SSAGCN: Social Soft Attention Graph Convolution Network for Pedestrian Trajectory Prediction

Pei Lv, Wentong Wang, Yunxin Wang, Yuzhen Zhang, Mingliang Xu, Fellow, IEEE, and Changsheng Xu, Member, IEEE

Abstract—Pedestrian trajectory prediction is an important technique of autonomous driving. In order to accurately predict the reasonable future trajectory of pedestrians, it is inevitable to consider social interactions among pedestrians and the influence of surrounding scene simultaneously, which can fully represent the complex behavior information and ensure the rationality of predicted trajectories obeyed realistic rules. In this article, we propose a new prediction model named social soft attention graph convolution network (SSAGCN), which aims to simultaneously handle social interactions among pedestrians and scene interactions between pedestrians and environments. In detail, when modeling social interaction, we propose a new social soft attention function, which fully considers various interaction factors among pedestrians. Also, it can distinguish the influence of pedestrians around the agent based on different factors under various situations. For the scene interaction, we propose one new sequential scene sharing mechanism. The influence of the scene on one agent at each moment can be shared with other neighbors through social soft attention; therefore, the influence of the scene is expanded both in spatial and temporal dimensions. With the help of these improvements, we successfully obtain socially and physically acceptable predicted trajectories. The experiments on public available datasets prove the effectiveness of SSAGCN and have achieved state-of-the-art results. The project code is available at https://github.com/WW-Tong/ssagcn_for_path_prediction.

Index Terms—Graph convolutional network (GCN), social and scene interactions, social soft attention, trajectory prediction.

I. INTRODUCTION

PREDICTING the future trajectory of pedestrians in video is one significant task in autonomous driving [1], [2], [3], [4], [5], [6], which attracts plenty of scientists and engineers in different research fields nowadays. The successful prediction of the future trajectory of pedestrians is closely related to the accurate modeling of influence factors of the surrounding pedestrians and the physical scene. Moreover, trajectory prediction is a key component of autonomous driving system. To achieve the safe and flexible driving in complex scenario, essential requirements for trajectory prediction in terms of speed, accuracy, and generalization, are raised in [7] and [8]. Therefore, it remains a great challenging task due to the diversity and complexity of interactions.

Early works aggregate the hidden state of the recurrent neural network to model pedestrian interactions according to the location information [1], [9], [10], [11], [12], [13]. Afterward, the attention mechanism [14] is introduced to calculate the degree of the influence of different pedestrians on other agents [15], [16], [17]. Recently, graph structure is involved to represent the topological relationship of pedestrians [18], [19], [20], [21] and achieved competitive performance. However, the real situation is too complex to be modeled by a simple graph structure, where the social factors affecting the future trajectory of agent include relative speed, position, perspective, and so on. People tend to be influenced by their neighbors in the field of view, and the effect will increase with larger relative speed and closer distance. Unfortunately, most of the previous methods mainly consider the relative distance and lack of comprehensive consideration on the aforementioned factors.

However, there are two obvious shortcomings in this manner. First, the physical scene usually plays an important role in the whole process of pedestrian movement, not just in fixed period of time. Second, the interactions among pedestrians usually occur simultaneously with that between pedestrians and scene. The above observations motivate us to fully consider the continuous influence of physical scene in our method and share the impact of scene on the agents with the neighbors in time.

In view of the excellent performance of graph-based methods [18], [20], we continue this way in our work. Different from previous work, we further analyze more social factors that may influence the trajectory of pedestrians, such as relative speed, position, and angle of view. For example, pedestrians who meet from opposite directions will have more impact on each other, and the faster the relative speed, the larger the impact. Due to the limited field of view, pedestrians who leave in opposite directions will have less influence on each other, even if they are very close. Based on the above practical laws, we propose a new social soft attention function...
3) We propose a sequential scene attention sharing mechanism that aggregates the physical effects of real environment as part into graph node features and shares these effects through a unified social propagation. This model greatly enriches the semantic information of graph nodes, which helps to produce more physically acceptable predicted trajectories.

4) The performance of the proposed method is estimated on the ETH/UCY dataset, which outperforms previous approaches on the two metrics average displacement error (ADE) and final displacement error (FDE). More remarkably, compared with Y-net with the best performance, our method can reduce by 27.78% on ADE and 11.11% on FDE.

II. RELATED WORK

With the rapid development of deep learning and extensive attention paid to autonomous driving technology, the pedestrian trajectory prediction has also been greatly promoted. Since there is huge amount of related work in this research area, we mainly focus on the following aspects and demonstrate them in detail.

A. Social Interaction Modeling Among Pedestrians

Researchers have noticed from the very beginning that it is inevitable to consider the influence of the surrounding neighbors on one agent to predict its movement in the future. Helbing and Molnar [24] first proposed the concept of social force, which used the repulsive force and attraction between pedestrians to model social interaction. To calculate the impact of motionless crowds on pedestrian movement, Yi et al. [25] introduced personality attributes to classify pedestrians into different categories. Social-LSTM [11] was proposed to integrate the original vanilla long short-term memory (LSTM) network with time-step pooling mechanism, and this model successfully simulated the social interaction of pedestrians. Since the importance of social influence is mainly determined by the distance between pedestrians and their neighbors, Xue et al. [26] proposed to construct a circular occupancy map for each pedestrian to capture the impact of other pedestrians. Zhang et al. [13] proposed a weighted algorithm similar to the attention mechanism, to extract social influence information from neighbors’ current behavioral intentions to establish an interaction model. Sadeghian et al. [15] proposed one social attention mechanism, which models the interaction between pedestrians by calculating the attention of the LSTM hidden state of nearby pedestrians. In recent years, some work [18], [20], [27], [28], [29], [30], [31] began to use graph structures to model different types of social interactions. Huang et al. [18] proposed STGAT, which combined graph network and attention mechanism based on distance to share information among different pedestrians and assigned different weights to various nodes. Social-STGCNN (SSTGCNN) [20] used GCN with more rich spatial–temporal information to encode the interactions among pedestrians. For those methods [27], [28], [29], [30], [31], [32], they constructed various excellent spatial–temporal interaction models for group to cover these situations and use it to construct a weighted adjacency matrix to represent different degrees of influence among pedestrian nodes in the constructed spatial graph. In addition, we also propose a sequential scene attention sharing mechanism to model scene interaction, as shown in Fig. 1. The scene features at each moment rather than that at a certain fixed moment are embedded into the corresponding graph node in the constructed interaction graph. By calculating the influence of the scene in continuous moments, the effects of scene are first extended in time domain and then shared among pedestrians through social interactions to further simultaneously enforce the constraints of social interaction and scene interaction.

Moreover, the various social interactive factors and physical scene information are fully considered and simultaneously embedded into the unified graph node as never before. The graph convolutional network (GCN) [22] is utilized to process the node feature aggregation on the constructed graph, and the temporal convolutional neural networks (TCNs) [23] are used to generate the predicted trajectory. The contributions can be summarized as follows.

1) We propose a new trajectory prediction model, social soft attention graph convolution network (SSAGCN), which can deal with more comprehensive social interactions among pedestrians, and also explore a new sequential scene attention mechanism to expand the usage of scene information uncovered by previous work. Through fully considering the interactions among pedestrians and the impact of the scene on pedestrians, the performance of SSAGCN has reached the state-of-the-art results on public datasets.

2) We propose a social soft attention function to calculate the mutual influence of pedestrian nodes in one customized graph, which is constructed to depict the interaction between multiple objects both in spatial and temporal dimensions. This function fully considers the influence of relative speed, position, and angle of view on the future trajectories of pedestrians.
activity recognition, which have great inspiration for trajectory prediction tasks with the same typical spatial–temporal characteristics.

The above methods have achieved satisfactory prediction results through modeling the social interactions among pedestrians. However, they did not fully consider various influence factors mentioned before. In our method, we propose a new social soft attention function for graph convolution, which considers the influence of position, relative speed, and perspective factors comprehensively, to model the complex social interactions by calculating the degree of interaction between pedestrian nodes in a graph structure.

B. Scene Interaction Modeling Between Pedestrians and Environment

The scene plays a vital role in the trajectory prediction task. Recently, many methods [15], [16], [33], [34], [35], [36], [37] focused on how to extract context features from the scene to constrain the movement of agents. Kitani et al. [38] used the hidden variable Markov decision process to model the interaction between human and the scene and infer the passable area for pedestrians. Sadeghian et al. [34] considered the dependencies between the historical trajectories of agents and the spatial navigation environment and then presented an interpretable trajectory prediction framework. Similar to the above methods, Sadeghian et al. [15] further proposed the Sophie framework, which simultaneously modeled physical and social interactions by leveraging the history trajectory of agents and the scene context features. Moreover, some approaches [16], [35] processed the original images of the scene into the form of binary segmentation to distinguish the roads and obstacles. Li et al. [36], [37] proposed social attention and spatial topology module for image understanding and activity analysis, which can be applied in the field of trajectory prediction. Haddad et al. [16] took into account the interaction with static physical objects and dynamic pedestrians in the scene. They presented a new spatial–temporal graph based on the LSTM network, which could avoid the potential collision in crowded environments efficiently.

In our work, we use a physical scene attention mechanism to calculate the continuous effects of scenes on agents in sequential frames rather than specific ones. Moreover, in order to consider the influence of scene context and social interaction at the same time, we share the influence of scene at each moment with each adjacent node in the social interaction graph. Compared with previous methods, our approach expands the influence of the physical scene both in time and space.

C. Graph Neural Network

Graph neural networks (GNNs) [39] are powerful neural network architecture for machine learning on graphs. Through combining GNNs and convolutional neural network (CNN), the GCN [40], which allows for assigning various weights to different neighbors according to distances, has been widely used for various vision tasks. In recent years, Gao et al. [41] used GCN to handle vision tracking tasks. Sun et al. [42] used GCN to process social interaction in video trajectory prediction tasks. Yan et al. [22] utilized an improved model, STGCNN, to recognize human interactions. Mohamed et al. [20] formulated a kernel function to attach social attributes to STGCNN to predict pedestrian trajectory. However, these works solidify the degree of influence into the same value when modeling the influence of any two pedestrians on each other with graph structure, which is unreasonable in practical situations. In many cases, the two pedestrians do not have exactly the same influence on each other. The emergence of graph attention networks (GATs) [43], which brings attention to GNNs, solves this problem well. Subsequently, STGAT [18] and Social-Bigat [19] extended GAT in trajectory prediction; in particular, the latter appended physical scene information to the final prediction stage and enforced the effect from the physical scene on pedestrian in the last observation frame. Shi et al. [44] used a self-attention mechanism to calculate an asymmetric attention score matrix and applied it to sparse graph convolution.

Although the GAT-based approaches have made great progress in trajectory prediction, they require a huge amount of data to learn the differences between nodes in the graph and consume much more computing resources. Our method is more hand-designed and has faster speed than that of the GAT-based method. Due to the thorough consideration, our method can better distinguish different influences between various nodes.

III. SSAGCN

In this section, we aim to develop a prediction model to derive a set of possible future trajectories for pedestrians. In Section III-A, we explain the problem definition of pedestrian trajectory prediction. In Section III-B, we give a brief overview of the proposed model. In Section III-C, we introduce the details of the design of social soft attention function. In Section III-D, we demonstrate the sequential scene attention sharing mechanism. Finally, the method of trajectory generation is introduced in Section III-E.

A. Problem Definition

Similar to [10] and [45], we assume that there are $N$ pedestrians during one period of time $[1, T_{\text{pred}}]$. After preprocessing the trajectories of pedestrians in the video, the position of each pedestrian $i$ at each time step $t$ can be expressed as a pair of spatial coordinates $(x_i^t, y_i^t)$, where $t \in \{1, 2, 3, \ldots, T_{\text{pred}}\}$, $i \in \{1, 2, 3, \ldots, N\}$. The scene image is represented by $I_{\text{scene}}$. We take the coordinate sequences and the scene information of $I_{\text{scene}}$ during the period of time interval $[1, T_{\text{obs}}]$ as the input and predict the coordinate sequences in $[T_{\text{obs}+1}, T_{\text{pred}}]$.

B. Model Overview

As shown in Fig. 2, first, we process the coordinates $(x_i^t, y_i^t)$ into graph representation, with each node representing an agent. Then, we estimate the influence matrix among graph nodes using social soft attention function according to the relative position, speed, and direction of agents. Meanwhile, scene attention $C_i$ is calculated according to the
coordinates \((x^t_i, y^t_i)\) of agent and the features of the scene image \(I_{scene}\) at each past moment. The scene attention \(C^t_i\) will be further embedded into the graph node at each time step.

Through the implementation of GCN and social soft attention, scene information is shared among agents through social interaction. Finally, we input the graph sequence processed by GCN into the TCNs [23] to estimate the future trajectory of the agent. We will explain the details in the following.

C. Social Attention Graph Representation

1) Node Representation: We construct a sequence of graphs to represent the pedestrian trajectories. At time step \(t\), all pedestrians are connected to form a complete graph \(G_t = (V_t, E_t)\). \(V_t = \{v^t_i | \forall i \in \{1, \ldots, N\}\}\) is the set of vertices of \(G_t\), representing all pedestrians at time step \(t\). The initial value of \(v^t_i\) is the observed coordinate position \((x^t_i, y^t_i)\). \(E_t = \{e^t_{ij} | \forall i, j \in \{1, \ldots, N\}\}\) is the set of edges of \(G_t\). In our work, the edges contained in \(E_t\) are represented by an adjacency matrix \(A_t\) calculated according to the social soft attention function \(F_{ssa}(\cdot)\).

2) Social Soft Attention Function: Due to various movement behaviors existing in actual situation and being affected by different social factors, they cannot be modeled accurately with only one single interaction pattern. Thus, the interactions among pedestrians can be divided into three categories, as shown in Fig. 3: meeting from opposite directions, leaving in the opposite direction, and walking abreast. Although the current position of the agent is the same in the three cases, it is obvious that they have different interaction behaviors and corresponding trajectories. Most of the previous prediction methods used the relative position as the basis for social interaction, for example, according to position pooling or calculating attention weight according to distance; however, the simple position relationship could not accurately determine the social interaction in the case of \(a\) and \(c\). In fact, it is not determined by distance alone. Therefore, we design the social soft attention function to cover these three cases explicitly through the prior knowledge summarized in practice. In order to cover these typical situations, we propose one new social soft attention function \(F_{ssa}(\cdot)\) to calculate the element \(a^t_{ij}\) in \(A_t\) to represent the attention weight between pedestrians. The process is expressed as follows:

\[
a^t_{ij} = F_{ssa}(u^t_i, u^t_j, \cos \alpha^t, \cos \beta^t, l^t_{ij})
\]
As shown in Fig. 3, $u'_i$ indicates the speed vector of node $v'_i$ at time step $t$, which is obtained by $(x'_i - x'_{i-1}, y'_i - y'_{i-1})$. Similarly, $u'_j$ represents the speed vector of the node $v'_j$ at time step $t$. $\alpha$ and $\beta$ represent the angle between the speed vectors $u'_i$ of the node $v'_i$ and $u'_j$ of the node $v'_j$, respectively. $l'_{ij}$ represents the Euclidean distance between node $v'_i$ and node $v'_j$. $\theta$ is a hyperparameter, which represents the self-attention of one agent. When we implement the social soft attention function, the value of $\theta$ is set into the interval $[0.04, 0.16]$ depending on the maximum value of elements in the matrix $A_t$. Equation (2) is the normalization process to obtain the new adjacency matrix $\tilde{A}_t$.

In Fig. 3(a), two people are facing each other. $\alpha$ and $\beta$ are both acute angles. The output of $F_{ssa}()$ is larger than 0, which means that these two people have an influence on each other. In Fig. 3(b), the two people are leaving in opposite directions. $\alpha$ and $\beta$ are both obtuse angles. The output of $F_{ssa}()$ is limited to 0, which indicates that the two people do not influence each other. In Fig. 3(c), two people are walking abreast, where $\alpha$ is an acute angle and $\beta$ is an obtuse angle. If the speed of the yellow pedestrian is much higher than that of the blue one, there will be a risk of collision in the future, then the output of $F_{ssa}()$ is larger than 0. If the speed of yellow pedestrian is less than that of the blue one, there will be no collision risk, and the output of $F_{ssa}()$ is limited to 0. We put the distance $l'_{ij}$ in the denominator to reflect the magnitude of the risk of collision. When $\alpha$ and $\beta$ take different values, even 0° or 180°, our function still can distinguish different situations of pedestrian interaction so that the features of social interaction can be better extracted by GCN.

D. Sequential Scene Attention Sharing Mechanism

1) Extended Scene Attention: Since the movement behavior of pedestrians is constrained by physical scene in real situations, it is inevitable to consider social interactions and the influence of surrounding scene simultaneously to accurately predict the reasonable future trajectory of pedestrians. To fully leverage scene information, we utilize the pretrained CNN to extract the features of scene images. The method adopted in our work is similar to [16]. We use the pretrained VGG19 [15] as the backbone network, and the extracted features are represented by $V_{ph}$. Since the images from the dataset are captured by fixed cameras, we only need to calculate $V_{ph}$ for each dataset. In previous work [15], [18], they usually chose to calculate the scene attention $C^{T_{obs}}$ when $t = T_{obs}$. In order to extend the influence of the scene, we calculate the scene attention $C'_t$ at each moment $t \in \{1, 2, 3, \ldots, T_{obs}\}$ according to the feature $V_{ph}$ and the position $X'_t$ of agent $i$.

$$V_{ph} = \text{VGG19}(I_{scene}; W_{vgg19})$$

where $I_{scene}$ represent the images from datasets, $W_{vgg19}$ is the weight of pretraining network, $X'_t$ is the position $(x'_t, y'_t)$ of the agent $i$ at time step $t$, and $W_{att}$ contains the parameters of the scene attention. The SceneAtt used in our model is the same as in [16], and the structure is shown in Fig. 4, where FC represents the fully connected layer, MLP represents the multilayer perception, and "X" represents the product of tensors.

Then, we embed $C'_t$ with coordinate information to form new graph node feature $v'_i$ by the fully connected layer

$$v'_i = \phi(X'_i, C'_i; W_e)$$

where $\phi(\cdot)$ is the embedding layer and $W_e$ is the weight of embedding layer.

2) Graph Neural Network: With the above graph representation, we perform graph convolution operation at each time step $t$. The convolution operation of graph at single moment $t$ is defined in GCN [46] as follows:

$$\tilde{A}_t = A_t + E$$
$$\tilde{D}_t = D_t + E$$
$$V'_t = \sigma(\tilde{D}_t^{-\frac{1}{2}}\tilde{A}_t\tilde{D}_t^{-\frac{1}{2}}V_tW) = \sigma(\tilde{L}_tV_tW)$$

where $A_t$ is the adjacency matrix of the graph at time step $t$, $D_t$ is the degree matrix of the graph, and $E$ is the identity matrix; $V_t$ is a matrix composed of $v'_i$ as row vectors at the same time step. $\tilde{L}_t$ is a graph displacement operator used to aggregate adjacent nodes, which is the result of normalization of $A_t$.

For a single node, the graph convolution operation is

$$v''_i = \sigma \left( \sum_{v_j \in N(v'_i)} \tilde{L}_t[i, j](Wv'_j) \right)$$

where $N(v'_i)$ is the set of first-order neighbors of node $v'_i$ and $\tilde{L}_t[i, j]$ is the element in the $i$ row and $j$ column of matrix $\tilde{L}_t$. During the process of aggregating neighbor features of $v'_i$, $\tilde{L}_t[i, j]$ represents the weight of $v'_j$.
3) Scene Attention Sharing: The node features \( v_i^t \) involved in the graph convolution operation is embedded with \( C_i^t \) in (5). Therefore, in the operation of (9), \( C_i^t \) participates in the convolution operation as a part of \( v_i^t \). The aggregation and sharing process of \( C_i^t \) in the convolution operation of the graph can be obtained by substituting into (5)–(9), as shown in the following equation:

\[
C_i^t = \sigma \left( \sum_{C_j \in N(C_i)} \bar{L}_t(i, j)(W\phi(X_i^t, C_i^t; W_c)) \right).
\] (10)

The impact of scene on agents during this process is shared through social interactions. In Fig. 5, the direct effect of the barrier on agent A is \( C_1 \), and that of the barrier on agent B is \( C_2 \). In the sharing process, the effect of obstacles on A is transmitted to B through the social relationship between A and B. The impact of obstacles on Agent B is aggregated into \( C_2^t = C_2 + WW_cC_1 \).

In order to better distinguish the interaction between different pedestrians, we use \( F_{sna} \) to calculate matrix \( \bar{A}_t \) and replace \( L_t \) after normalization, so the operation of GCN in our model algorithm can be expressed as

\[
V_i^t = \sigma \left( N(\bar{A}_t)V_i^t W \right)
\] (11)

where \( \bar{A}_t \) is given by (2) and \( N(\bar{A}_t) \) means to normalize \( \bar{A}_t \).

E. Trajectory Generation

After dealing with social interaction and scene interaction, we choose TCNs to model the temporal dependence of social graph sequences \( V = \{ v_{i}^{t'} \mid \forall i \in \{1, \ldots, N\} \} \). Just like SSTGCNN, we treat the time dimension as feature channels input \( v_{i}^{t'} \) into TCNs. After the convolution operation of TCNs, we obtain the parameters of the 2-D Gaussian distribution about the sequence of coordinates during \([T_{obs}+1, T_{pred}] \)

\[
\left[ \mu_x^t, \mu_y^t, \sigma_x^t, \sigma_y^t, \rho^t \right] = \text{TCNs}(V_i^{t'}W_c)
\] (12)

where \( \mu_x^t \) and \( \mu_y^t \) are the mean of the coordinates and \( \sigma_x^t \) and \( \sigma_y^t \) are variance, \( \rho^t \) is the correlation coefficient between \( x \) and \( y \). \( W_c \) is the weight of TCNs. \( \tau \) is time steps in \([T_{obs}+1, T_{pred}] \). Then, we construct a 2-D Gaussian distribution \( \mathcal{N} \) by \([ \mu_x^t, \mu_y^t, \sigma_x^t, \sigma_y^t, \rho^t ] \), and the future coordinates can be obtained by sampling the distribution

\[
(x_i^t, y_i^t) \sim \mathcal{N}(\mu_x^t, \mu_y^t, \sigma_x^t, \sigma_y^t, \rho^t).
\] (13)

Loss Function: We use the ground-truth values \( x_i^t, y_i^t \) and the predicted parameters of Gaussian distribution to calculate the negative log-likelihood loss to guide the model training

\[
\mathcal{L} = -\sum_{t=T_{obs}+1}^{T_{pred}} \log(\mathbb{P}(x_i^t, y_i^t | \sigma_x^t, \sigma_y^t, \rho^t)).
\] (14)

IV. EXPERIMENT AND EVALUATION

We perform the evaluation experiments on widely used benchmark datasets and compare the results with other state-of-the-art methods. In detail, we use ETH [47] and UCY [48] datasets, which contain the following scenarios, ETH, HOTEL, UNIV, ZARA1, and ZARA2. The data attributes consist of frame number, pedestrian number, and 2-D position of trajectory coordinates. These trajectories are spaced 0.4 s apart from each other. Similar to Social-LSTM [10], we also input the trajectory of eight time steps (3.2 s) and predict the next 12 time steps (4.8 s). To verify that the proposed method is able to deal with various scenarios, we also conduct experiments on the Stanford Drone dataset (SDD) [49], which contains a large number of different scenes. The tracking coordinates in this dataset are measured in pixels. We use the same standard data segmentation settings as in [50].

A. Experiment

1) Evaluation Criteria: Following the strategy adopted by other baseline methods, we use the leave-one-out method to conduct the evaluation experiments, train on four datasets, and test on the remaining one. ADE and FDE are used as the standard metrics, which are defined as

\[
\text{ADE} = \frac{\sum_{n \in N} \sum_{t \in T_{obs}} \| \hat{Y}_n^t - Y_n^t \|_2}{N \times T_{pred}}
\] (15)

\[
\text{FDE} = \frac{\sum_{n \in N} \| \hat{Y}_n^{T_{pred}} - Y_n^{T_{pred}} \|_2}{N}
\] (16)

2) Implementation Details: The structure of GCN used in SSAGCN is similar to [20]. We use coordinate information in data processing to calculate the weighted adjacency matrix through the social soft attention function. When the weighted adjacency matrix is normalized, \( R \) is set to 0.10. The dimension of scene attention is 8, and the coordinate dimension is 2.
| TABLE I |

**QUANTITATIVE COMPARISON WITH BASELINE METHODS ON ETH AND UCY DATASETS. ALL THESE METHODS ARE USED TO PREDICT THE TRAJECTORY OF THE FUTURE 12 FRAMES BASED ON THE PREVIOUS EIGHT FRAMES. THE EVALUATION METRICS USED IN THIS TABLE ARE ADE AND FDE. K REPRESENTS THE NUMBER OF PREDICTED TRAJECTORIES. THE DATA IN THIS TABLE ARE FROM THE RESULTS REPORTED IN THEIR WORK. SSAGCN-W/o-sen, SSAGCN-W/o-seq, AND SSAGCN-W/o-ssa ARE OUR MODELS WITH DIFFERENT EXPERIMENTAL SETTINGS. K = 1 MEANS ONE GENERATED TRAJECTORY AND K = 20 MEANS 20 GENERATED TRAJECTORIES. IT IS TO BE MENTIONED THAT, SOME METHODS ONLY REPORT THEIR RESULTS FOR K = 1 AND OTHERS REPORT RESULTS FOR K = 20. (a) ADE/FDE K = 1. (b) ADE/FDE K = 20 |

| Method       | ETH  | HOTEL | UNIV | ZARA1 | ZARA2 | AVG  |
|--------------|------|-------|------|-------|-------|------|
| S-LSTM [10]  | 1.09 | 2.35  | 0.79 | 1.76  | 0.67  | 1.40 | 0.47 | 1.00 | 0.56 | 1.17 | 0.72 | 1.54 |
| SR-LSTM [13] | 0.63 | 1.25  | 0.37 | 0.74  | 0.51  | 1.10 | 0.41 | 0.90 | 0.32 | 0.70 | 0.45 | 0.94 |
| STGAT [18]   | 0.75 | 1.55  | 0.43 | 0.88  | 0.31 | 0.66 | 0.25 | 0.53 | 0.21 | 0.44 | 0.39 | 0.81 |
| GAT [19]     | 0.68 | 1.29  | 0.68 | 1.40  | 0.57 | 1.29 | 0.29 | 0.60 | 0.37 | 0.75 | 0.52 | 1.07 |
| Social-BiGAT [19] | -   | -     | -    | -     | -    | -   | -   | -   | -   | -   | 0.61 | 1.33 |
| RSBG [42]    | 0.80 | 1.53  | 0.33 | 0.64  | 0.59 | 1.25 | 0.40 | 0.86 | 0.30 | 0.65 | 0.48 | 0.99 |
| Star [51]    | 0.56 | 1.11  | 0.26 | 0.50  | 0.40 | 0.89 | 0.31 | 0.71 | 0.52 | 1.13 | 0.41 | 0.87 |
| SCAN [21]    | 0.57 | 0.78  | 0.43 | 0.85  | 0.61 | 1.28 | 0.39 | 0.84 | 0.34 | 0.74 | 0.46 | 0.89 |
| Trajectorn++ [52] | 0.71 | 1.68  | 0.22 | 0.46  | 0.41 | 1.07 | 0.30 | 0.77 | 0.23 | 0.59 | 0.37 | 0.91 |
| SSAGCN-W/o-sen | 0.50 | 0.87  | 0.31 | 0.49  | 0.35 | 0.72 | 0.27 | 0.45 | 0.24 | 0.36 | 0.33 | 0.58 |
| SSAGCN-W/o-seq | 0.45 | 0.76  | 0.29 | 0.52  | 0.35 | 0.68 | 0.25 | 0.43 | 0.26 | 0.40 | 0.32 | 0.51 |
| SSAGCN-W/o-ssa | 0.38 | 0.63  | 0.25 | 0.48  | 0.26 | 0.55 | 0.22 | 0.42 | 0.23 | 0.36 | 0.27 | 0.49 |
| SSAGCN       | 0.30 | 0.59  | 0.22 | 0.42  | 0.25 | 0.47 | 0.20 | 0.39 | 0.14 | 0.28 | 0.22 | 0.43 |

| TABLE II |

**EVALUATION RESULTS OF SSAGCN ON SDD TAKING ADE AND FDE AS EVALUATION CRITERIA AND PIXEL AS SCALE. THE DATA IN THIS TABLE ARE FROM THE RESULTS REPORTED IN THEIR WORK.**

| Metric   | SGAN [45] | Sophie [15] | PMP-NMMP [50] | GTTPO [55] | PECnet [54] | LB-EBM [57] | Y-Net [58] | SSAGCN |
|----------|------------|-------------|---------------|------------|-------------|-------------|------------|--------|
| ADE      | 27.25      | 16.27       | 14.67         | 10.13      | 9.96        | 8.87        | 7.85       | 10.36  |
| FDE      | 41.44      | 29.38       | 26.72         | 15.35      | 15.88       | 15.61       | 11.85      | 11.80  |

Therefore, the input dimension of the embedding layer is 10. We set the output dimension of the embedding layer as 5, corresponding to the number of parameters in the Gaussian distribution. The input and output dimensions of TCNs correspond to the time dimensions of the input sequence and the prediction sequence, respectively. To avoid oversmoothing,
TABLE III
AVERAGE PERCENTAGE OF HUMAN COLLIDING FOR EACH SCENE IN ETH AND UCY DATASETS. THE HUMAN COLLISION IS DEFINED IN [15]

|        | GT  | Liner | SGAN | Sophie | SCAN | Ours |
|--------|-----|-------|------|--------|------|------|
| ETH    | 0.000 | 3.137 | 2.509 | 1.757 | 0.793 | 1.145 |
| HOTEL  | 0.092 | 1.568 | 1.752 | 1.936 | 1.126 | 0.573 |
| UNIV   | 0.124 | 1.242 | 0.559 | 0.621 | 0.481 | 0.277 |
| ZARA1  | 0.000 | 3.776 | 1.749 | 1.027 | 0.852 | 0.634 |
| ZARA2  | 0.732 | 3.631 | 2.020 | 1.464 | 3.109 | 1.428 |
| **Avg** | **0.189** | **2.670** | **1.717** | **1.361** | **1.272** | **0.811** |

we only use one layer of GCN and six layers of TCNs. Finally, we use the stochastic gradient descent (SGD) optimizer and learning rate with 0.001 to train the SSAGCN model for 200 epochs on Tesla V100 GPU.

B. Quantitative Analysis

We choose the state-of-the-art methods as the baselines.

S-LSTM [10]: It is an LSTM network based on prediction method using an ingenious pooling mechanism.

SGAN [45]: It is one method introducing generative adversarial network (GAN) into pedestrian trajectory prediction and using global pooling for interaction.

SR-LSTM [13]: It is a prediction method using states refinement for LSTM network.

Sophie [15]: It is a GAN-based prediction method considering both scene factors and social factors using an attention mechanism.

STGAT [18]: It is a spatial–temporal GAT based on a sequence-to-sequence architecture to predict future trajectories of pedestrians.

Social-BiGAT [19]: It is one method adopting a cyclic confrontation structure and introducing one GAT to calculate the impact between pedestrians.

STSTGCN [20]: It is one method modeling pedestrian trajectories into graphs and using STGCN to deal with social interactions.

RSBG [42]: A recursive social behavior graph combined with GCN is proposed to model social interaction.

NNMP [50]: A neural motion message passing is proposed for interactive modeling, which can predict future trajectories in a variety of scenarios.

Star [51]: A novel spatial graph transformer is introduced to capture the interaction between pedestrians.

PECNet [54]: It is a two-stage prediction framework used to predict the endpoint of the trajectory and, then, a reasonable path planning.

SCAN [21]: It is a spatial context attentive network that can jointly predict socially acceptable multiple future trajectories for all pedestrians in a scene.

CARPe [53]: It is a convolutional approach for real-time pedestrian path prediction, which utilizes a variation of graph isomorphism networks in combination with an agile CNN design.

SGCN [44]: It is a pedestrian path prediction method using sparse graph convolution and self-attention mechanism to calculate asymmetric attention score matrix.

GTPPO [55]: It is a graph-based pseudo-oracle trajectory predictor (GTPPO), which encodes pedestrian movement patterns using short and long memory units and introduces temporal attention to highlight specific temporal steps.

Trajectron++ [52]: It is an approach designed to be tightly integrated with robotic planning and control frameworks, which can produce predictions that are optionally conditioned on ego-agent motion plans.

Introvert [56]: A pedestrian trajectory prediction method using 3-D visual attention mechanism to capture dynamic scene context.

LB-EBM [57]: It is a latent belief energy-based model (LB-EBM) for diverse human trajectory forecast.

Y-Net [58]: It is a scene compliant trajectory forecasting model that exploits the proposed epistemic and aleatoric structure for diverse trajectory predictions across long prediction horizons.

We compare the experimental results with above baseline methods, and their ADE and FDE are shown in Table I. Compared with these methods, our method achieves the best performance at $K = 1$ and $K = 20$. $K = 1$ means that the model generates only one trajectory, while $K = 20$ means that the model predicts 20 trajectories or samples 20 times. It can be seen from Table I that Trajectron++ has been at the forefront in the past due to its advanced predictive framework and unique modeling approach, and has a decrease of nearly 45% on ADE and FDE compared to the methods during the same period. Recent methods, such as SGCN, Introtvet, and LB-EBM, provide new research ideas for trajectory prediction.

SGCN introduces a sparse graph convolution network and combines with a self-attention mechanism, which reduced ADE and FDE by 15% compared with other GCN-based methods. Introvert converts trajectory prediction to 3-D domain, focusing on dynamic scene context, with a 17% FDE reduction compared to Trajectron++. Our model uses a customized social soft attention function, which covers various pedestrian social interaction factors, enabling our model to better learn the social interaction between pedestrians. In addition, our model has the additional input of the scene image and extends the role of the scene in time and space. As a result of the above novel contributions, our model achieved state-of-the-art performance for diverse trajectory predictions across long prediction horizons.

Table II shows the performance of our approach compared to others on the SDD dataset. It can be seen that the performance of SSAGCN on SDD is also competitive compared with other baseline methods, and the FDE is at its lowest level. Our model performs social interaction and scene interaction at the same time, which tends to cause a relatively large error in the dataset where social interaction and scene interaction are unevenly distributed. In fact, the obstacle and road information...
is abundant in the SDD dataset, and the input of scene information can provide great help for trajectory prediction. The Y-Net model contains the input of the scene image, and its results are 20% less in ADE and 25% less in FDE than previous methods. Our SSAGCN also takes as the input scene image and enhances the role of physical scene in spatial and temporal space, which makes our prediction result 0.05 lower than that of Y-Net on FDE.

In addition, we use the percentage of near collisions (whether the distance between two pedestrians is less than 0.1 m) used by SoPhie [15] to further evaluate our results, which are shown in Table III. In the aspect of this new evaluation methodology, our method is completely beyond other methods, which indicates that it can generate better socially and physically acceptable trajectories for each pedestrian.

C. Ablation Study

We set up two groups of comparative experiments to prove the effectiveness of the new social soft attention function. The SSAGCN-w/o-ssa in Table I shows the results obtained by directly using the adjacent matrix $A_i^j$ to aggregate node feature without our social soft attention function. The results show that the role of social soft attention function is positive. In addition, we insert the modules in our model to other baseline methods for experiments. The results are shown in Table IV. We can see that the addition of $F_{ssa}$ module to SSTGCNN can reduce ADE/FDE by 6%/12.5%, and the addition of scene module to SSTGCNN can reduce ADE/FDE by 8%/7.9%. Furthermore, both of these two modules can help to reduce by 16%/17%. The addition of $F_{ssa}$ module to the Sophie model can reduce ADE/FDE by 11%/19.1%, and that of scene module by 16.7%/25.2%, respectively, and both of them to reduce 20%/27%. Moreover, the addition of $F_{ssa}$ module to the SGCN model can reduce by 2.63% on ADE. When only adding the scene module, it will affect the function of the self-attention in SGCN and further has a negative effect on the accuracy of trajectory prediction. While simultaneously embedding the $F_{ssa}$ and scene modules into SGCN, it can improve the performance by 7.89%/1.47% on ADE and FDE, respectively. Nonetheless, from these quantitative results, we conclude that the proposed module can play a good role in the GCN framework to improve social interaction and scene interaction.

In addition, when we implement the social soft attention function, we use a hyperparameter $\theta$, to specify the values of the diagonal elements in the matrix $A_i^j$. The value of the diagonal element refers to the impact of pedestrians on themselves. When we implement SSAGCN, we take the $\theta$ value in the interval [0.04, 0.16], the mid-value of which is the maximum value of the off-diagonal element calculated by the social soft attention function. In Fig. 6, we take seven different values of $\theta$ from the above interval for experiment. The performance of the model increases as $\theta$ decreases from 0.16 to 0.1. The model starts to look bad as we go down further. As $\theta$ continues to decrease from 0.1, the performance of the model degrades.

In order to prove the rationality of social soft attention function, we used different combinations of speed, direction, and distance to conduct ablation experiments. The experimental results are shown in Table V. It can be seen that better results cannot be achieved by simply considering more factors. The influence between nodes is still dominated by the distance between them. However, it is not reasonable to distinguish the influence of different nodes only by distance. Our function achieves better results by combining more influences in a suitable form.

In order to further verify that our method is more effective in exploring scene information, we implement the ablation experiment of sequential scene attention sharing mechanism (shown in Table I). SSAGCN-w/o-sen in Table I shows the evaluation results without the sequential scene attention sharing and
TABLE VI
COMPARATIVE EXPERIMENTS ON THE EFFICIENCY OF GRAPH STRUCTURE-BASED METHODS. THE FIRST ROW SHOWS THE NUMBER OF PARAMETERS CONTAINED IN EACH MODEL. THE REMAINING LINES ARE THE TIME TAKEN BY EACH MODEL TO PREDICT THE TRAJECTORY ON DIFFERENT DATASETS

| Parm | STGAT | SSAGAT | SGCN | SSTGCNN | SSAGCN |
|------|-------|--------|------|---------|--------|
| ETH  | 44.630| 12.575 | 25.369| 7.563   | 7.578  |
| HOTEL | 4.606 | 9.533  | 5.155 | 4.635   | 5.060  |
| UNIV | 9.935 | 15.345 | 6.233 | 4.545   | 4.585  |
| ZARA1| 29.653| 40.445 | 22.252| 19.545  | 17.106 |
| ZARA2| 18.135| 24.255 | 9.705 | 6.755   | 6.635  |
| AVG  | 17.845| 24.825 | 11.545| 9.045   | 8.595  |

the SSAGCN is the result of our complete model. It can be seen that due to the efficient scene sharing, the average ADE and FDE are increased by 16% and 11%, respectively. In addition, in order to further demonstrate the advantage of sequential scene attention, we also designed an experimental setup called STGCNN-w/o-seq that used only the last frame of the historical trajectory to calculate the impact of the scene and compare it to the full version. The results show that it is more advantageous to consider the impact from sequential attention sharing not just one single frame.

The structure of GCN and social soft attention function in our method is based on prior knowledge. Compared with learning methods such as GAT and SGCN, it has fewer parameters and is faster. We conduct a set of experiments to replace the GCN and the social soft attention function in our method with GAT. Since SGCN and SSTGCNN do not contain modules about the scene, the result of our model shown in Table VI is also the version without the scene information. It can be seen that the methods of modeling pedestrian interaction based on prior knowledge, SSAGCN and STGCNN, have fewer parameters and faster inference speed than that using the attention mechanism in the previous columns. It is to be mentioned that our model and SSTGCNN have a similar number of parameters, but the prediction speed of our model is faster than that of other methods in this table.

D. Qualitative Analysis

The quantitative analysis indicates that our model has high accuracy. Furthermore, we implement a qualitative analysis to show that the generated trajectories have better social acceptability and physical rationality. As shown in Figs. 7 and 8, we select some representative cases from testing datasets.

1) Social Acceptability: In Fig. 7, we choose the parallel, encounter, and hybrid cases. In the first column of Fig. 7, parallel pedestrians generally maintain the same state of motion. The SGAN model built on location-based pooling in this case produces a false prediction of over-avoidance. Attention-based methods, such as Sophie and SGCN, tend to maintain the original state due to self-attention and are insensitive to turning. SSTGCNN successfully predicts the motion trend, but the accuracy is not high. In the second column of Fig. 7, the pedestrian meeting from opposite directions should have some tendency to avoid collisions. However, only our method and SSTGCNN show a tendency to avoid collisions. In the third column of Fig. 7, pedestrians should strike a balance between avoiding collision and maintaining the original state in the mixed case of pedestrian parallelism and encounter. This is a huge challenge for trajectory prediction. It can be seen from Fig. 7 that all models fail to accurately predict future trajectories in this complex situation. However, our approach produces fewer errors than other methods. The above advantages are attributed to the fact that the SSA function in our method can capture the pedestrian interaction rules more closely to the real situation.

In Fig. 8, we visualize the graph relationships of three methods based on weighted matrices and GCN. Our method in the first column does a good job of differentiating and filtering the relationships between agents. Agents with higher collision risk also have larger edge weights. The weights of edges between agents with low collision risk are small or zero. Agents who are walking abreast do not generate edges due to the limitation of cosine in our social soft attention function. In the second column, the filtering effect of the kernel function of SSTGCNN is so small that the constructed graph is similar to a full connection graph. Moreover, the weight differences between edges in this graph are very small. This makes SSTGCNN unable to judge which neighbor has a greater influence on the agent in the face of complex situations. In the third row for SGCN, most edges between adjacent nodes are removed, and only a small number of edges with low weights are left. Due to the influence of self-attention mechanism, in most of the graph relational adjacency matrices established by SGCN, only the diagonal elements have weights. Therefore, the prediction results of SGCN generally tend to maintain the original motion state. From the comparison results in Fig. 8, it can be seen that the graph established by our method is more reasonable and more capable of distinguishing and filtering social interactions.

2) Physical Rationality: It is widely known that reasonable predictions should conform to realistic physical rules. In Fig. 9, we visualize the predicted results on the SDD dataset to show the model’s response to the scene. The SDD contains a large number of roads and obstacles and there are relatively few pedestrians. Therefore, the input of scene image plays a more important role in trajectory prediction on SDD. Our model takes as the input scene images in the form of binary semantics and calculates scene attention at each moment; therefore, the whole trajectory in our prediction results will be affected by the scene. In addition, the scene sharing mechanism in our model enables the influence of the scene not to be limited to local areas. For example 2 in Fig. 9, the restriction of the road is transmitted by the pedestrian at the road boundary to the pedestrian in the middle of the road, thus keeping the pedestrian in the middle of the road at a distance from the boundary.

In the third column of Fig. 9, examples 3 and 6 show the failure cases. In example 3, the two pedestrians met directly and did not dodge to keep their distance, while in our prediction, the two pedestrians should turn to avoid the collision. In example 6, a pedestrian walked directly toward the obstacle and then stopped to rest, while our prediction...
concludes that he should bypass the obstacle. Even in this challenging case, the trajectory predicted by our method remains consistent with the rules of physics. We predict that two people who are meeting will follow the road and avoid collisions, while a person walking straight will steer clear of obstacles. This shows that our model can produce reasonable
Fig. 8. Visualization of graph relationships in different methods. Each blue node corresponds to one pedestrian and the edge connected to the node itself is not drawn. The blue dotted line represents the historical trajectory (eight frames) of the pedestrian, and the red dotted line represents the ground truth (12 frames). The color of the edge is set according to the color bar on the right, which describes the weight of the edge. The greater the weight of the edge, the greater the influence between the pedestrians.

Fig. 9. Visualization of the trajectory prediction of our method on SDD. The green dotted line is the single trajectory predicted by our method, the red dotted line is the ground truth (12 frames), and the blue dotted line is the historical trajectory (eight frames). The unmarked pedestrians in this figure are eliminated in the data processing.

prediction trajectories in most cases. In Fig. 10, we visualize the predicted results on the ETH/UCY dataset to show the model’s positive results and negative results. From the positive results, it can be seen that SSAGCN can reasonably calculate the influence between different individuals when dealing with social interactions, thereby producing prediction results that
Fig. 10. Visualization of the trajectory prediction of our method on ETH/UCY. The orange dotted line is the single trajectory predicted by our method, the red dotted line is the ground truth (12 frames), and the green dotted line is the historical trajectory (eight frames). The unmarked pedestrians in this figure are eliminated in the data processing.

conform to realistic scenarios. The failure cases of SSAGCN mainly focus on the situation where scene interaction and social interaction are concentrated at the same time, which shows the disadvantage of SSAGCN model. In certain situations, the derived predictions tend to keep distance from obstacles when the scene interaction dominates, while in other situations, they tend to keep social distance from neighbors to avoid collisions due to the profound influence of social interaction. For these cases, although our model can obtain prediction results that conform to social laws and the physical constraints of the scene, there is still a certain deviation with the actual trajectory because the proposed model cannot sensitively handle the unbalancing problem between social interaction and scene interaction. This causes our model to be wrong on some agents when predicting the future trajectory of agents close to the obstacle. However, as the final positive example in Fig. 10, single agent close to an obstacle would not make this error.

V. CONCLUSION

In this article, we propose SSAGCN, a trajectory prediction model, which comprehensively considers a variety of social factors and sequential scene information to obtain accurate and reasonable prediction trajectories. When dealing with the social interaction of pedestrians, we use coordinate data to calculate the relationship of pedestrian speed, direction, and distance and then use the social soft attention function to calculate the influences between pedestrians. When considering the context of the scene, we realize that the scene information should influence the trajectory at each moment, and the scene interaction should occur simultaneously with social interaction. Therefore, we calculate the attention at each moment in the scene and then embed the sequential scene attention into the social graph so that the influence of the scene is shared and spread in the graph according to the social relations. Experimental and evaluation results on public datasets show that our method achieves the best results on most datasets due to the social soft attention function and the scene attention sharing mechanism. A qualitative analysis proves that our model can predict trajectories that are socially and physically acceptable to all pedestrians. In our method, the social soft attention function is an extension of the social force. According to the practical rules, the interaction between pedestrian and collision risk is quantified and applied to the distinction and filtering of adjacency relationship in GCN. We show the rationality of social soft attention function in the visualization of graph relation.

Our prediction model can predict reasonable results in various situations, so it can be used for trajectory prediction in complex scenes and provide a reference for collision avoidance for autonomous driving platforms. In our efficiency comparison experiment, we can see that for most prediction methods, their efficiency is not so high in the face of complex and dense scenes. Besides, since the function is designed based on the inherent graph structure, it is not suitable for the direct application to those non-GCN methods. However, the underlying motivation of the social soft attention function can be transferred and applied to non-GCN-based methods, and we will explore potential applications in future work. Moreover, we will also focus on improving the prediction efficiency and accuracy of prediction methods in scene with dense traffic. A promising improvement is to use subgraph modeling of social relationships in complex scenarios to reduce computational redundancy and improve efficiency and accuracy.

REFERENCES

[1] H. Bai, S. Cai, N. Ye, D. Hsu, and W. S. Lee, “Intention-aware online POMDP planning for autonomous driving in a crowd,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2015, pp. 454–460.
[2] Y. Luo, P. Cai, A. Beta, D. Hsu, W. S. Lee, and D. Manocha, “PORCA: Modeling and planning for autonomous driving among many pedestrians,” IEEE Robot. Automat. Lett., vol. 3, no. 4, pp. 3418–3425, Oct. 2018.
[3] P. Raksincharoensak, T. Hasegawa, and M. Nagai, “Motion planning and control of autonomous driving intelligence system based on risk potential optimization framework,” Int. J. Automot. Eng., vol. 7, no. 14, pp. 53–60, 2016.
H. Xue, D. Q. Huynh, and M. Reynolds, “SS-LSTM: A hierarchical S. Yi, H. Li, and X. Wang, “Understanding pedestrian behaviors from S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic S. Yan, Y. Xiong, and D. Lin, “Spatial temporal graph convolutional net- J. Sekhon and C. Fleming, “SCAN: A spatial context attentive network S. Tang, X. Shu, R. Yan, and L. Zhang, “Coherence constrained graph
[52] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, “Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data,” in Computer Vision—ECCV (Lecture Notes in Computer Science). Glasgow, U.K.: Springer, 2020, pp. 683–700.

[53] M. Mendieta and H. Tabkhi, “CARPe Posterum: A convolutional approach for real-time pedestrian path prediction,” 2020, arXiv:2005.12469.

[54] K. Mangalam et al., “It is not the journey but the destination: Endpoint conditioned trajectory prediction,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2020, pp. 759–776.

[55] B. Yang, G. Yan, P. Wang, C.-Y. Chan, X. Song, and Y. Chen, “A novel graph-based trajectory predictor with pseudo-oracle,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 12, pp. 7064–7078, Dec. 2022.

[56] N. Shafiee, T. Padir, and E. Elhamifar, “Introvert: Human trajectory prediction via conditional 3D attention,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 16815–16825.

[57] B. Pang, T. Zhao, X. Xie, and Y. N. Wu, “Trajectory prediction with latent belief energy-based model,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 11814–11824.

[58] K. Mangalam, Y. An, H. Girase, and J. Malik, “From goals, waypoints & paths to long term human trajectory forecasting,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 15233–15242.

Pei Lv received the Ph.D. degree from the State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, China, in 2013.

He is currently an Associate Professor with the School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China. His research interests include computer vision and computer graphics. He has authored more than 50 journal articles and conference papers in the above areas, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON COMPUTER VISION AND PATTERN RECOGNITION, the ACM Transactions on Computer and Multimedia Systems, the ACM Transactions on Multimedia Computing, Communications and Applications, ACM Transactions on Multimedia Systems, and the IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS.

Wentong Wang received the B.S. degree in computer science and technology from Shenyang Ligong University, Shenyang, China, in 2019. He is currently pursuing the master’s degree with the Henan Institute of Advanced Technology, Zhengzhou University, Zhengzhou, China.

His current research is focused on deep learning algorithms and applications for autonomous driving, including driving behavior learning, pedestrian trajectory prediction, and pedestrian intention prediction.

Yunxin Wang received the B.S. degree in information and computing sciences from Northeast Electric Power University, Jilin, China, in 2018. He is currently pursuing the master’s degree with the Henan Institute of Advanced Technology, Zhengzhou University, Zhengzhou, China.

His research interests mainly include graph convolutional networks, pedestrian trajectory prediction, and action recognition.

Authorized licensed use limited to: University of Nebraska. Downloaded on July 15, 2022 at 09:42 For personal use only. No other uses without permission. All rights reserved.