EasyChair Preprint

№ 9345

Socially Acceptable Trajectory Prediction for Scene Pedestrian Gathering Area

Rongkun Ye, Zhiqiang Lv, Aite Zhao and Jianbo Li

November 21, 2022

Socially Acceptable Trajectory Prediction for Scene Pedestrian Gathering Area

Rongkun Ye[1], Zhiqiang Lv[1, 2], Aite Zhao[1]*, Jianbo Li[1]*

[1] College of Computer Science and Technology, Qingdao University, Qingdao 266071, China
[2] Institute of Ubiquitous Networks and Urban Computing, Qingdao University, Qingdao 266701, China
zhaoaite@qdu.edu.cn,lijianbo@ubinet.cn

Abstract. Dense areas of pedestrians in complex crowded scenes tend to disrupt the proper path of the agents. The agents usually avoid gathering areas to find a reasonable pedestrian-sparse path, slow down the speed to walk, and wait for the gathering pedestrians to disperse. The accurate trajectory prediction in gathering areas is a challenging problem. This work introduces a new feature that affects trajectories to address this problem. The area gathering feature that allows agents to plan future paths based on the gathering level of pedestrians. The gathering areas as well as indicate the degree of gathering in the areas by means of a dynamic pedestrian filtering method to generate a trajectory heat map. Besides, the convolutional neural network is used to extract the corresponding area gathering feature. Furthermore, a new approach is proposed for inter-agent interactions that makes full excavation of deep interaction information and takes into account a more comprehensive interaction behavior. This work predicts trajectories by incorporating multiple factors such as area-dense features, social interactions, scene context, and individual intent. The prediction accuracy is significantly enhanced and outperforms state-of-the-art methods.

Keywords: Trajectory Prediction, Gathering Areas, Trajectory Heat map, Social Interactions.

1 Introduction

Trajectory prediction predicts the path for a while in the future by studying the motion behavior of the agent. For pedestrians, the uncertainty of personal intentions, the complexity of social relationships among pedestrians, and the variability of pedestrians' surrounding environment. This makes the prediction task challenging.

In trajectory prediction, some researches [1, 2] use a social interaction pool to model the interaction between pedestrians, considering the influence of other pedestrians on the target pedestrian from the local and global aspects of the scene, respectively. However, the proposed social interaction model often identifies incorrect interaction agents. Sadeghian et al. [3] encoded the interactions between agents by a more reliable feature extraction strategy from a bird's eye view to learn the scene context. And the proposed attention mechanism to combine scene context [4, 5] and social interaction to generate

accurate interpretable social and physical feasible paths. However, the social attention mechanism is unable to memorize the social interactions of long-time pedestrians. Sun et al. [6] has built more interpretable social interaction graphs based on the relationships between pedestrians. And the deep social interaction features are extracted by graph convolutional neural network. However, the building of social interaction graphs consumes a large amount of time and resources. Besides, some work [7, 8] has used the novel thought of dividing the prediction into target points and way lines, improving the accuracy of prediction significantly.

Fig. 1. GA-GAN uses the trajectory heat map to highlight gathering areas, and combines the the agent interaction algorithm to predict a reasonable trajectory.

This work proposes the GA-GAN method to solve the above problems. As in Fig. 1, GA-GAN considers the influence of the pedestrian gathering area on the agent's future path. Using the method of trajectory heat map to highlight gathering areas, thereby there are more attention for them in the procession of model prediction. And making full excavation of deep interaction information, GA-GAN simulates complex social interactions by pedestrians considering each agent's distance and intention. GA-GAN makes the prediction of each agent's path more reasonable and accurate by combining the information of scene context, pedestrian gathering areas, and social interactions between agents. The main contributions of this work are as follows:

- This work innovatively proposes trajectory prediction based on the fusion of gathering area features with other features and achieve excellent results.
- This work proposes a method that generates a trajectory heat map to represent the gathering areas and proposes the dynamic pedestrian filtering method (DPFM) fuzzy motion pedestrian to solve the problem of gathering area uncertainty.
- This work proposes a comprehensive and interactive approach to social interaction.
- The GA-GAN is the state-of-the-art model on the ETH/UCY dataset.

2 Related Work

2.1 Gathering Area

The study of gathering areas is generally applied in the fields of urban traffic, trajectory hotspot area discovery, and area gathering density analysis. Most of the research methods of gathering areas involve several traditional methods (K-means and some of its improved methods). [9] have detected scene hotspots by a two-stage clustering ap-

proach, in which spatial clustering of trajectory points with spatial and temporal attributes [10-12] was applied using spatial clustering based on temporal density. Choi et al. [13] have generated a heat map based on the relational features gathered by agents in areas, and predicts the relational features of the future target location in the heat map form. In this paper, the GA-GAN generates a trajectory heat map with time dependence to represent the gathering level of an area of pedestrians at a certain period. Then combine other features to make the model predict a more accurate and reasonable path.

2.2 Generative Modeling

Goodfellow et al. [14] have proposed a new type of generative model, generative multifunctional networks, which is trained as a very small game between the generative and discriminative models. In Gupta et al. [2], the S-GAN has been proposed to combine GAN with trajectory prediction for the first time and proposed a social interaction model with multivariate losses to encourage GAN generative networks to extend their normal distribution and cover the space of possible paths. However, since S-GAN does not fully utilize the deep interaction information of pedestrians in the social interaction model. Many approaches have improved it by using attention mechanisms [15], adding feasibility constraints, and learning more accurate sample distribution to synthesize pedestrian interaction models and explore the influencing factors of pedestrian trajectories. In this paper, the GA-GAN makes full use of social interaction information and incorporates an attention mechanism to enhance scene interaction.

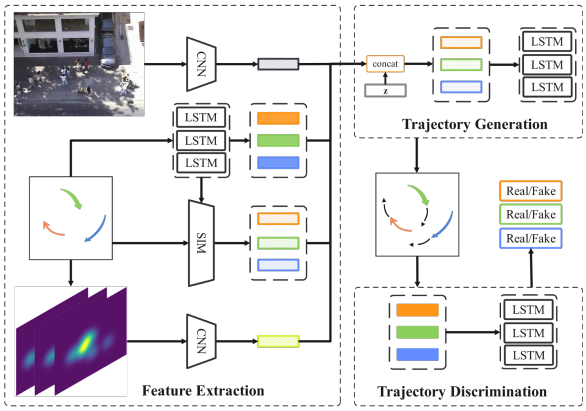


Fig. 2. Structure figure of GA-GAN, which consists of a feature extraction module, a trajectory generation module and a trajectory discrimination module.

3 Method

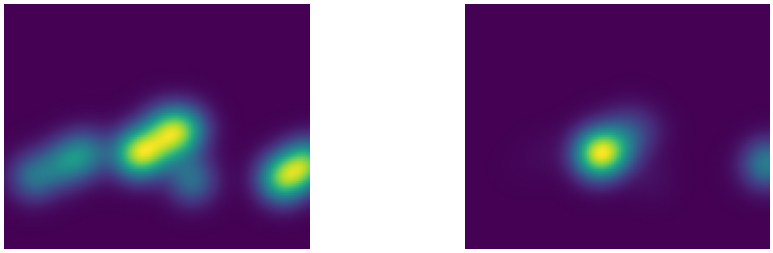
3.1 Problem Definition

Trajectory prediction is the process of learning the first obs time walker trajectories to predict the future path of the next $pred$ time steps. This work defines the model input

as $\{X_t^i\}\ i \in N, t \in [1, obs]$, i.e., the trajectory coordinates of the i-th person in the current scene at the observed t-th time step, where $X_t^i = (x_t^i, y_t^i)$. The model output is $\{\hat{Y}_t^i\}\ i \in N, t \in [1, pred]$, In addition, we define the true trajectory coordinates of the latter $pred$ time step as $\{\bar{Y}_t^i\}\ i \in N, t \in [1, pred]$.

3.2 GA-GAN

The proposed model is based on the GAN consisting of a generator and a discriminator. as shown in Fig. 2. The generator consists of a feature extraction module and a trajectory generation module, which are continuously trained to learn the path distribution of the agent and generate reasonable future trajectory samples for the agent. The discriminator consists of an LSTM-based encoder, which distinguishes whether the generated agent trajectory is feasible or not.



(a) (b)

Fig. 3. The (a) represents the heat map of the trajectory generated without the DPFM. The (b) represents the heat map of the trajectory generated with the DPFM.

Feature Extraction.

Area Gathering Feature. We use the deep learning [16, 17] method to obtain the influence of the gathering area on the agent. Firstly, the trajectory coordinates X_t of pedestrians in space need to be pre-processed, which can prevent the problem that the generated spatial heat map is inconsistent with the original scene coordinate distribution due to the inconsistent coordinate scaling of different scenes.

Next, we represent the spatially gathered areas by transforming the trajectory coordinates of pedestrians into a trajectory heat map S_t, as in Eq. 1. The S_t is of fixed size, with width W and height H respectively. $GauKe$ is a method we propose to generate a Gaussian kernel centered at the k-th pedestrian trajectory coordinate X_t^k at time step t_s, with r as the scope of influence of the agent k. The n_{ped} is all the pedestrians in the scene at the current time steps.

$$S_t = \frac{\sum_{k=1}^{k\epsilon[1,n_{ped}]} GauKe(X_t^k, r)}{n_{ped}} \tag{1}$$

This work proposes DPFM to solve the uncertainty of the gathering area of walking people at the current time by blurring the moving pedestrians. The DPFM reduces the impact of pedestrians with a speed greater than v on the scene to highlight the areas where people are gathered, making the model easier to extract the features of the gathering areas. Here v is a threshold value. If the moving speed v_{mov} is above v, the r of

$GauKe$ function is reduced by 2-1/$Sigmoid(v_{mov}\text{-}v)$ times. The effect is shown in Fig. 3. Besides, we take the trajectory heat map connection of the first t_s time steps to prevent the gathering areas from changing. At last, as Eq. 2 using convolutional neural network to down-sample the trajectory heat map to extract the area gathering features.

$$A_t = CNN(S_{t-t_s:t}; W_{cnn}) \tag{2}$$

Individual Behavioral Feature. The GA-GAN uses a long short-term memory (LSTM) network as an encoder to capture the temporal dependencies of the observed agent trajectory coordinates. The previous t-step trajectory coordinate of agent i X_t^i is then input to the $LSTM_{en}$ to obtain the behavioral features B_t^i, as in Eq. 3.

$$B_t^i = LSTM_{en}(X_{:t}^i, h_{en}{}_{t-1}^{\ i}; W_{en}) \tag{3}$$

Environmental Feature. The GA-GAN uses VGG-16 to process each frame of the video I to extract the environmental features E_t, and uses a self-attention mechanism to emphasize areas with high impact, as in Eq. 4.

$$E_t = ATTEN\left(VGG-16(I_t, h_t; W_{vgg-16})\right) \tag{4}$$

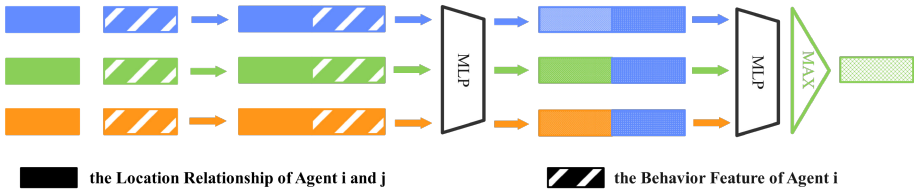


Fig. 4. The internal structure diagram of SIM about agent i

Social Interaction Feature. The social interaction of the agent is related to the spatial relationship between the agents in the scene and the intention of the agents. The social interaction module (SIM) of GA-GAN obtains the social interaction representation of agent i based on the location relationship L_{rel} between agent i and other agents and the personal intention of agent i to get the social interaction representation of agent i, as shown in Eq. 5. In SIM, the influence of other agents on agent i in terms of position and behavior is inferred from the relative position of agent i with other agents and their behavior, as shown in Fig. 4. Then the pedestrian with the greatest influence among the influence of other agents over agent i based on the behavior of agent i is selected as the interaction object of agent i.

$$S_t^i = SIM(L_{rel}{}_t^{\ i}, B_t^i; W_{SIM}) \tag{5}$$

Trajectory Generation.

The trajectory generation fuses the above features and white noise vector z sampled from a multivariate normal distribution to obtain a multi-source feature that can describe and constrain agents. The multi-source features are fed into the multilayer perceptron to obtain a semantic vector, which adequately represents the multi-source information. And then the vector is input into the LSTM-based decoder, which generates

reasonable paths that conform to the realistic trajectory pattern and are robust by learning the relationship between the multi-source features of agent i in the temporal dimension, as shown in Eq. 6.

$$\hat{Y}_{t+1}^{i} = Generation([B_t^i, A_t, E_t, S_t^i, z_t^i], {h_{gen}}_t^i; W_{gen}) \quad (6)$$

Trajectory Discrimination.
The trajectory discrimination consists of an LSTM-based encoder and a MLP. The encoder estimates the time-dependent future state of the trajectory of pedestrian i by learning the distribution of the trajectory of the input pedestrian i. The MLP discriminates the trajectory of pedestrian i based on the temporal dependencies of the trajectory of pedestrian i estimated by the encoder. Discriminator final output C^i, as shown in Eq. 7. If C^i is closer to 1 then the input trajectory is more like the real trajectory, else is closer to 0 then the input trajectory is more like the faked trajectory.

$$C^i = Discriminator(Y^i, {h_{dis}}^i; W_{dis}) \quad (7)$$

4 Experiment

4.1 Dataset and Implementation Details

In this work, we use ETH/UCY dataset to test the interaction and capture function to the aggregation area of the model, and then evaluate the accuracy and rationality of the model. ETH/UCY dataset has a large number of rich interaction scenarios for everyone, such as gathering, following, pooling, and collision. We sample coordinate points in meters for pedestrians every 0.4s, and use Average Displacement Error (ADE) and Final Displacement Error (FDE) as evaluation metrics.

This work implement GA-GAN in PyTorch, and perform all experiments with Nvidia 3090 GPUs. The default H and W of trajectory heat map S_t are 144 and 180 respectively, and t on S_t is 3. In addition, the r in *GauKe* is 40, the v in DPFM is 0.5. We use the Adam optimizer with default parameters and initial learning rate 1×10^{-3}.

4.2 Quantitative Analysis

Table. 1. Results of GA-GAN with baselines under the ETH/UCY dataset with a prediction time of 3.2s or 8 time steps.

Model	**S-LSTM**		**S-GAN**		**SoPhie**		**Y-net**		**Ours**	
Metric	**ADE**	**FDE**	**ADE**	**FDE**	**ADE**	**FDE**	**ADE**	**FDE**	**ADE**	**FDE**
ETH	1.09	2.35	0.81	1.52	0.70	1.43	0.28	**0.33**	**0.23**	0.34
HOTEL	0.79	1.76	0.72	1.61	0.76	1.67	**0.10**	**0.14**	0.17	0.25
UNIV	0.67	1.40	0.60	1.26	0.54	1.24	0.24	0.41	**0.14**	**0.23**
ZARA1	0.47	1.00	0.34	0.69	0.30	0.63	0.17	0.27	**0.14**	**0.22**
ZARA2	0.56	1.17	0.42	0.84	0.38	0.78	**0.13**	0.22	**0.13**	**0.18**
AVG	0.72	1.54	0.58	1.18	0.54	1.15	0.18	0.27	**0.16**	**0.24**

We compare GA-GAN against several similar and SOTA baselines (S-LSTM, S-GAN, SoPhie, Y-net) which have introduced in the chap 1.

In the UCY/ETH dataset, the results of comparing GA-GAN with baselines are shown in Table. 1. The effect of GA-GAN is significant in UNIV, ZARA1, and ZARA2 scenes with pedestrian gathering and complex interactions, which all outperform other models. Especially in UNIV, ADE, and FDE are significantly decreased compared with baselines. As UNIV has the most and densest pedestrians in all scenes, which helps GA-GAN focus more on the interactions between agents and the gathering areas, thus enabling GA-GAN to extract more important interaction features and gathering features. In addition, the environmental feature of the interaction between self-attention pedestrians and the environment extracted by GA-GAN play a guiding role in predicting the trajectory. The combination of these features can generate a more reasonable trajectory in the prediction process.

4.3 Qualitative Analysis

In this section, we investigate the ability of GA-GAN to model pedestrian interaction and perceive gathered areas. First two subsections investigate the impact of region gathering features extracted by GA-GAN on future trajectories. Last subsection investigates the impact of GA-GAN's improved pedestrian interaction method on future trajectories. We select ZARA with high pedestrian gathering and strong interactive behavior.

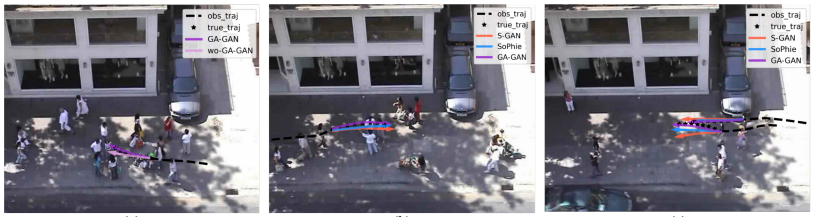


Fig. 5. Comparative graph of qualitative analysis results

Gathering Versus No-Gathering.

In this subsection, we investigate the impact of the presence or absence of gathering modules on trajectory prediction, to understand the importance of aggregation modules on GA-GAN prediction. In this scenario, pedestrians pushing baby strollers choose to bypass the gathering crowd when they encounter the gathering pedestrians in front of them. The prediction results are shown in Fig. 5 (a). It is clear that the GA-GAN identifies and perceives the gathering area in advance by the DPFM, and combines the constraints of the surrounding environment with the behavior implication of the surrounding agents, so as to focuses its prediction more on the pedestrians gathered in front. Therefore, GA-GAN adjusts the motion trajectory to avoid collision with the forward gathering pedestrians before the agent passes the gathering area. However, the GA-GAN with no gathering module (wo-GA-GAN) does not notice the forward gathering area. Therefor the gathering module of GA-GAN has a powerful function of perceiving

the gathering area in advance. And then combine with multiple feature information to make the corresponding behavior of avoiding or approaching the gathering area.

Gathering Module.
In this subsection, we choose S-GAN and SoPhie, which are based on GAN but do not include the gathering module, to compare with GA-GAN. In this scenario, the labeled agents change their walking direction to avoid pedestrians gathered in front of them when they encounter them. The prediction results are shown in Fig. 5 (b). Thanks to the inclusion of dynamic crowd gathering area feature extraction and DPFM, the GA-GAN pre-perceives the gathering area and makes avoidance behavior. However, the S-GAN, which only considers interactions related to distance and individual behavior, does not perform well with SoPhie, which fails to balance physical constraints with social interactions. Therefore, the trajectories predicted by the gathering module in complex and pedestrian-aggregated scenes are significantly enhanced in terms of both accuracy and plausibility.

Social Interaction Module.
In this subsection, we compare SGAN and SoPhie with interaction modules against GA-GAN. In this scenario, a step of pedestrian walks slowly waiting for another pedestrian, and converges with him. The effect is shown in Fig. 5 (c). As GA-GAN uses a social interaction module with mutual influence between agents' behaviors, it can accurately measure the influence of the agent and other agents in location and behavior on this agent, and then determine the corresponding interaction behavior and the interacting pedestrians. The SoPhie measures the importance of distance and interaction by the self-attention mechanism. In this scene, the SoPhie tends to interact with the rear pedestrians but the tendency is tiny. The S-GAN selects the pedestrian in the global interaction who has the most influence on the target pedestrian in terms of distance and behavior, but not the rear converging pedestrians. Therefore, GA-GAN has greater interaction capability and easier to observe the interaction between pedestrians.

5 Conclusion

This paper proposes a trajectory heat map representation of the gathering areas to solve the problem of failing to model complex relationships in crowded scenes. Besides, we propose a more interactive method for pedestrians to consider each other's intentions. This solves the problem of inconsistent interaction goals caused by both agents not considering each other. Experiments show that the GA-GAN is a state-of-the-art method.

Acknowledgement

National Natural Science Foundation of China under Grant No.62106117, and Shandong Provincial Natural Science Foundation under Grant No.ZR2021QF084.

References

1. Alahi, A., Goel, K., Ramanathan, V., et al.: Social lstm: Human trajectory prediction in crowded spaces. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 961-971. IEEE, New York, USA (2016).
2. Gupta, A., Johnson, J., Fei-Fei, L., et al.: Social gan: Socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2255-2264. IEEE, Salt Lake City, USA (2018).
3. Sadeghian, A., Kosaraju, V., Sadeghian, A., et al.: Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1349-1358. IEEE, Long Beach, USA (2019).
4. Lv, Z., Li, J., Li, H., et al.: Blind travel prediction based on obstacle avoidance in indoor scene. Wireless Communications and Mobile Computing, 1-14 (2021).
5. Jiang, B., Li, Y.: Construction of Educational Model for Computer Majors in Colleges and Universities. Wireless Communications and Mobile Computing, 1-9 (2022).
6. Sun, J., Jiang, Q., Lu, C.: Recursive social behavior graph for trajectory prediction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 660-669. IEEE, Seattle, USA (2020).
7. Mangalam, K., An, Y., Girase, H., et al.: From goals, waypoints & paths to long term human trajectory forecasting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 15233-15242. IEEE, Montreal, Canada (2021).
8. Wang, C., Wang, Y., Xu, M., et al.: Stepwise goal-driven networks for trajectory prediction. IEEE Robotics and Automation Letters, 1-11 (2022).
9. Li, F., Shi, W., Zhang, H.: A two-phase clustering approach for urban hotspot detection with spatiotemporal and network constraints. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 3695-3705 (2021).
10. Lv, Z., Li, J., Dong, C., et al.: DeepSTF: A deep spatial–temporal forecast model of taxi flow. The Computer Journal, 1-16 (2021).
11. Cheng, Z., Rashidi, T H., Jian, S., et al.: A Spatio-Temporal autocorrelation model for designing a carshare system using historical heterogeneous Data: Policy suggestion. Transportation Research Part C: Emerging Technologies 141, 103758 (2022).
12. Wang, Y., Lv, Z., Zhao, A., et al.: A deep spatio-temporal meta-learning model for urban traffic revitalization index prediction in the COVID-19 pandemic. Advanced Engineering Informatics, 1-17 (2022).
13. Choi, C., Dariush, B.: Looking to relations for future trajectory forecast. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 921-930. South Korea (2019).
14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. Advances in neural information processing systems 27, 1-10 (2014).
15. Song, Y., Bisagno, N., Hassan, S. Z., et al.: Ag-gan: An attentive group-aware gan for pedestrian trajectory prediction. In: 2020 25th International Conference on Pattern Recognition, pp. 8703-8710. IEEE, Milan, Italy (2021).
16. Lv, Z., Li, J., Dong, C., et al.: Deep learning in the COVID-19 epidemic: A deep model for urban traffic revitalization index. Data & Knowledge Engineering 135, 101912 (2021).
17. Zhao, A., Wang, Y., Li, J.: Transferable Self-Supervised Instance Learning for Sleep Recognition. IEEE Transactions on Multimedia, 1-15 (2022).