Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 1, No. 3, 2022

Review of pedestrian trajectory prediction methods

Xiaochuan Tan*, Ruiyuan Liu, Shuai Zhang, Jiaojiao Li, Pengcheng Ma

School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China.

* **Corresponding author**: Xiaochuan Tan (Email: 915486554@qq.com)

Abstract: Urban driverless vehicles will inevitably interact with pedestrians in the process of driving. In order to avoid path conflict with pedestrians, the research on pedestrian trajectory prediction is of great significance. This paper mainly summarizes the technical classification and research status of pedestrian trajectory prediction at this stage in detail. According to the different modeling methods, the existing trajectory prediction methods are divided into shallow learning-based trajectory prediction methods and depth learning based trajectory prediction methods. The advantages and disadvantages of the depth learning based trajectory prediction methods are compared, and the current mainstream pedestrian trajectory prediction public dataset is summarized, and the performance of the mainstream pedestrian trajectory prediction methods is compared in the dataset. Finally, the challenges and development trend of pedestrian trajectory prediction at this stage are prospected.

Keywords: Pedestrian trajectory prediction; Automatic driving; Deep learning; Prediction method; Neural network.

1. Introduction

With the rapid development of the global economy and the acceleration of urban processes, the global automotive industry is developing rapidly. Global car ownership is expected to reach 2.5 billion vehicles by 2050. The popularity of automobiles brings convenience to people 's lives. At the same time, traffic congestion and frequent traffic accidents are also increasing, which brings great hidden dangers to driving safety. Pedestrians and cyclists, as vulnerable groups among road traffic participants, are extremely vulnerable without any safety protection devices when participating in traffic activities. According to the data of China 's National Bureau of Statistics, the number of road vulnerable groups killed by traffic accidents in China accounts for 26 % of the total number of accidents every year. Therefore, to further enhance road traffic safety and pedestrian safety, reduce casualties and property losses is the future we need to solve the problem. Among them, improving the trajectory prediction of pedestrians crossing the street by autonomous vehicles is an important measure to avoid traffic accidents.

Understanding human motion is a key skill for the coexistence and interaction between intelligent systems and humans, which involves representation, perception and motion analysis. Prediction plays an important role in human motion analysis. As time goes by, the model can predict scenes involving multiple agents and integrate this scene information in an active way, that is, to enhance the effects of active perception, predictive planning, model predictive control or human-computer interaction. Therefore, in recent years, pedestrian trajectory prediction has been the focus of research in many fields, such as autonomous vehicles, service robots, intelligent transportation, command cities, etc.

Ensuring the safety of road users in traffic scenarios is a prerequisite for the widespread use of autonomous vehicles [1]. If the autonomous vehicle can accurately predict the location of the surrounding pedestrians in the traffic scene, it can avoid traffic accidents and ensure pedestrian safety. At present, there are three main difficulties in pedestrian crossing trajectory prediction. (1) As a more flexible participant in the traffic scene, it is a difficult task to accurately predict the future trajectory of pedestrians. However, by observing the trajectory of its historical moments, some advanced algorithms can be used to roughly predict the future trajectory of pedestrians. However, in actual traffic scenarios, compared with traffic participants such as vehicles, pedestrian movement is more flexible and can turn, stop and move at any time. It is difficult to establish a kinematic model suitable for pedestrians, and sometimes even drivers are difficult to predict the future trajectory of pedestrians. (2) In the actual traffic scene, the future movement of a pedestrian is not only dominated by personal will, but also by the surrounding traffic environment and the influence of the surrounding pedestrians (such as walking together, interaction, etc.). According to the research of Moussaid M et al. [2], 70 % of pedestrians tend to walk together, and the accompanying pedestrians often interact in the same time and space. This interaction between pedestrians is very abstract and difficult to model accurately in the algorithm. (3) The realization of the conventional pedestrian trajectory prediction algorithm model is to find a function mapping from input to output. For the trajectory prediction model, it corresponds to the mapping between different sequences. The conventional model or training method is easy to make the model prediction result fall into a compromise state (the prediction result tends to predict a compromise trajectory). Obviously, the conventional training model cannot effectively and accurately predict the pedestrian trajectory.

Due to the influence of pedestrian 's subjective intention and objective environment, the interaction between pedestrians and between pedestrians and the environment becomes complex and abstract. The traditional pedestrian trajectory prediction model has been unable to meet the interaction in responsible scenarios, and the environmental adaptability is poor, which limits the prediction performance of the model. With the development of deep learning, neural networks have made major breakthroughs in the fields of image recognition, classification, and tracking. Its complete theoretical system and rich network models provide the necessary conditions for deep learning to be applied to the field of pedestrian trajectories. In particular, recurrent neural network (RNN), generative adversarial network (GAN) and graph convolutional network (GCN) for sequence learning have become the main networks for pedestrian trajectory prediction modeling.

The high dynamics and randomness of pedestrian trajectories and the complex interaction with traffic environment agents make trajectory prediction challenging. However, it is still necessary to predict pedestrian trajectories for a long time, which has a great impact on the active planning and decision-making of autonomous vehicles. At present, the research on pedestrian trajectory prediction is increasing at home and abroad, and it is necessary to review the related technologies and literatures in this field. In this paper, the problem description, research development and problem processing flow of pedestrian trajectory prediction are sorted out, and the research progress of pedestrian trajectory prediction, especially the research results in recent years, is reviewed.

The main contents of the article are as follows: Part 1 introduces the current pedestrian trajectory prediction problem and leads to the common pedestrian trajectory prediction methods; in the second part, the existing trajectory prediction methods are classified, and the methods based on shallow learning and machine learning are briefly summarized. The advantages and disadvantages of pedestrian trajectory prediction methods based on different neural networks are compared. The third part introduces the current mainstream trajectory prediction public data set in the industry. After testing some trajectory prediction methods in the data set, the methods to obtain excellent test results are analyzed and compared. Part 4 and 5 summarize the existing problems and future research directions of pedestrian trajectory prediction. The last part summarizes the full text.

2. Related work

2.1. Pedestrian Trajectory Prediction Problem Description

The essence of the pedestrian trajectory prediction problem is to infer the location and possible state of the pedestrian in the future based on the pedestrian characteristic information and the environmental information. The problem can be regarded as a sequential decision problem, that is, through the historical trajectory of the pedestrian observed in the past, the historical information of the self-movement, etc., through the establishment of the model, the machine learns some rules generated by the behavior reasoning, the interaction with others, the influence of the surrounding environment, etc.[3,4]to understand the human movement in the complex environment, so as to predict the pedestrian's position coordinates and motion trajectory in the short time (such as 8s) in the future according to the trajectory of the pedestrian in the past time period.

Due to the complexity and uncertainty of pedestrian-pedestrian and environment interaction, pedestrian trajectory prediction is difficult. Traditional methods have conducted a series of studies on pedestrian interaction through social force model [5-8], multi-model method and mixed estimation [9-10], but the above methods have poor universality and have certain limitations. In recent years, with the development of deep learning, pattern-based methods can learn to fit different functions (such as neural networks) from data to learn human interaction perception, greatly improving the flexibility and generalization ability of the model. Since the trajectory of pedestrians is not only affected by the surrounding pedestrians, but also by the scene environment, the methods based on scene interaction mainly include static obstacle avoidance method [11], map perception method [12] and semantic graph method [13-16], which can improve the prediction performance of pedestrian trajectory to a certain extent.

Specifically, the processing flow of typical pedestrian trajectory prediction problems includes: pre-dataset collection, dataset preprocessing and input, feature coding, extraction and aggregation, trajectory prediction visualization and prediction result output, as shown in Figure 1.

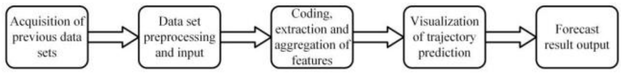

Fig. 1 Process of pedestrian trajectory prediction problem

2.2. Introduction of pedestrian trajectory prediction method

Pedestrian trajectory prediction methods can be roughly divided into methods based on traditional probability prediction and methods based on deep learning. Based on the traditional probability prediction method, the pedestrian prediction is transformed into a probability prediction problem. The pedestrian kinematics model is established to estimate the change of pedestrian motion state with time, so as to obtain the pedestrian trajectory results. Nicolas Schneider et al. [17] used dynamic Bayesian network to determine the trajectory of pedestrians according to the head posture of pedestrians, the nearest collision point between pedestrians and vehicles, and the distance between pedestrians and roadsides when the scene is relatively certain, that is, pedestrians are ready to cross the road. QM Raul et al. [18] extracted the three-dimensional time-related information of the key points of the pedestrian skeleton into a low-dimensional space based on the balanced Gaussian process dynamics model (B-GPDMs), and obtained multiple models of the four behaviors of pedestrian walking, stopping, starting and standing during use, and selected the most similar model to predict the future path, posture and intention of pedestrians. However, these methods cannot capture the complex and changeable motion characteristics and rich environmental characteristics of pedestrians in practical use, and are sensitive to input noise. They can only obtain the prediction results in a short time range, which has great problems in practical use.

With the continuous development of deep learning, researchers began to introduce deep learning methods into pedestrian trajectory prediction tasks. At present, the common pedestrian trajectory prediction methods based on deep learning mainly include RNN-based pedestrian trajectory prediction, GAN-based pedestrian trajectory prediction and GCN-based pedestrian trajectory prediction.

Pedestrian trajectory prediction based on RNN. The recurrent neural network RNN is the earliest model for pedestrian trajectory prediction. It determines the output by inputting and storing information in the historical network. Through this feature, RNN can predict future values based on historical sequence information. It can be said that RNN is designed for sequence modeling and has a recursive organizational structure, showing strong modeling capabilities in time analysis and sequence learning. However, as the length of the time series increases, the shortcomings of RNN gradually emerge. RNN cannot achieve long-term memory of the data state, resulting in the network layer stopping learning. In terms of effectiveness, RNN will store all historical information in the network, which will lead to gradient disappearance or gradient explosion in large networks during training. In terms of operating efficiency, because the current state of the RNN model depends on the hidden state of the previous moment, parallel processing cannot be realized, resulting in low training and reasoning speed of the model. In pedestrian trajectory prediction, a large number of network nodes and huge data sets are needed to train the network to improve the accuracy of prediction. Therefore, the traditional RNN will not meet the needs of pedestrian trajectory prediction. However, researchers have developed related variants based on RNN networks, such as long short-term memory networks (LSTM) and gate cycle units (GRUs), which regulate the flow and selection of information through gating mechanisms to solve the problem of information transmission in long sequences. In the field of pedestrian trajectory prediction, the above variant methods have focused on RNN research and achieved remarkable success [19-22].

Pedestrian trajectory prediction based on GCN. The graph convolutional neural network GCN is a method that can perform deep learning on graph data by using the edge and node data of the graph as input for learning and training. Although RNN has significant sequence modeling ability, it lacks intuitive high-level spatio-temporal structure. In the field of pedestrian trajectory prediction, the number of pedestrians is uncertain, and the interaction between pedestrians is irregular. The graph structure is a natural method to represent the interaction between pedestrians. It is more intuitive and effective than the RNN method based on aggregation. GCN has a great effect on graph data processing in non-Euclidean space [23]. Its core idea is to map nodes or edges in graph structure to vector space through deep learning methods, and then perform clustering and classification. At present, many methods [24-26] take graph structure as the basic component. These methods usually represent pedestrians as nodes, use their interactions as connections, and combine deep sequence models such as long short-term memory networks for modeling. [27] By adding spatio-temporal data to pedestrian trajectory prediction, GCN can understand pedestrian behavior and accelerate the modeling progress of social interaction. Therefore, GCN has great application prospects in trajectory prediction, but there are still some problems such as shallow network, unstable structure and weak adaptive ability.

Pedestrian trajectory prediction based on GAN. The generative adversarial network GAN is a deep learning model of unsupervised learning. The GAN network overcomes the difficulty of calculating the generation probability by the game between the generation model and the discriminant model. Adding GAN network to pedestrian trajectory prediction can solve the defect that only one ' optimal ' trajectory can be predicted in the past. The network can predict multiple feasible trajectories and further optimize the prediction accuracy through game theory. At the same time, drawing on the successful experience of GAN network in the fields of super-resolution, image conversion and image synthesis, the trajectory sampler [28] combines the GAN input random vector with the hidden representation of other pedestrian trajectories to deal with the interaction between all observed pedestrians. However, the neural network based on GAN model is prone to problems such as slow convergence and mode collapse.

3. Comparison of pedestrian trajectory prediction methods

Since the 1990s, a large number of trajectory prediction networks based on shallow learning have been proposed. These models have high requirements on the calculation examples of computing equipment, and lack unified evaluation criteria. The quality of the data sets used to test the model is also uneven. Therefore, there is little review of trajectory prediction models based on shallow learning. In recent years, machine learning technology, especially deep learning technology, has gradually emerged. Due to the excellent performance of recurrent neural network in processing time series data, time series prediction models based on recurrent neural network emerge in endlessly. The following will classify and summarize the main pedestrian trajectory prediction methods at home and abroad.

According to the modeling method of the prediction model, pedestrian trajectory prediction methods are roughly divided into trajectory prediction methods based on shallow learning and trajectory prediction methods based on deep learning. Among them, the kinematics-based method in shallow learning is the first to be applied in the field of trajectory prediction. Such methods generally need to model the kinematic characteristics (speed, position and angular velocity, etc.) of pedestrians and combine them with Bayesian filters, Markov networks or Bayesian networks to propagate the current state to the future state for prediction. According to whether the interaction between pedestrians is considered, the trajectory prediction model based on deep learning can be divided into single trajectory prediction model and interactive trajectory prediction model. According to whether to generate deterministic pedestrian trajectory, it can be divided into deterministic trajectory prediction model and acceptable trajectory prediction model. With the introduction and optimization of the model, the current trajectory prediction method based on deep learning tends to be modular, the information that needs to be considered and utilized is gradually improved, and the accuracy and real-time performance of the prediction are gradually improved. The classification of pedestrian trajectory prediction methods is shown in Figure 2.

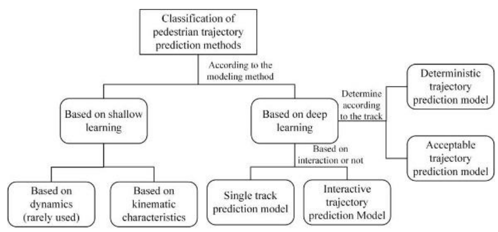

Fig. 2 Classification of pedestrian trajectory prediction methods

3.1. Trajectory prediction method based on shallow learning

The initial trajectory prediction method was to combine the basic kinematic model with the Bayesian filter and its extensions to propagate the current state to the future state [29]. Schneider et al. [30] compared the method based on a single kinematic model with the method based on multi-model interaction. The results show that the method based on multi-model interaction can reduce the lateral position estimation error of 30 cm during maneuvering. However, because the model assumes of constant speed, it also makes it difficult for the model based on Bayesian filter to capture the switching dynamics of pedestrians, and the number of data sets and motion types used for testing are limited, which is not enough to support the establishment and prediction of more complex motion models. Pavlovic et al. [31] used the switched linear dynamic system (SLDS) model [32] to describe nonlinear and time-varying dynamics. The model is based on a Markov chain for probability transfer, and switches between multiple linear kinematic models to predict nonlinear motion in the actual situation. However, the motion feature information based on this model is sometimes insufficient to support the model to switch states, and its effect is limited for some more complex motion models. It is necessary to build larger motion capture data sets to meet the accuracy requirements of complex motion model testing. Kooij et al. [33] established a context based dynamic Bayesian network (DBN) model for pedestrian path prediction. This model combines the context information (pedestrian head direction, emergency degree and environmental space layout) as the potential state to the top of the SLDS model, so as to control the switching state of the SLDS model. Compared with the SLDS model only, this model can make more accurate predictions. However, both SLDS model-based prediction methods and DBN model-based prediction methods always require a lot of calculation in the process of model prediction reasoning and mathematical model building, which is a huge computing power consumption for computing equipment, and cannot well reflect additional scenes (such as traffic lights, crosswalks) and basic motion types (such as turning) based on SLDS model.

Helbing et al. [34] proposed an attractive and repulsive pedestrian motion model, called the social force model. This model is widely used in robotics and activity understanding [35]. Alahi et al. [36] proposed presenting social affinity features by learning their relative position from pedestrian trajectories in a crowd, while Yi et al. [37] proposed using human attributes to improve predictions in a crowd.

With the rapid development of machine learning, kinematics-based methods use some machine learning-based tracking algorithms to improve tracking and prediction when making predictions, such as Kalman filter (KF), Markov model (MM) and Gaussian process [38] (GP). The advantage of the KF model is that it can effectively process the trajectory data of noise-free points when dealing with the pedestrian trajectory prediction problem, and has high prediction accuracy in a short time. On the contrary, the prediction error for a long time is large, and the model complexity is high, which seriously affects the prediction accuracy. The KF model becomes more sensitive with the increase of noise, and the prediction accuracy is approximately linearly reduced. The MM has a good effect on the state prediction of pedestrian movement process, but it is sensitive to the fluctuation of pedestrian trajectory and is not suitable for medium and long-term pedestrian trajectory prediction. The first-order MM only considers the influence of the current pedestrian trajectory point on the future trajectory point, and the data information of the historical trajectory point cannot be used as much as possible, while the high-order MM greatly increases the complexity of the model calculation. The GP model provides a non-parametric model for probability prediction by assuming that the latent variables obey the Gaussian distribution. The trajectory predicted based on the GP model is learned from the historical trajectory data [39], and the uncertainty involved in the trajectory modeling of line prediction is clarified by specifying an appropriate kernel function in the GP model. GP model can effectively predict the trajectory data with noise points, and avoid the lack of discrete nature of trajectory data, and effectively express the statistical characteristics of pedestrian trajectory distribution.

In summary, the trajectory prediction method based on shallow learning has achieved some results in the early stage and contributed to the development of pedestrian trajectory prediction. However, due to the limitations of kinematics-based methods, insufficient extraction of motion feature information, lack of specific scene information, and complexity of model construction, there is a big gap between the final prediction effect and the actual situation. It is difficult for traditional methods to accurately predict more complex pedestrian motion models and scenes.

3.2. Trajectory Prediction Method Based on Deep Learning

Trajectory prediction methods based on shallow learning require complex and rigorous modeling of the model, while trajectory prediction methods based on deep learning generally do not need to assume a fixed mathematical model, and their networks can learn more reasonable mapping relationships with large-scale data sets. In recent years, with the rise of deep learning, a variety of models for processing time series data have been proposed, making trajectory prediction algorithms based on neural networks popular, and the prediction effect of these algorithms has been greatly improved compared with traditional algorithms. Trajectory prediction methods based on deep learning mainly include three types: trajectory prediction method based on RNN, trajectory prediction method based on GAN and trajectory prediction method based on GCN.

3.2.1. Trajectory Prediction Method Based on RNN

The RNN encodes the incoming sequence data into a fixed-size hidden representation, and then uses another RNN to decode the hidden representation to generate a sequential time representation of the output. Although LSTM has the ability to learn and replicate long sequences, they cannot capture dependencies between multiple related sequences. To this end, the social pooling layer of S-LSTM [20] was proposed to heuristically aggregate the information of neighborhood pedestrians. S-LSTM is the first application of recurrent neural networks to simulate human-to-human interaction in crowded scenes and has become a baseline for pedestrian trajectory prediction. The comparison of pedestrian trajectory prediction methods based on RNN is shown in Table 1.

According to the way of obtaining neighbor information, the prediction model can be divided into models based on current results (speed, location, etc.) [13,22,40-41] and models based on previous states [13,20,40-41,42-46]. In order to distinguish the importance of neighbors around pedestrians, some RNN-based pedestrian trajectory prediction methods use the attention mechanism to provide different ways for calculating the weights of neighbors. For example, the soft attention score is calculated based on the hidden state, and the pairwise speed correlation is calculated [42, 45], using an attention model that combines ' soft attention ' and ' hard attention '. Based on LSTM 's position-velocity-time-attention model [47], two temporal attention mechanisms are designed to calculate the hidden state vectors of the position and velocity LSTM layers. The application of attention mechanism enhances the interaction between pedestrians and improves the credibility of trajectory prediction. The RNN-based model has the problems of gradient disappearance and difficulty in expansion during training. Many predictions work combines convolutional neural network (CNN), which confirms that CNN is competitive in trajectory prediction. However, it cannot simulate spatial interaction between pedestrians. In addition, some models combine RNN with other structures. For example, the Gaussian process [48] is combined to predict the complete distribution of pedestrian future trajectories. Due to the uncertainty of the future, CGNS [46] used GRU to infer from the perspective of probability and proposed feasible hypotheses.

Table 1. Comparison of pedestrian trajectory prediction methods based on RNN

Method	Advantage	Shortcoming	Research direction	Loss function
SS-LSTM [19]	Scene Layout	Lack of interaction	Attention mechanism	Mean square error
S-LSTM [20]	Social pooling	Local interaction	Interactive optimization	Negative logarithmic likelihood
ST-LSTM [21]	Spatiotemporal recursion	Loss of interaction	Perceptual interaction	Negative logarithmic likelihood
SR-LSTM [22]	Refinement of intention	Data deviation	Strong adaptability	Mean square error
DESIRE [40]	Inverse optimal control	Spatial constraints	Social association	Cross entropy regression
S-Attention [41]	Social Attention	Maintain the whole drawing	Optimize promotion	Negative logarithmic likelihood
MX-LSTM [42]	Joint forecasting	Narrow promotion	Enrich the scene	Negative logarithmic likelihood
SNS-LSTM [43]	Multimode input	Pay attention to average	Application promotion	Negative logarithmic likelihood
Varshneya[44]	Extent pooling	Various parameters	Multi class objects	Negative logarithmic likelihood
Fernando [45]	Soft and hard attention	Local interaction	Domain expansion	Mean square error

3.2.2. Prediction method based on GAN

GAN-based pedestrian trajectory prediction method introduces the idea of confrontation into the task of pedestrian trajectory prediction by combining sequence prediction model and GAN network, showing the sample space of all possible solutions. The application of GAN network overcomes the shortcomings of most previous methods based on optimizing the distance between pedestrians and predicting only one trajectory.

In the process of generating discrimination, the GAN network is prone to non-differentiable operations. In order to overcome this limitation, S-GAN extends social pooling into a multi-layer perceptron network, predicting multiple socially acceptable trajectories for pedestrians. However, S-GAN is not only simple in modeling pedestrian interaction, but also does not make full use of the deep interaction information of pedestrians. To this end, the subsequent methods explore the influencing factors of pedestrian trajectory by using attention mechanism [13,49-51], increasing scene interaction [13,52-54], feasibility constraints [46], etc. S-GAN and SoPhie are single behavior patterns with high variance, which are limited by social behavior and cannot learn the real multimodal distribution of pedestrians. To this end, a graph-based generative adversarial network Social-BiGAT[51] was proposed to construct a generative model for learning multimodal trajectory distributions. Since the GAN network is prone to mode collapse and fall, Info-GAN in Social Ways [55] can better improve multimodal pedestrian trajectory prediction and avoid these problems. GAN-based methods usually deal with future uncertainty by sampling a latent variable. Previous studies have rarely discussed these latent variables in depth. TPPO [56] designed a latent variable prediction model to estimate the distribution of latent variables from observations and real trajectories, and achieved excellent prediction performance. The pedestrian trajectory prediction method proposed above is mainly to learn and predict in two-dimensional image space. Since human motion occurs in the three-dimensional world, experiments prove that [57] is more effective in learning and predicting pedestrian trajectories in three-dimensional space, which also provides a new direction for the promotion and application of GAN network in the field of trajectory prediction. The comparison of pedestrian trajectory prediction methods based on GAN is shown in Table 2. L2 is the least square error and KL is the relative entropy.

Table 2. Comparison of pedestrian trajectory prediction methods based on GAN

Method	Advantage	Shortcoming	Research direction	Loss function
SoPhie[13]	Semantic scenario	Manual sorting	Flexible performance	L2+confrontation
S-GAN [28]	Reasonable and diverse	Slow sampling	Enrichment mode	L2+confrontation
GCANS [46]	Probability distribution	Complex and changeable	Wide application type	Reconstruction+confrontation+KL
Zhiyuan Zhang [49]	Rich features	Insufficient generalization	Scene interaction	L2+confrontation
Yasheng Sun [50]	Social Attention	Time cost	Information fusion	L2+confrontation
STG-GAN [52]	Global interaction	Inter graph disassociation	Optimize and expand	L2+confrontation
Sun [54]	Reciprocal constraint	Slow fitting	Application promotion	Confrontation+reciprocity
Social Ways [55]	Multimode distribution	Application restrictions	Decision optimization	Countermeasure+information
TPPO [56]	Potential variables	Poor reliability	Improve control	L2+confrontation+KL
Zhong [57]	3D space	High complexity	Generalization performance	L2+confrontation

3.2.3. Prediction method based on GCN

(1) Trajectory prediction based on space-time graph. Based on the wide application of graph convolutional networks in behavior recognition [58], traffic prediction [59], demand prediction [60], etc., many studies have tried to apply spatio-temporal graphs [61,52,60-65] to pedestrian trajectory prediction tasks and achieved good prediction performance.

The spatio-temporal map in the prediction task can be divided into two dimensions: space and time. The space dimension models the interaction between the target pedestrian and its neighbors, while the time dimension models the historical trajectory of the pedestrian. Some methods are extended on this basis. For example, recursive social behavior graph [25] recursively updates individual features within the interaction range to enhance the interaction relationship. Zhang et al. [22] dynamically constructed a directed social graph in the direction of position and speed to effectively capture the interaction behavior of pedestrians. STAR [65] processed the spatio-temporal modeling of graphs based on the enhanced attention mechanism of transformer structure. Liang et al. [24] designed RNN on the spatial graph to encode the inductive bias of pedestrian motion patterns. However, these graph network-based prediction methods do not have multimodal modeling capabilities. For this reason, Ivanovic et al. [61] demonstrated a highly multimodal multi-person scenario that guarantees performance in trajectory prediction modeling. Since these methods introduce a graph at each time step, they can handle graphs that change between prediction steps. However, this is only an implicit ability to handle dynamic edges and cannot explicitly handle dynamic nodes and edges.

(2) Trajectory prediction based on graph attention mechanism. Pedestrian trajectories are often affected by surrounding pedestrians and their obstacles (buildings, sidewalks, grasslands, etc.), which may change or limit human activities. Therefore, it is necessary to pay attention to the factors that have great influence on pedestrian trajectory prediction. In order to capture the interaction between pedestrians and neighbors, some methods calculate the influence factor of pedestrians according to the distance between pedestrians [19,28]. When aggregating the interaction behavior of modeling, they are encoded by pooling, symmetric function or geometric relationship, which is neither intuitive nor direct, and cannot fully interpret the interaction between pedestrians. Graph attention network (GAT) uses soft attention or shift mechanism to distinguish the importance of neighbors, and realizes effective weighted message passing between nodes and better group understanding. Trajectory prediction based on graph attention network [51,60,62-65] By capturing the importance of the surrounding pedestrians to the target pedestrians, it breaks the order dependence of the RNN network and provides a more intuitive method for reproducing the topology of pedestrians in the shared space. However, most of the existing methods are based on the study of the spatial and temporal characteristics of the trajectory. The spatial information representation is single, ignoring the deep motion characteristics in the process of pedestrian movement, such as motion speed, motion direction, motion state, etc., and this multi-feature information is more in line with the real scene. Pedestrian motion state has more application value. The comparison of GCN-based pedestrian trajectory prediction methods is shown in Table 3.

Table 3. Comparison of pedestrian trajectory prediction methods based on GCN

Method	Advantage	Shortcoming	Research direction	Loss function
RSGB [25]	Supervised recursion	Insufficient interaction	Scene interaction	Index L2
Zhang [49]	Time light code	Poor stability	Intertextuality	Reconstruction+ KL
S-BiGAT[51]	Multimodal trajectory	Insufficient features	Deep interaction	Reconstruction+confrontation+KL+L2+classification
STGAT [60]	Serial GAT	Simplify interaction	Supervisory attention	L2 loss
Ivanovic [61]	Multi model and multi person	Computing overhead	Building dynamic map	Negative logarithmic likelihood
S-STGCNN [62]	Application of Graph	Time domain independence	Application expansion	Negative logarithmic likelihood
GraphTCN[63]	Gating adaptation	Single feature	Lack of flexibility	ADE loss
Haddad [64]	Dynamic static interaction	Local interaction	Type promotion	Negative logarithmic likelihood
STAR [65]	Novel framework	Scene missing	Application promotion	Mean square error

4. Data set and Performance comparison

4.1. Data set Introduction

In the pedestrian trajectory prediction method based on depth learning, the data sets involved mainly include ETH/UCY, SDD, DUT, ActEV/VIRAT, Town Center, PETS09S2, Edinburgh, Interaction, DUT and the data sets of Osaki Station, Grand Central and CUHK for crowded scenes.

The ETH and UCY combined data set is a widely used public dataset for evaluating pedestrian trajectory prediction methods. The data set include the global trajectory coordinates of pedestrians in various types of social interaction scenarios. These data sets include the trajectory coordinates of pedestrian interaction, nonlinear trajectory, collision avoidance, standing and group pedestrians, etc., as well as five unique outdoor environment information recorded from the fixed top view. Table 4 shows the detailed introduction of ETH/UCY data set.

Table 4. ETH/UCY Dataset

Scene	Frames	People	Groups	Obstacles
ETH	1448	360	243	44
HOTEL	1168	390	326	25
UNIV	541	434	297	16
ZARA1	866	148	91	34
ZARA1	1052	204	140	34

The SDD data set is a video shot by Robicquet A et al. [66] using UAVs over the Stanford University campus from an overhead perspective, covering 20 different scenes, which involve both dynamic objects and the coordinate positions of static obstacles and objects; ActEV/VIRAT dataset is a public dataset for behavior detection, involving 12 scenarios. The total duration of the video is more than 12h, and the video contains rich annotations, which can last up to 4.5h. The Town Center dataset (1 video), the PETS09S2 dataset (3 videos), and the Grand Central dataset (1 video) were originally used for target tracking and crowd behavior analysis. These three datasets have a short video time, but they contain a lot of social interaction. The Edinburgh data set is the pedestrian movement track collected on the information forum of Edinburgh University. The collection time lasts for several months, including 92000 tracks in total. The interaction data set is jointly collected by UAV and vehicle mounted lidar, and the observation track is obtained through visual inspection technology. The DUT dataset includes the sections with typical mixed traffic characteristics in the main campus of Dalian University of Technology acquired through aerial photography, including 17 intersection scene clips and 11 roundabout scene clips, including 1793 tracks in total.

The Osaki Station data set is the trajectory data collected in Osaka Station, Tokyo, Japan, using two-dimensional laser sensors. The laser sensors can also accurately obtain the position of pedestrians in groups when pedestrians block each other. CUHK dataset is a crowd dataset with different density and different angles shot in many scenes, including real track, group status, crowd video classification, etc.

4.2. Evaluation index

As a subproblem of the prediction problem, pedestrian trajectory prediction is essentially a sequence generation problem. The input sequence of the problem corresponds to all the observed pedestrian positions, and the output sequence represents the position of the pedestrian in the future. Assuming that at any time T, the number of pedestrians is n, the actual position of pedestrian i in the scene is expressed as (xiT,yiT), the position sequence of all pedestrians from time T=1 to T=Tobs (obs represents the length of observation sequence), then the observation time T=Tobs+1 to T=Tobs+pred (pred represents the length of prediction sequence), and the prediction position of all pedestrians (xipT,yipT). When evaluating the performance of trajectory prediction methods, the Average displacement error (ADE) and the Final displacement error (FDE) are usually used to predict and evaluate the trajectory of agents in each scene.

(1) ADE

ADE refers to the average European distance difference between the real track and the predicted track position sequence of each pedestrian in each time step, which is used to evaluate the overall performance of the prediction process, as shown in Formula (1).

$$ADE = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{pred}\sum_{T=T_{obs+1}}^{T_{obs+pred}}\sqrt{(x_i^T - x_{ip}^T)^2 + (y_i^T - y_{ip}^T)^2}$$

(2) FDE

FDE refers to the average Euclidean distance difference between the real track and the predicted track of each pedestrian at the destination, as shown in Formula (2).

$$FDE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i^{T_{obs+pred}} - x_{ip}^{T_{obs+pred}})^2 + (y_i^{T_{obs+pred}} - y_{ip}^{T_{obs+pred}})^2}$$

4.3. Performance comparison

At present, ETH and UCY datasets are widely used in the field of pedestrian trajectory prediction. Although many data sets for trajectory prediction also appeared in the later stage, due to the relatively late appearance of these data sets, the validation of some early published trajectory prediction methods on these data sets lacks sufficient experimental data, so it is difficult to accurately test the effectiveness of these data sets. Therefore, this paper only compares the performance of models based on ETH and UCY data sets. Table 5 shows the performance of some methods mentioned in this article on ETH and UCY datasets.

It can be seen from Table 5 that the trajectory prediction method based on deep learning has better effect, and the recognition accuracy is much higher than that based on traditional shallow learning method, which shows extremely excellent performance in ETH and UCY data sets. With the development of generative network and graph based neural network technology and the combination of these two networks and various semantic scene information, the test performance of many trajectory prediction methods on the public data set is increasingly excellent. Among them, the learning method based on generative countermeasure network represented by Social ways has more significant performance and relatively high recognition rate. The ADE value and FDE value of Social ways method reach 0.45 and 0.81 respectively, the performance of the proposed method is greatly improved. It can also be found that the trajectory prediction methods based on graph networks (GAT, STGAT, Social BiGAT, TPNet-20, Social STGCNN) are developing rapidly. When S-RNN was proposed, the performance of S-RNN was only 50% of that based on LSTM. However, with the improvement of graph models and the application of GCN and GAT models, the prediction accuracy of trajectory prediction methods based on graph networks was also improved, Its prediction effect has exceeded the trajectory prediction method based on the generative network, and the prediction accuracy is more than 50% higher than S-LSTM. In a word, by reasonably designing the network structure and loss function, the trajectory prediction method based on deep learning can relatively accurately predict the next trajectories of travelers, and can better establish the basic framework of trajectory prediction task. Problems and challenges in pedestrian trajectory prediction

In recent years, although significant achievements have been made in pedestrian trajectory prediction methods, there are still some problems.

The prediction algorithm is not adaptable to the environment. Existing social perception methods assume that all observed pedestrian behaviors are similar, and their movements can be predicted with the same model and characteristics, and the capture and reasoning of high-level social attributes are not strong. Most models are designed for specific scenes, tasks, or motions. These methods perform well when the spatial structure is specific and the motion mode is fixed, for example, when the motion mode is significant in the environment and the spatial structure and pedestrian target are known, but the performance is poor in undefined and changing situations.

The interaction lacks interpretability. When training the network model, the data used are objectively measurable data, which cannot accurately grasp the pedestrian movement intention, and lack of data that rely on human subjective judgment to train the algorithm. Some models have made some attempts to use head posture [42] in combination with pedestrian behavior prediction [24], but the way to obtain data is single, and there is little research on the subjective intentions of pedestrians. Therefore, the current model is still dependent on data-driven because of its lack of interpretability for calculated interactions.

The dynamic graph lacks time sequence correlation. In the network architecture based on graph structure, in the process of building dynamic graph in time sequence, there is a lack of tracking and updating of relevant information of targets at different times. That is, the model can clearly obtain the position of the target (such as obstacles) at each time point, but the current algorithm does not associate the target in time sequence, and the network cannot understand the corresponding relationship between the two-time targets, which reduces the interaction performance and leads to the instability of the graph network structure.

The understanding of semantic scenarios is not deep enough. The semantic understanding of scenes is to enable computers, like human brains, to correctly understand natural scenes and content. Many existing models or methods are limited to a small range of context information modeling, and can only learn local features, resulting in the semantic understanding of the scene is hindered.

5. Future outlook of pedestrian trajectory prediction

Learn and summarize various algorithms to improve the adaptability of the model. Pay attention to transfer learning and method promotion, and combine the different advantages of multiple prediction frameworks to achieve more reliable prediction. In the new environment, learn inductive models, mine and reason pedestrian motion patterns and collision avoidance rules and norms. At the same time, the prediction and control are combined to improve the overall robustness of the system.

Enhance context feature analysis and deeply understand semantic scenarios. The combination of context and surrounding environment can realize long-term prediction and help to explain pedestrian's motion intention. However, the context understanding of dynamic and static environment features and their semantics is still an unexplored field for better trajectory prediction. For scene understanding, it is necessary to make full use of the global context information and high-level semantic features to describe the scene content for the limitations of local context and other issues.

Expand access to data and enhance data fusion. Combined with the functional application of cameras, laser radars and other hardware sensors, data fusion is completed in cascade. Through software recognition algorithm or face recognition technology, accurately identify pedestrian posture, judge the subjective intention of pedestrians, and enhance the interpretability of the model, so as to make the perception and decision-making of the model more effective.

Table 5. Performance of each trajectory prediction method on ETH and UCY data set

Method	Reference	Evaluating indicator (ADE/FDE)					
		ETH	Hotel	Univ	Zara1	Zara2	Average
SoPhie	[13]	0.70/1.44	0.76/1.68	0.54/1.24	0.30/0.64	0.38/0.78	0.54/1.15
S-LSTM	[20]	1.09/2.35	0.79/1.76	0.67/1.42	0.47/1.02	0.56/1.18	0.72/1.56
SR-LSTM	[22]	1.05/2.21	0.81/1.68	0.71/1.45	0.47/1.02	0.64/1.25	0.74/1.52
S-attention	[41]	3.62/4.71	0.79/1.44	1.30/2.66	0.95/2.06	1.02/2.14	1.53/3.53
MX-LSTM	[42]	1.05/2.22	0.81/1.68	0.71/1.46	0.47/1.03	0.64/1.26	0.74/1.53
S-GAN	[28]	0.82/1.52	0.72/1.62	0.61/1.25	0.34/0.70	0.42/0.84	0.58/1.18
S-RNN	[30]	2.72/4.60	0.85/1.35	1.05/2.23	1.61/3.52	1.45/3.01	1.53/2.94
STG-GAN	[52]	0.64/1.12	0.79/0.85	0.45/0.78	0.34/0.54	0.31/0.48	0.44/0.75
Linear	[43]	1.33/2.94	0.39/0.72	0.82/1.59	0.62/1.21	0.77/1.48	0.78/1.57
Social ways	[55]	0.39/0.63	0.39/0.66	0.55/1.32	0.44/0.64	0.52/0.94	0.46/0.82
GAT	[47]	0.68/1.30	0.68/1.41	0.57/1.31	0.29/0.60	0.37/0.76	0.52/1.08
Social-BiGAT	[51]	0.69/1.31	0.49/1.02	0.55/1.33	0.31/0.62	0.36/0.75	0.48/1.02
STGAT	[54]	0.71/1.36	0.37/0.65	0.59/1.23	0.35/0.69	0.32/0.66	0.47/0.95
TPPO	[56]	0.84/1.74	0.24/0.47	0.42/0.95	0.34/0.76	0.26/0.61	0.42/0.90
NEXT	[60]	0.64/1.12	0.79/0.86	0.44/0.78	0.34/0.53	0.31/0.48	0.44/0.76

6. Conclusion

In recent years, with the progress of science and technology, the research of pedestrian trajectory prediction based on deep learning has made significant progress. However, due to the complexity and uncertainty of the interaction between pedestrians, pedestrians and the environment, pedestrian trajectory prediction still has some problems and challenges. Starting from the essence and challenges of trajectory prediction, this paper summarizes the work in the field of trajectory prediction this year, classifies the current pedestrian trajectory prediction methods based on the structural design and optimization of models, and summarizes the advantages and disadvantages of different algorithms. The research of pedestrian trajectory prediction involves many fields, such as automatic driving, behavior recognition, target detection and tracking, etc. It is not only a major challenge in the field of computer vision, but also plays a significant role in the development of artificial intelligence.

References

[1] Wang, Jinxiang, Dai, et al. Output-feedback robust control for vehicle path tracking considering different human drivers' characteristics[J]. Mechatronics: The Science of Intelligent Machines, 2018, 50:402-412.

[2] MOUSSAID M, PEROZO N, GARNIER S, et al. The walking behaviour of pedestrian social groups and its impact on crowd dynamics[J]. PLoS ONE, 2010, 5(4): e10047.

[3] LEFEVRE S, VASQUEZ D, LAUGIER C. A survey on motionprediction and risk assessment for intelligent vehicles[J]. Robomech Journal, 2014, 1(1): 1-14.

[4] Su Lumin Research on trajectory prediction method based on machine learning [D] Beijing: Beijing University of Posts and Telecommunications, 2019: 1-65.

[5] Helbing D, Molna´r P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.

[6] Ferrer G, Sanfeliu A. Behavior estimation for a complete framework for human motion prediction in crowded environments[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Hong Kong: IEEE, 2014: 5940-5945.

[7] Yan X, Kakadiaris I A, Shah S K. Modeling local behavior for predicting social interactions towards human tracking[J]. Pattern Recognition, 2014, 47(4): 1626-1641.

[8] Ye H R, Liu M Y, Zheng W S, et al. Trajectory prediction for multi-class target based on preferred speed[J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2017, 45(10): 100-104.

[9] Schneider N, Gavrila D M. Pedestrian path prediction with recursive bayesian filters: A comparative study[C].Proceedings of the German Conference on Pattern Recognition. Saarbrucken: Springer, 2013: 174-183.

[10] Schulz A T, Stiefelhagen R. A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction[C]. Proceedings of the IEEE International Conference on Intelligent Transportation Systems. Gran Canaria: IEEE, 2015: 173-178.

[11] Luo Y, Cai P, Hsu D, et al. GAMMA: A general agent motion prediction model for autonomous driving[J]. arXiv preprint arXiv:1906.01566, 2019.

[12] Chen Y F, Liu M, Everett M, et al. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017: 285-292.

[13] Sadeghian A, Kosaraju V, Sadeghian A, et al. Sophie: An attentive gan for predicting paths compliant to social and physical constraints[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 1349-1358.

[14] Kuhnt F, Schulz J, Schamm T, et al. Understanding interactions between traffic participants based on learned behaviors[C]. Proceedings of the IEEE Intelligent Vehicles Symposium. Gotenburg: IEEE, 2016: 1271-1278.

[15] Hong J, Sapp B, Philbin J. Rules of the road: Predicting driving behavior with a convolutional model of semantic interactions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 8454-8462.

[16] Sun L, Yan Z, Mellado S M, et al. 3D of pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Brisbane: IEEE, 2018: 5942-5948.

[17] Nachiket Deo, Mohan M. Trivedi et al Learning and Predicting On-road Pedestrian Behavior Around Vehicles[J]. ITSC,2017

[18] Raúl Quintero Mínguez, Ignacio Parra Alonso et al. Pedestrian Path, Pose, and Intention Prediction Through Gaussian Process Dynamical Models and Pedestrian Activity Recognition.TITS,2018

[19] Xue H, Huynh D Q, Reynolds M. Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction[C]. Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe: IEEE, 2018: 1186-1194.

[20] Alahi A, Goel K, Ramanathan V, et al. Social lstm: Human trajectory prediction in crowded spaces[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE CS, 2016: 961-971.

[21] Liu Q, Wu S, Wang L, et al. Predicting the next location: A recurrent model with spatial and temporal contexts[C]. Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016: 194-200.

[22] Zhang P, Ouyang W L, Zhang P F, et al. SR-LSTM: State refinement for lstm towards pedestrian trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 12077-12086.

[23] Kong W, Liu Y, Li H, et al. A survey of action recognition methods based on graph convolutional network[J]. Control and Decision, 2021, 36(7): 1537-1546.

[24] Liang J W, Jiang L, Niebles J C, et al. Peeking into the future: predicting future person activities and locations in videos[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE CS, 2019: 2960-2963.

[25] Sun J H, Jiang Q H, Lu C W. Recursive social behaviorgraph for trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE CS, 2020: 657-666.

[26] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//Proceedings of the European Semantic Web Conference. Berlin: Springer Press, 2018: 593-607.

[27] Ivanovic B, Pavone M. The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs[C]. Proceedings of the IEEE International Conference on Computer Vision. Seoul: CV/IEEE, 2019: 2375-2384.

[28] Gupta A, Johnson J, Li F F, et al. Social gan: Socially acceptable trajectories with generative adversarial networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE CS, 2018: 2255-2264.

[29] KELLER C G, GAVRILA D M. Will the pedestrian cross? A study onpedestrian path prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2): 494-506.

[30] SCHNEIDER N, GAVRILA D M. Pedestrian path prediction with recursive Bayesian filters: a comparative study[C]//Proceedings of the 35th German Conference on Pattern Recognition. Berlin: Springer Press, 2013: 174-183.

[31] PAVLOVIC V, REHG J M, MACCORMICK J. Learning switching linear models of human motion[C]//Proceeding of the Conference and Workshop on Neural Information Processing Systems. New York: Curran Associates Press, 2000: 981-987.

[32] FOX E B, SUDDERTH E B, JORDAN M, et al. Bayesian nonparametric inference of switching dynamic linear models[J]. IEEE Transactions on Signal Processing, 2011, 59(4): 1569-1585.

[33] KOOIJ J F P, SCHNEIDER N, FLOHR F, et al. Context-based pedestrian path prediction[C]//Proceedings of the 13th European Conference on Computer Vision. Berlin: Springer Press, 2014: 618-633.

[34] HELBING D, MOLNAR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.

[35] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. Computer Science, 2014.

[36] ALAHI A, RAMANATHAN V, LI F F. Socially-aware large-scale crowd forecasting[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 2211-2218.

[37] YI S, LI H S, WANG X G. Understanding pedestrian behaviors from stationary crowd groups[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3488-3496.

[38] RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning[M]. Cambridge: MIT Press, 2005.

[39] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[40] Lee N, Choi W, Vernaza P, et al. Desire: Distant future prediction in dynamic scenes with interacting agents[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2165-2174.

[41] Vemula A, Muelling K, Oh J. Social attention: Modeling attention in human crowds[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Brisbane: IEEE, 2018: 4601-4607.

[42] Hasan I, Setti F, Tsesmelis T, et al. MX-LSTM: Mixing tracklets and vislets to jointly forecast trajectories and head poses[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE CS, 2018: 6067-6076.

[43] Lisotto M, Coscia P, Ballan L. Social and scene-aware trajectory prediction in crowded spaces[C]. Proceedings of the IEEE International Conference on Computer Vision Workshop. Seoul: IEEE, 2019: 2567-2574.

[44] Varshneya D, Srinivasaraghavan G. Human trajectory prediction using spatially aware deep attention models[J]. arXiv preprint arXiv:1705.09436, 2017.

[45] Fernando T, Denman S, Sridharan S, et al. Soft+hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection[J]. Neural Networks, 2018, 108: 466-478.

[46] Li J, Ma H, Tomizuka M. Conditional generative neural system for probabilistic trajectory prediction[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019: 6150-6156.

[47] Xue H, Huynh D Q, Reynolds M. A locationvelocity-temporal attention lstm model for pedestrian trajectory prediction[J]. IEEE Access, 2020, 8: 44576-44589.

[48] Su H, Zhu J, Dong Y P. Forecast the plausible paths in crowd scenes[C]. Proceedings of the International Joint Conference on Artificial Intelligence. Melbourne, Australia, 2017: 2772-2778.

[49] Zhang Z Y, Diao Y H. Pedestrian trajectory prediction model with social features and attention[J]. Journal of Xidian University, 2020, 47(1): 10-17.

[50] Sun Y S, Jiang Q, Hu J, et al. Attention mechanism based pedestrian trajectory prediction generation model[J]. Journal of Computer Applications, 2019, 39(3): 668-674.

[51] Kosaraju V, Sadeghian A, Martín-Martín R, et al. Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks[C]. Proceedings of the Annual Conference on Neural Information Proceedingsessing Systems. Vancouver: NeurIPS, 2019: 1-10.

[52] Zou X Y, Sun B, Zhao D, et al. Multi-modal pedestrian trajectory prediction for edge agents based on spatial-temporal graph[J]. IEEE Access, 2020, 8: 83321-83332.

[53] Zhao T Y, Xu Y F, Monfort M, et al. Multi-agent tensor fusion for contextual trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 12118-12126.

[54] Sun H, Zhao Z Q, He Z H. Reciprocal learning networks for human trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE CS, 2020: 7414-7423.

[55] Amirian J, Hayet J B, Pettre J. Social ways: Learning multi-modal distributions of pedestrian trajectories with gans[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE CS, 2019: 2964-2972.

[56] Yang B, Yan G, Wang P, et al. TPPO: a novel trajectory predictor with pseudo oracle[J]. arXiv preprint arXiv:2002.01852, 2020.

[57] Zhong J Q, Sun H, Cao W M, et al. Pedestrian motion trajectory prediction with stereo-based 3d deep pose estimation and trajectory learning[J]. IEEE Access, 2020, 8: 23480-23486.

[58] Yan S J, Xiong Y J, Lin D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018:7444-7452.

[59] Yao H, Tang X, Wei H, et al. Modeling spatial-temporal dynamics for traffic prediction. arXiv 2018[J]. arXiv preprint arXiv:1803.01254.

[60] Huang Y F, Bi H K, Li Z X, et al. STGAG: Modeling spatial-temporal interactions for human trajectory prediction[C]. Proceedings of the IEEE International Conference on Computer Vision. Seoul: CV/IEEE, 2019: 6271-6280.

[61] Ivanovic B, Schmerling E, Leung K, et al. Generative modeling of multimodal multi-human behavior[C].Proceedings of the IEEE International Conference onIntelligent Robots and Systems. Madrid: IEEE, 2018:3088-3095.

[62] Mohamed A, Qian K, Elhoseiny M, et al. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE CS, 2020: 14412-14420.

[63] Wang C, Cai S, Tan G. Graphtcn: Spatio-temporal interaction modeling for human trajectory prediction[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3450-3459.

[64] Haddad S, Wu M, Wei H, et al. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[J]. arXiv preprint arXiv:1902.05437, 2019.

[65] Yu C J, Ma X, Ren J W, et al. Spatio-temporal graph transformer networks for pedestrian trajectory Prediction[C]. Proceedings of the European Conference on Computer Vision. Glasgow: Springer, 2020: 507-523.

[66] ROBICQUET A, SADEGHIAN A, ALAHI A, et al. Learning social etiquette: human trajectory prediction in crowded scenes[C]//Proceedings of the 14th European Conference on Computer Vision. Berlin: Springer Press, 2016: 549-565.

[67] Mangalam K, Girase H, Agarwal S, et al. It is not the journey but the destination: Endpoint conditioned trajectory prediction[C]//European conference on computer vision. Springer, Cham, 2020: 759-776.

[68] Zhu Y, Ren D, Fan M, et al. Robust trajectory forecasting for multiple intelligent agents in dynamic scene[J]. arXiv preprint arXiv:2005.13133, 2020.

[69] Xu Y Y, Piao Z X, Gao S H. Encoding crowd interaction with deep neural network for pedestrian trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE CS, 2018: 5275-5284.