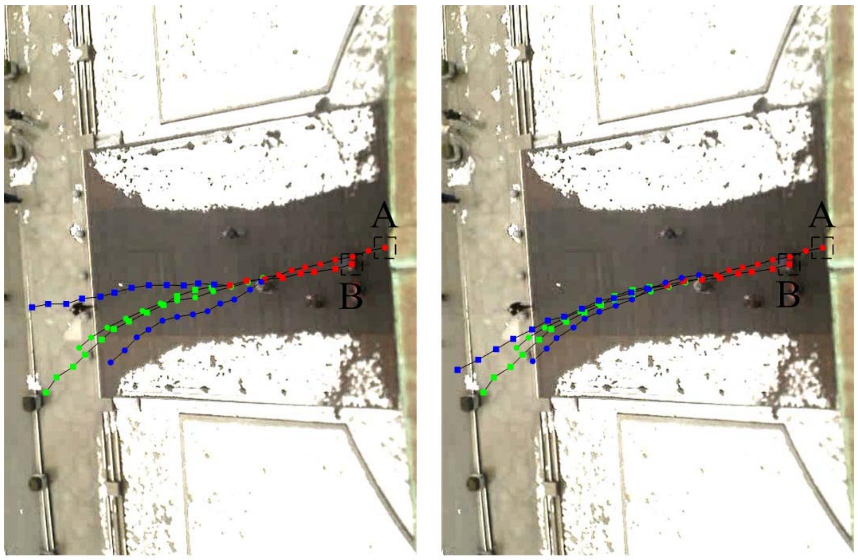
(a) Social-STGCNN (b) VDRGCN

Fig. 7. The predicted trajectories in HOTEL, where the shapes painted by red, green, and blue indicate the observed positions, future positions, and predicted positions based on ADE, respectively.

moving rates, and walking directions. In Fig. 6, the pedestrians A and B are walking in parallel. It can be observed from Fig. 6(a), though the predicted trajectories obtained by Social-STGCNN can keep a close distance for the group, there are large deviations from the ground-truth trajectories to the predicted paths. Compared with Social-STGCNN, the prediction of our model VDRGCN in Fig. 6(b) shows that A and B will keep walking in

TABLE II
THE ADE / FDE RESULTS OF THE GRAPH TOPOLOGIES

	G_u	G_v	G_d	G_r	G_f
ETH	0.71 / 1.21	0.66 / 1.07	0.68 / 1.15	0.68 / 1.20	0.62 / 0.81
HOTEL	0.35 / 0.47	0.28 / 0.40	0.29 / 0.41	0.29 / 0.42	0.27 / 0.37
UNIV	0.42 / 0.68	0.39 / 0.60	0.40 / 0.61	0.39 / 0.62	0.38 / 0.58
ZARA1	0.35 / 0.52	0.30 / 0.43	0.30 / 0.43	0.31 / 0.44	0.29 / 0.42
ZARA2	0.25 / 0.37	0.21 / 0.34	0.22 / 0.35	0.23 / 0.35	0.21 / 0.32



(a) G_u (b) G_f

Fig. 8. The visual comparison on G_u and G_f, where the shapes marked by red, green, and blue indicate observed positions, future positions, and predicted positions based on ADE, respectively.

parallel, and the predicted trajectories also have smaller deviations to the ground-truth trajectories. This demonstrates the superiority of modeling motion behavior using the directed graph topologies.

- Following people: When one person is following other persons, they often have similar moving rates and directions. Fig. 7 shows that the pedestrian A is following the group of B and C. Intuitively, pedestrians that are followed by others should be not easy to be impacted by the following persons. However, in Fig. 7(a), due to the symmetrical interaction of Social-STGCNN, the person C seems to be impacted by A, and this causes a potential collision risk between B and C. Instead, the predicted trajectories of VDRGCN in Fig. 7(b) show that the motion behavior of the group can be kept and the predicted path of the person A can well follow that of C. It indicates that the asymmetrical interactions are effective for the trajectory perdition.

3) Comparison on Graph Topologies: To verify the effectiveness of the proposed directed graph topologies, we also conduct an additional experiment by comparing different graphs. We consider the undirected graph G_u, view graph G_v, direction graph G_d, rate graph G_r, and fused graph G_f in this experiment. The numerical results and visual comparison are shown in Table II and Fig. 8, respectively.

- From Table II, we can learn that the performance of G_v, G_d, and G_r are always better than G_u on all the datasets. This demonstrates the importance of modeling the asymmetric influences. It is worth noting that G_v is slightly better than G_d and G_r and it indicates that the view fields of agents impact the motion behaviors most. When the three directed graphs are simultaneously incorporated, it achieves the best performance as shown in the last column in Table II. This shows that all the proposed directed graphs are helpful for the social interaction modeling.
- Fig. 8 visually shows the superiority of our proposed directed graphs, where the pedestrian A is following B. We can find that the predicted paths of the undirected graph G_u largely deviate from the true trajectories of A and B. It represents that the symmetrical interaction incorrectly describes the spatial relationship of the two persons. While in Fig. 8(b), the predicted positions based on G_f have small gaps to the ground truths and this demonstrates the superiority of the proposed directed graphs due to the asymmetric social interaction modeling.

4) Comparison for Losses: In this subsection, we experimentally verify the superiority of the proposed loss function based on Cauchy distribution for multimodal trajectories. For a fair comparison, we directly embed the loss function based

TABLE III
THE RESULTS OF ADE / FDE OBTAINED BY GAUSSIAN DISTRIBUTION AND CAUCHY DISTRIBUTION

	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
Gaussian Distribution	0.64 / 0.98	0.38 / 0.44	0.39 / 0.50	0.36 / 0.42	0.32 / 0.38	0.42 / 0.54
Cauchy Distribution	0.62 / 0.81	0.27 / 0.37	0.38 / 0.58	0.29 / 0.42	0.21 / 0.32	0.35 / 0.50

TABLE IV
THE RESULTS OF ADE / FDE OBTAINED BY TPNSTA, SOCIAL-STGCNN, AND VDRGCN

	BOOKSTORE	DEATHCIRCLE	GATES	NEXUS	AVG
TPNSTA	18.7 / 38.7	25.8 / 52.9	14.0 / 26.6	**8.1** / 14.7	16.7 / 33.2
Social-STGCNN	34.3 / 38.4	56.7 / 96.4	12.7 / 18.4	10.4 / 14.3	28.5 / 41.9
VDRGCN	**13.4 / 20.6**	**20.0 / 27.9**	**8.2 / 12.2**	8.3 / **10.7**	**12.5 / 17.9**

on Gaussian distribution adopted in [37] into our proposed model, and compare the results obtained by different losses. Table III reports the numerical results. Our proposed loss based on Cauchy distribution has 17% and 7% relative performance improvement in terms of the average ADE and FDE, respectively. Except FDE on UNIV, the predicted accuracies achieved by the loss based on Cauchy distribution are always higher or equal to those obtained by the loss based on Gaussian distribution. Compared with Gaussian distribution, Cauchy distribution with the fat tail property has a larger possibility to generate diverse positions for reducing ineffective redundancies.

E. Multi-Type Agent Trajectory Forecasting

We further verify our model in complex transportation environments for multi-type agents' trajectory prediction.

1) Dataset and Experimental Setting: The four scenarios from Stanford Drone Dataset [44], including the intersection, the roundabout, fork road, and entrances of parking lots are used. There are totally 4463 agents containing challenging social behaviors, and the types of agents have pedestrian, biker, skater, cart, car, and bus.

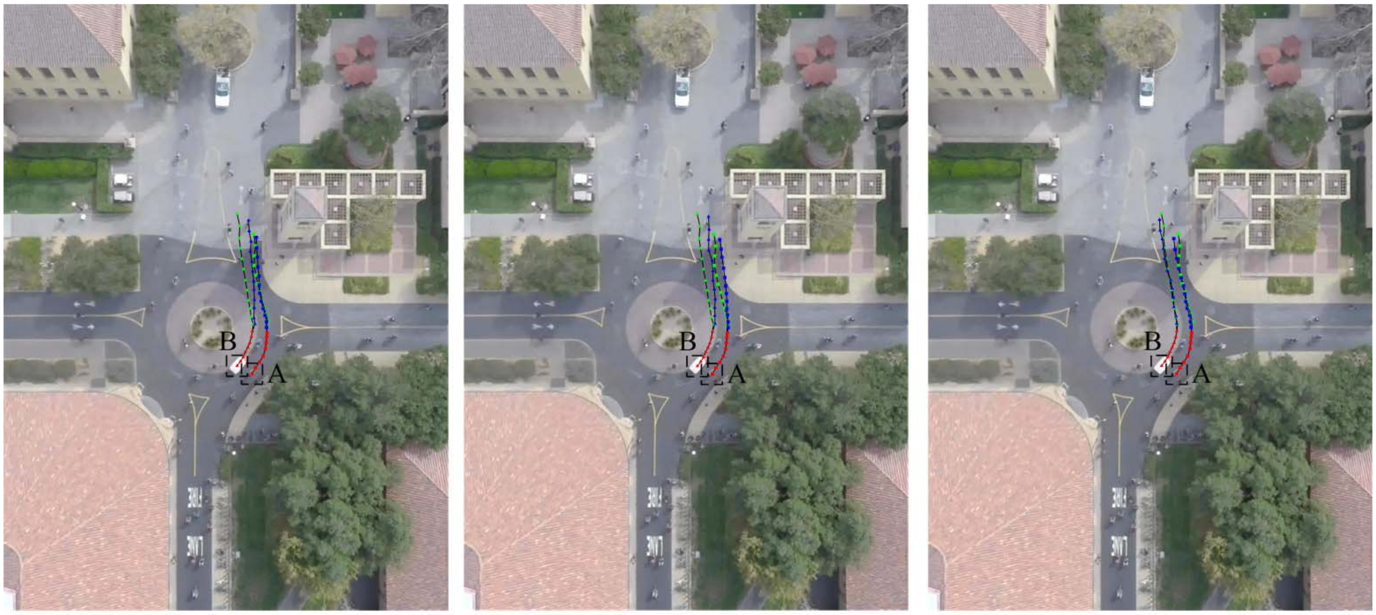
The true positions of each agent are recorded in every ten frames, as implemented in ETH [40] and UCY [41]. The time step numbers of the observed trajectories and predicted trajectories are set to 8 (about 2.6 seconds) and 12 (about 4 seconds), respectively. The used hyperparameters are the same to those in the previous experiments. We compare VDRGCN with TPNSTA and Social-STGCNN, due to their competitive performance in Table I.

2) Quantitative Analysis: Table IV shows the prediction results obtained by TPNSTA, Social-STGCNN, and VRDGCN on the four scenes, and we also report the average results for each method in the last column, where the **best** results are highlighted by **bold** fonts.

Due to the directed graphs for multi-type agents, VDRGCN gets a better performance than Social-STGCNN in all the scenes, which gains about 56% and 57% relative performance improvement in terms of the average ADE and FDE, respectively. Except ADE on NEXUS, VDRGCN always achieves better predicted accuracies when compared with TPNSTA. Moreover, the average ADE and FDE are also relatively improved about 25% and 46%, respectively. It indicates that the proposed directed graphs can also better characterize the asymmetric interactions for multi-type agents.

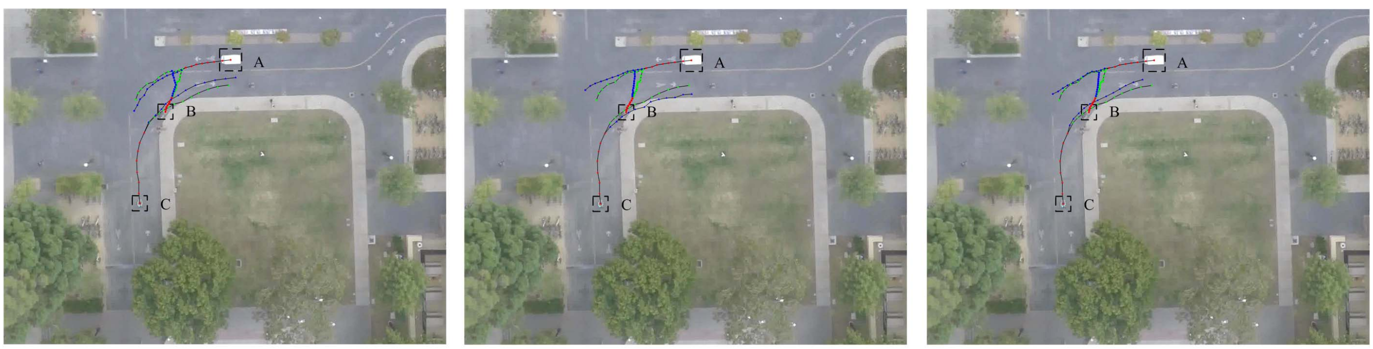
3) Qualitative Analysis: We conduct two cases to show how VDRGCN successfully models the social behaviors of multi-type agents and predicts the trajectories, and we qualitatively compare the prediction results obtained by TPNSTA, Social-STGCNN, and VDRGCN, as shown in Figs. 9-10.

- Parallel motion: When two or more agents are moving parallel in the road, it is important to keep a secure distance and prevent collusion. Different from parallel motions of persons, the prediction for the behaviors of multi-type agents is more difficult due to their different properties. In Fig. 9, the bike A and the car B are parallel passing through the roundabout. It can be obtained from Fig. 9(a) that the predicted trajectories of TPNSTA cannot keep a secure distance between A and B, and the predicted path of B is far away from the ground truths. From Fig. 9(b), though Social-STGCNN can keep a secure distance for the agents A and B, some predicted positions of A are out of the road. While in Fig. 9(c). the predicted trajectories achieved by VDRGCN can well keep a secure distance for A and B and are closer to the true paths than those of TPNSTA and Social-STGCNN. It indicates VDRGCN can well predict parallel motions of multi-type agents.
- Complex interactions: Fig. 10 shows that the car A and the bike C are opposite driving on two lanes of the road, and the person B is crossing the lanes. For TPNSTA, some predicted positions of B are far away from the true positions and the predicted path of C has a large gap to the ground truths. While in Fig. 10(b), Social-STGCNN is difficult to correctly predict the future trajectories, in which some predicted positions of C deviate from the lane and the predicted paths of A and B are not close to the true trajectories. Even for the complex interaction environment, our proposed VDRGCN can still appropriately address the asymmetric interactions between multi-type agents and well predict future trajectories, as shown in Fig. 10(c).



(a) TPNSTA (b) Social-STGCNN (c) VDRGCN

Fig. 9. The predicted trajectories in DEATHCIRCLE, where the shapes painted by red, green, and blue indicate the observed positions, future positions, and predicted positions based on ADE, respectively.



(a) TPNSTA (b) Social-STGCNN (c) VDRGCN

Fig. 10. The predicted trajectories in BOOKSTORE, where the shapes painted by red, green, and blue indicate the observed positions, future positions, and predicted positions based on ADE, respectively.

From the above cases, the superiority of our proposed model is also verified for multi-type agents in cooperative transportation systems. It demonstrates that the three directed graphs can effectively model the social interactions between multi-type agents and the proposed loss function is good for generating diverse multimodal trajectories.

IV. Conclusion

It is important and challenging to predict the future trajectories of the moving agents in advance for cooperative intelligent transportation systems. As the complex social influences between the moving agents are highly asymmetric, existing methods are difficult to model these cooperative interactions. To solve this issue, we propose a trajectory prediction algorithm based on a directed graph convolutional neural network, called VDRGCN, for cooperative intelligent transportation systems. By encoding different prior knowledge of social behaviors, we propose three directed graph topologies to exploit three types of social interactions between agents, i.e., View Graph, Direction Graph, and Rate Graph. The three directed graphs characterize the asymmetric influences of the view fields, walking directions, and moving rates of agents, respectively. We have shown that our proposed directed graph topologies are more intuitive for modeling the complex social interactions. Moreover, for characterizing the multimodal property of the future trajectories, we have also introduced a new loss function based on Cauchy distribution. The experimental results have been provided to verify the superiority of our proposed model over the existing ones.

In the future work, we will consider using scene features learned by a convolutional neural network as auxiliary information to interact with agents and roads for the improvement of robustness.

References

[1] S.-Y. Yu, A. V. Malawade, D. Muthirayan, P. P. Khargonekar, and M. A. A. Faruque, "Scene-graph augmented data-driven risk assessment of autonomous vehicle decisions," *IEEE Trans. Intell. Transp. Syst.*, early access, May 4, 2021, doi: 10.1109/TITS.2021.3074854.

[2] K. Yu, L. Lin, M. Alazab, L. Tan, and B. Gu, "Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4337–4347, Jul. 2021.

[3] Z. Ning *et al.*, "A cooperative quality-aware service access system for social Internet of Vehicles," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2506–2517, Aug. 2018.

[4] X. Liu, H. Song, and A. Liu, "Intelligent UAVs trajectory optimization from space-time for data collection in social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 853–864, Apr. 2021.

[5] M. Lv, Z. Hong, L. Chen, T. Chen, and S. Ji, "Temporal multi-graph convolutional network for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3337–3348, Jun. 2020.

[6] A. Alahi, V. Ramanathan, and L. Fei-Fei, "Socially-aware large-scale crowd forecasting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2203–2210.

[7] Y. Miao, C. Chen, L. Pan, Q.-L. Han, J. Zhang, and Y. Xiang, "Machine Learning–based cyber attacks targeting on controlled information: A survey," *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–36, Sep. 2022.

[8] N. Sun, J. Zhang, P. Rimba, S. Gao, Y. Xiang, and L. Y. Zhang, "Data-driven cybersecurity incident prediction: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1744–1772, 2nd Quart., 2018.

[9] Y. Li, J. Zhou, J. Chen, J. Tian, L. Dong, and X. Li, "Robust matrix factorization via minimum weighted error entropy criterion," *IEEE Trans. Comput. Social Syst.*, early access, Dec. 29, 2021, doi: 10.1109/TCSS.2021.3135654.

[10] B. Hu, G.-P. Gao, L.-L. He, X.-D. Cong, and J.-N. Zhao, "Bending and on-arm effects on a wearable antenna for 2.45 GHz body area network," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 378–381, 2016.

[11] Y. Li, J. Zhou, J. Tian, X. Zheng, and Y. Y. Tang, "Weighted error entropy-based information theoretic learning for robust subspace representation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 19, 2021, doi: 10.1109/TNNLS.2021.3056188.

[12] A. Scibior, V. Lioutas, D. Reda, P. Bateni, and F. Wood, "Imagining the road ahead: Multi-agent trajectory prediction via differentiable simulation," 2021, *arXiv:2104.11212*.

[13] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Oct. 2018, pp. 2255–2264.

[14] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "STGAT: Modeling spatial-temporal interactions for human trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6272–6281.

[15] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 961–971.

[16] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2375–2384.

[17] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, "SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2019, pp. 12085–12094.

[18] R. Liang, Y. Li, X. Li, Y. Tang, J. Zhou, and W. Zou, "Temporal pyramid network for pedestrian trajectory prediction with multi-supervision," 2020, *arXiv:2012.01884*.

[19] X. Shi, H. Qi, Y. Shen, G. Wu, and B. Yin, "A spatial-temporal attention approach for traffic prediction," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 8, pp. 4909–4918, Apr. 2020.

[20] Y. Xu, Z. Piao, and S. Gao, "Encoding crowd interaction with deep neural network for pedestrian trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5275–5284.

[21] Y. Xu, J. Yang, and S. Du, "CF-LSTM: Cascaded feature-based long short-term networks for predicting pedestrian trajectory," in *Proc. AAAI Conf. Art. Intel.*, 2020, vol. 34, no. 7, pp. 12541–12548.

[22] Z. Huang, J. Wang, L. Pi, X. Song, and L. Yang, "LSTM based trajectory prediction model for cyclist utilizing multiple interactions with environment," *Pattern Recognit.*, vol. 112, Feb. 2021, Art. no. 107800.

[23] N. Deo and M. M. Trivedi, "Trajectory forecasts in unknown environments conditioned on grid-based plans," 2020, *arXiv:2001.00735*.

[24] H. Cheng, W. Liao, M. Y. Yang, B. Rosenhahn, and M. Sester, "AMENet: Attentive maps encoder network for trajectory prediction," *ISPRS J. Photogramm. Remote Sens.*, vol. 172, pp. 253–266, Feb. 2021.

[25] C. Choi, J. H. Choi, J. Li, and S. Malla, "Shared cross-modal trajectory prediction for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Oct. 2021, pp. 244–253.

[26] X. Hu *et al.*, "Emotion-aware cognitive system in multi-channel cognitive radio ad hoc networks," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 180–187, Apr. 2018.

[27] J. Li, H. Ma, and M. Tomizuka, "Conditional generative neural system for probabilistic trajectory prediction," in *Proc. Int. Conf. Intel. Robot. Sys.*, 2019, pp. 6150–6156.

[28] J. Amirian, J.-B. Hayet, and J. Pettré, "Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Workshops*, Dec. 2019, pp. 2964–2972.

[29] T. van der Heiden, N. Shankar Nagaraja, C. Weiss, and E. Gavves, "Safe-Critic: Collision-aware trajectory prediction," 2019, *arXiv:1910.06673*.

[30] T. Zhao *et al.*, "Multi-agent tensor fusion for contextual trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12126–12134.

[31] X. Li *et al.*, "Vehicle trajectory prediction using generative adversarial network with temporal logic syntax tree features," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 3459–3466, Apr. 2021.

[32] Y. Li, R. Liang, W. Wei, W. Wang, J. Zhou, and X. Li, "Temporal pyramid network with spatial-temporal attention for pedestrian trajectory prediction," *IEEE Trans. Netw. Sci. Eng.*, early access, Mar. 9, 2021, doi: 10.1109/TNSE.2021.3065019.

[33] A. Vemula, K. Muelling, and J. Oh, "Social attention: Modeling attention in human crowds," in *Proc. IEEE Int. Conf. Robot. Autom.*, Oct. 2018, pp. 4601–4607.

[34] L. Zhang, Q. She, and P. Guo, "Stochastic trajectory prediction with social graph network," 2019, *arXiv:1907.10233*.

[35] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, S. Hamid Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks," 2019, *arXiv:1907.03395*.

[36] K. Guo, Y. Hu, Z. Qian, Y. Sun, J. Gao, and B. Yin, "Dynamic graph convolution network for traffic forecasting based on latent network of Laplace matrix estimation," *IEEE Trans. Intell. Transp. Syst.*, early access, Sep. 9, 2020, doi: 10.1109/TITS.2020.3019497.

[37] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14424–14432.

[38] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," 2018, *arXiv:1801.07455*.

[39] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[40] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 261–268.

[41] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 655–664, 2007.

[42] J. Liang, L. Jiang, J. C. Niebles, A. G. Hauptmann, and L. Fei-Fei, "Peeking into the future: Predicting future person activities and locations in videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Mar. 2019, pp. 5725–5734.

[43] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1349–1358.

[44] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 549–565.

Yuchao Su (Student Member, IEEE) received the M.S. degree from Shenzhen University, China, in 2019, where he is currently pursuing the Ph.D. degree. His research interests include trajectory prediction, artificial neural networks, and evolutionary computation.

Jie Du (Member, IEEE) received the B.S. degree in computer science and technology from Shihezi University, Shihezi, China, in 2013, and the Ph.D. degree from the Department of Computer and Information Science, University of Macau, Macau, China, in 2019. She is currently an Assistant Professor with the Health Science Center, School of Biomedical Engineering, Shenzhen University, Shenzhen, China. Her research interests include machine learning methods and medical image processing.

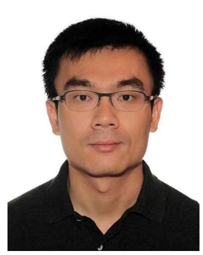
Yuanman Li (Member, IEEE) received the B.Eng. degree in software engineering from Chongqing University, Chongqing, China, in 2012, and the Ph.D. degree in computer science from the University of Macau, Macau, in 2018. From 2018 to 2019, he was a Post-Doctoral Fellow with the State Key Laboratory of Internet of Things for Smart City, University of Macau. He is currently an Assistant Professor with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China. His current research interests include data representation, multimedia security and forensics, computer vision, and machine learning.

Xia Li (Member, IEEE) received the B.S. and M.S. degrees in electronic engineering and signal and information processing (SIP) from Xidian University, in 1989 and 1992, respectively, and the Ph.D. degree from the Department of Information Engineering, The Chinese University of Hong Kong, in 1997. Currently, she is a member of the Guangdong Key Laboratory of Intelligent Information Processing. Her research interests include intelligent computing and its applications, image processing, and pattern recognition.

Rongqin Liang (Student Member, IEEE) received the B.Eng. degree in communication engineering from Wuyi University, Guangdong, China, in 2018, and the M.S. degree from the College of Electronics and Information Engineering, Shenzhen University, China, in 2021, where he is currently pursuing the Ph.D. degree. His research interests include computer vision and deep learning.

Zhongyun Hua (Member, IEEE) received the B.S. degree from Chongqing University, Chongqing, China, in 2011, and the M.S. and Ph.D. degrees from the University of Macau, Macau, China, in 2013 and 2016, respectively, all in software engineering.

He is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China. His research interests include chaotic systems, chaos-based applications, and multimedia security. He has published more than 50 papers on the subject, receiving more than 2700 citations.



Jiantao Zhou (Senior Member, IEEE) received the B.Eng. degree from the Department of Electronic Engineering, Dalian University of Technology, in 2002, the M.Phil. degree from the Department of Radio Engineering, Southeast University, in 2005, and the Ph.D. degree from the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, in 2009. He held various research positions with the University of Illinois at Urbana-Champaign, The Hong Kong University of Science and Technology, and McMaster University. He is an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, and also the Interim Head of the newly established Centre for Artificial Intelligence and Robotics. He holds four granted U.S. patents and two granted Chinese patents. His research interests include multimedia security and forensics, multimedia signal processing, artificial intelligence, and big data. He has coauthored two papers that received the Best Paper Award at the IEEE Pacific-Rim Conference on Multimedia in 2007 and the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2016. He is serving as the Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON MULTIMEDIA.