



Robust Trajectory Prediction of Multiple Interacting Pedestrians via Incremental Active Learning

Yi Xi¹, Dongchun Ren¹, Mingxia Li², Yuehai Chen², Mingyu Fan³(✉),
and Huaxia Xia¹

¹ Meituan, Beijing 100102, China

{xuyi14, rengdongchun, xiahuaxia}@meituan.com

² Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

{alichelmx, cyh0518}@stu.xjtu.edu.cn

³ Wenzhou University, Wenzhou 325035, Zhejiang, China

fanmingyu@amss.ac.cn

Abstract. Pedestrian trajectory prediction is an important issue in many real applications, including autonomous driving, robot navigation, and intelligent monitoring. With rapid growing volume of pedestrian trajectory data, existing methods roughly learn pedestrian walking motion directly with increasing computation and time costs while neglecting checking the relative importance of the trajectory data. In order to address this issue, we propose a novel trajectory prediction model via incremental active learning, which is referred as “IAL-TP”. In this method, we utilize a simple and effective strategy to evaluate the candidate data samples and then select the more valuable and representative samples. An active set is determined by our proposed strategy such that both noisy and redundant samples are not selected. The active learning strategy is implemented iteratively to improve the generalization ability of the model. Experimental results on benchmark public datasets demonstrate that our model is able to achieve better performance than state-of-the-art methods with only a small fraction of the training data.

Keywords: Trajectory prediction · Incremental active learning · Select strategy

1 Introduction

Future path prediction, which aims at forecasting the future trajectories of multiple agents in the next few seconds, has received a lot of attention in the multimedia community [27, 32]. This is a fundamental problem in a variety of applications such as autonomous driving [7], long-term object tracking [21], monitoring, robotics, etc. Recently, Recurrent Neural Network (RNN) and its variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have demonstrated promising performance in modeling the trajectory sequences [1, 14].

Y. Xu and D. Ren are co-first authors. This work was done while Y. Xu was an intern in Meituan.

© Springer Nature Switzerland AG 2021

T. Mantoro et al. (Eds.): ICONIP 2021, CCIS 1516, pp. 141–150, 2021.

https://doi.org/10.1007/978-3-030-92307-5_17

Trajectory prediction is difficult because of its intrinsic properties: (1) The pedestrians walking in public often interact with each other and will change the walking paths to avoid collision or overtaking. (2) The pedestrians may follow several viable trajectories that could avoid collision, such as moving to the left, right, or stopping. Because of these properties, some trajectory prediction methods are proposed by modeling social interactions [1, 11, 23, 35]. Some methods include additional surrounding semantics maps into their model such that the predicted trajectories will comply the restrictions in the maps [13, 25, 26]. Given the observed trajectories of agents in a scene, a pedestrian may take different trajectories depending on their latent goals. Because of future is uncertain, many researches focus on multiple future trajectory prediction [11, 14, 18]. Some recent works also proposed probabilistic multi-future trajectory prediction [23, 28], which provide very useful quantity results. Moreover, some vision-based trajectory prediction methods apply raw camera data and LiDAR point cloud data directly for interaction modeling and trajectory prediction [12, 22].

In practice, one can drive a car with a LiDAR [17] or use overhead surveillance cameras on open road to collect as many as possible trajectories of the road users. However, not all trajectory data are helpful in training a robust and accurate prediction model. Some observed trajectories are noisy because of the car movements in some scenes are relatively simple where the road users are moving at constant speeds. In this study, we propose a data efficient trajectory prediction active learning method through the selection of a compact less noisy and more informative training set from all the observed trajectory data. To the best of our knowledge, it is the first time considering the trajectory prediction problem from the perspective of trajectory samples. The main contributions of this study are summarized as follows.

- This study proposes a simple and efficient active learning strategy, which could remove noisy and redundant trajectory data for a compact and informative training set. The storage and computation costs at training stage are greatly reduced.
- Our proposed method could actively learn the streaming trajectory data incrementally and efficiently, which is the first work that consider the value of trajectory data in trajectory prediction task.
- Our proposed active prediction method is able to achieve better performance than the previous state-of-the-arts with much smaller training dataset on five public pedestrian trajectory datasets.

2 Related Work

With rapid development of deep learning, various methods have been proposed. Social-LSTM [1] is one of the earliest methods of applying Recurrent Neural Network for pedestrian prediction. In Social-LSTM, a pooling layer is designed for sharing human-human interaction features among pedestrians. Later work [4, 19, 30, 34] followed this pattern, design different approaches for delivering information of human-human interactions. Instead of making only one determined trajectory of each pedestrian, Generative Adversarial Networks-based (GAN) methods [2, 9, 13, 16, 25, 29] has been designed for multiple plausible trajectories prediction. Moreover, auto-encoder-based methods [5, 20] have been developed for encoding important features of pedestrians and then making predictions with a decoder. Due to the big success of

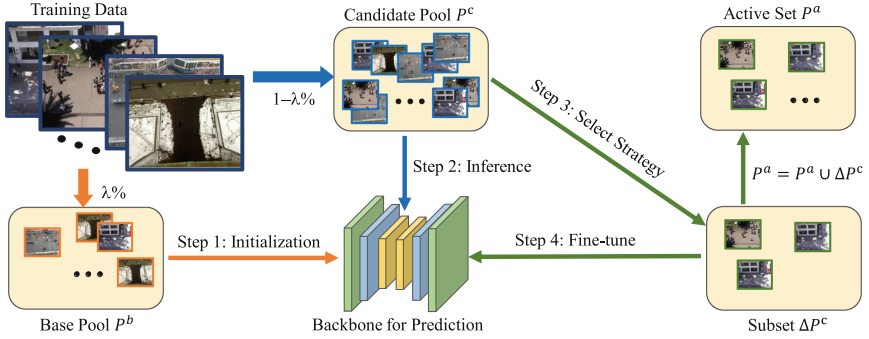


Fig. 1. The structure of our proposed IAL-TP model. At first, the model is initialized with samples from the base pool. At each iteration, a subset is selected out of the candidate pool based on the model inference and then added into the active set. Once the number of trajectory samples in active set meets the requirement, retrain the model with samples from the collected active set.

Transformer structure [29] in sequential processing [6]. Recent works [10, 33] seek to utilize this structure for pedestrian trajectory prediction and achieve competitive performance. However, these methods roughly utilize all the available trajectory data to attempt to understand the movement pattern for future trajectory prediction. We argue that not all available data are useful or meaningful during training and blindly using such large amount of data could damage the performance of models, not to mention the expensive computation and time costs.

3 Model

In this section, we first give the problem definition and then detailed introduce our proposed model. The pipeline of our IAL-TP method is illustrated in Fig. 1.

3.1 Construct Two Pools

In real-world applications, trajectory data are easy to be collected automatically by sensors such as LiDAR and cameras. Previous methods save all the collected trajectories and then train their models on the collected data. However, the raw trajectory data are noisy and redundant and it is unnecessary to save and train all the collected data. Some collected trajectories fluctuate drastically with time and some trajectories are straight lines with constant speeds. The fluctuation trajectories are noisy and the straight trajectories with constant speeds are redundant and too easy for the prediction model. To address this issue, we begin with constructing two non-intersect pools, a base pool \mathcal{P}^b only with a small amount of trajectory data for model initial learning, and a candidate pool \mathcal{P}^c with the remaining trajectory data for incremental active learning. Following the above problem formulation, the whole training trajectory data is $\Gamma^{obs} = \{\Gamma_i^{obs} | \forall i \in \{1, 2, \dots, N\}\}$, define the base pool $\mathcal{P}^b = \{\Gamma_j^{obs} | \forall j \in \mathcal{S}^b\}$, the candidate pool $\mathcal{P}^c = \{\Gamma_k^{obs} | \forall k \in \mathcal{S}^c\}$, where \mathcal{S}^b and \mathcal{S}^c are two disjoint subsets

Algorithm 1. Incremental Active Learning

Input: Untrained model M^0 , whole trajectory data Γ_{obs} .
Input: Hyper-parameters: $\lambda\%$, $p\%$, learning rate, etc.
Output: Trained model M^a .

- 1: Construct base pool \mathcal{P}^b and candidate pool \mathcal{P}^c following the percentage $\lambda\%$.
- 2: Train an initial model M^b using the trajectory samples of base pool \mathcal{P}^b .
- 3: Define an active set $\mathcal{P}^a = \emptyset$. Define N^a is the number of trajectory samples of \mathcal{P}^a . Calculate the expected number \hat{N}^a of trajectory samples with Equation 1.
- 4: **while** $N^a \leq \hat{N}^a$ **do**
- 5: Infer all trajectory data in candidate pool \mathcal{P}^c through model M^b .
- 6: Select a subset $\Delta\mathcal{P}^c$ from \mathcal{P}^c based on our incremental select strategy.
- 7: Fine-tune model M^b with the subset $\Delta\mathcal{P}^c$.
- 8: $\mathcal{P}^c \leftarrow \mathcal{P}^c - \Delta\mathcal{P}^c$, $\mathcal{P}^a \leftarrow \mathcal{P}^a \cup \Delta\mathcal{P}^c$.
- 9: Update N^a .
- 10: **end while**
- 11: Train model M^0 with the active set \mathcal{P}^a until convergence is realized.
- 12: **return** Trained model M^a .

and $\mathcal{S}^b \cup \mathcal{S}^a = \{1, 2, \dots, N\}$. We randomly select $\lambda\%$ trajectory data from the whole training data as the base pool \mathcal{P}^b and we set $\lambda = 5$ in our work.

3.2 Incremental Active Learning

Instead of learning from whole trajectory samples, we propose an active learning method to incrementally select partial “worthy” trajectory data from candidate pool merging in the active set \mathcal{P}^a , and then utilize these more valuable samples for model learning. Denote we expect to select $p\%$ trajectory samples from candidate pool, \hat{N}^a is the expected number of trajectory samples of active set, which is defined as follow:

$$\hat{N}^a = p\% \times N \quad (1)$$

At each iteration, we infer all the trajectory samples from candidate pool \mathcal{P}^c and rank these samples base on their inference errors, and choose the median ones as the subset $\Delta\mathcal{P}^c$. According to [3], the larger the error, the more noise the sample will have, the smaller the error is, the easier the sample can be learned. Therefore, our proposed select strategy is to select median trajectory samples merging in the active set. These median samples have less noise, and at the meantime, they are more representative than those with smaller error. In the experiments, we have explored different select strategies to shown the effectiveness of our proposed selection. Note that the subset will be removed from the candidate pool once selected. We iterate the above selection steps until the number of active set N^a equals to \hat{N}^a . The overall incremental active learning method is illustrated in Algorithm 1 followed with detailed explanation.

Beginning with the untrained model M^0 , whole observed trajectories Γ_{obs} , and hyper-parameter $\lambda\%$, we firstly construct the base pool and candidate pool (line 1), and then train an initial model M^b with trajectory samples from base pool \mathcal{P}^b (line 2). Before starting iteration, we define an empty active set \mathcal{P}^a , and calculate the number

Table 1. Experimental results of our proposed IAL-TP model with different $p\%$. We use four different $p\%$ to evaluate the influence of different numbers of trajectory samples that participate in the model training phase.

	Performance (ADE/FDE)					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	Average
$p\% = 100\%$	0.56/1.06	0.52/0.91	0.57/1.20	0.41/0.90	0.33/0.74	0.48/0.96
$p\% = 50\%$	0.56/1.04	0.42/0.89	0.59/1.29	0.42/ 0.89	0.35/0.76	0.46/0.95
$p\% = 33\%$	0.69/1.20	0.52/1.04	0.63/1.33	0.46/1.00	0.40/0.86	0.54/1.09
$p\% = 25\%$	0.75/1.25	0.60/1.18	0.63/1.34	0.56/1.17	0.39/0.85	0.59/1.16

N^a of the trajectory samples in \mathcal{P}^a . Also, we calculate the expected number \hat{N}^a of trajectory samples for final model learning (line3). At each iteration, we first inference all the samples of the candidate pool through model M^b and obtain the errors of these samples (line 5). Then, we select a batch of samples with median errors as the subset $\Delta\mathcal{P}^c$ from the candidate pool (line 6). Afterwards, the model M^b is fine-tuned with the subset $\Delta\mathcal{P}^c$ to ensure the model learn well on this subset, and thus avoid selecting similar samples at the next iteration (line 7). Finally, we update the \mathcal{P}^c , \mathcal{P}^a , and calculate the new number N^a in \mathcal{P}^a (line 8, 9). When the number N^a equals to our expected \hat{N}^a , the iteration is finished, and we obtain an active set \mathcal{P}^a within more valuable and representative trajectory samples. We retrain the model M^0 with the active set until convergence is realized to return the final model M^a .

3.3 Backbone for Prediction

In order to make accurate trajectory predictions, we utilize our previous state-of-the-art framework [31], which is able to extract global spatial-temporal feature representations of pedestrians for future trajectory prediction.

4 Experiments

We demonstrate the experimental results on two public datasets: ETH [24] and UCY [15]. We observe 8 frames and predict next 12 frames of trajectories.

4.1 Evaluation Metrics

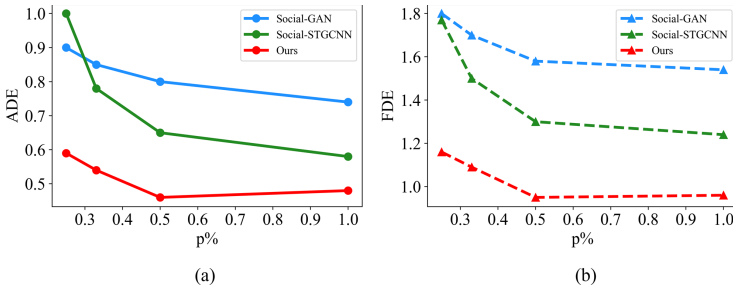
Similar with other baselines, we use two evaluation metrics: Average Displacement Error (ADE) and Final Displacement Error (FDE).

4.2 Data Efficiency and Model Robustness

In our work, the most important hyper-parameter is the percentage $p\%$ denoting the number \hat{N}^a of trajectory samples in the active set. The number \hat{N}^a represents the number of trajectory samples that are learned by the model at the training phase. Note that

Table 2. Experimental results of baselines compared to our proposed model. Original Social-STGCNN model is a probabilistic model and we try our best to adapt it to a deterministic model.

Method	Performance (ADE/FDE)					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	Average
Linear [1]	1.33/2.94	0.39/0.72	0.82/1.59	0.62/1.21	0.79/1.59	0.79/1.59
Social-LSTM [1]	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
Social-GAN [11]	1.13/2.21	1.01/2.18	0.60/1.28	0.42/0.91	0.52/1.11	0.74/1.54
TPNet [8]	1.00/2.01	0.31/0.58	0.55/1.15	0.46/0.99	0.33/0.72	0.71/1.08
TF-based [10]	1.03/2.10	0.36/0.71	0.53/1.32	0.44/1.00	0.34/0.76	0.54/1.17
Social-STGCNN*	0.98/2.10	0.54/1.10	0.57/1.20	0.46/1.00	0.37/0.78	0.58/1.24
IAL-TP (Ours)	0.56/1.04	0.42/0.89	0.59/1.29	0.42/0.89	0.35/0.76	0.46/0.95

**Fig. 2.** Model performance versus part of the training trajectory samples. The x-axis shows the percentage of trajectory samples used. The y-axis shows the corresponding ADE/FDE error.

$p\% = 100\%$ means all trajectory samples are used in the candidate pool for model learning, which is the same with the existing baselines. Table 1 shows the performance of our proposed model with different $p\%$.

We can observe that when $p\% = 50\%$, namely using only half of the trajectory samples from the candidate pool, our proposed IAL-TP model achieves the best performance with the smallest error. It indicates that there are a lot of redundant trajectory samples in the datasets and it also validates the necessity of our proposed active learning idea. In specific, the ADE error on dataset ETH is the same with the result with $p\% = 100\%$, the ADE error on dataset HOTEL with $p\% = 50\%$ outperforms the result with $p\% = 100\%$ by 19.2%, which is a significant improvement. One possible reason is that the datasets ETH and HOTEL are relatively more crowded than other datasets [1]. It reflects that with more training data, active learning is more necessary and effective.

For comparison, Fig. 2 demonstrates the results of several existing methods with part of the training trajectory samples. Note that the original Social-GAN [11] and Social-STGCNN [23] are two probabilistic models, and we adapt them to deterministic models. We can observe that our model consistently outperforms the Social-GAN and Social-STGCNN models on both ADE and FDE metrics with shrunk training

Table 3. Experimental results (ADE/FDE) of our proposed IAL-TP model with three different strategies, “Max”, “Median”, and “Min”.

Select strategy	Performance (ADE/FDE)
	Average
Max	0.52/1.06
Median	0.46/0.95
Min	0.54/1.11

trajectory data. In addition, with the same increase of $p\%$, our model has the least reduction on both ADE and FDE metrics, which validates the robustness of our proposed model. Note that without any specific indication, we set $p\% = 50\%$ in following sections.

4.3 Selection Strategy

In order to validate the effectiveness of our proposed “Median” selection strategy (introduced in Sect. 3.2), we design two others election strategies for comparison. One strategy is to select the samples with the largest error, which is referred as “Max”, the other strategy is to select the samples with the smallest error, which is referred as “Min”. Table 3 shows the results of three different select strategies. We can observe that the “Median” select strategy outperforms other strategies. This proves that samples with median inference errors are more representative. As discussed in Sect. 3.2, samples with larger error are more likely to have noise, which have negative influences on model learning. In addition, samples with smaller error are more likely to be too “easy” for model learning, namely these samples are less valuable. Thus our proposed “Median” select strategy is more appropriate while seeking the “worthy” trajectory samples.

4.4 Quantitative Analysis

Table 2 shows quantitative results of our proposed model and baseline. Overall, IAL-TP model outperforms all the baselines on the two metrics with only half of the training trajectory samples. The ADE metric improves by 14.8% compared to TF-based and the FDE metric improves by 12.0% compared to TPNet. In specific, the ADE error of our IAL-TP model on dataset ETH is 0.56, significantly improving by 42.8% compared to Social-STGCNN, the FDE error is 1.04, significantly improving by 48.3% compared to TPNet. It validates the necessity of our active learning idea in the pedestrian trajectory prediction problem. Additionally, it also validates the active set selected by our proposed strategy is a small but compact and representative training set.

5 Conclusion

In this paper, we propose a novel trajectory prediction model via incremental active learning (IAL-TP). In this model, we design a simple and effective method to iteratively select more valuable and representative trajectory samples for model learning,

which can filter out noisy and redundant samples. This incremental active learning method greatly improves the generalization ability and the robustness of the model. Experimental results on five public datasets validate the effectiveness of our model. Additionally, it can achieve better performance than the state-of-the-art methods with only a small fraction of the whole training data.

Acknowledgments. This work was supported by the Beijing Nova Program under No. Z201100006820046, the National Natural Science Foundation of China under Grant No. 61772373, and the Beijing Science and Technology Project under Grant no. Z181100008918018.

References

1. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S.: Social LSTM: human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 961–971 (2016)
2. Amirian, J., Hayet, J.B., Pettré, J.: Social ways: learning multi-modal distributions of pedestrian trajectories with GANs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2019)
3. Beygelzimer, A., Dasgupta, S., Langford, J.: Importance weighted active learning. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 49–56 (2009)
4. Bisagno, N., Zhang, B., Conci, N.: Group LSTM: group trajectory prediction in crowded scenarios. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11131, pp. 213–225. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11015-4_18
5. Cheng, H., Liao, W., Tang, X., Yang, M.Y., Sester, M., Rosenhahn, B.: Exploring dynamic context for multi-path trajectory prediction. arXiv preprint [arXiv:2010.16267](https://arxiv.org/abs/2010.16267) (2020)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding (2018)
7. Djuric, N., et al.: Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, pp. 2084–2093 (2020)
8. Fang, L., Jiang, Q., Shi, J., Zhou, B.: TPNNet: trajectory proposal network for motion prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6797–6806 (2020)
9. Fernando, T., Denman, S., Sridharan, S., Fookes, C.: GD-GAN: generative adversarial networks for trajectory prediction and group detection in crowds. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11361, pp. 314–330. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20887-5_20
10. Giuliari, F., Hasan, I., Cristani, M., Galasso, F.: Transformer networks for trajectory forecasting. In: Proceedings of the International Conference on Pattern Recognition (2020)
11. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., Alahi, A.: Social GAN: socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2255–2264 (2018)
12. Hong, J., Sapp, B., Philbin, J.: Rules of the road: predicting driving behavior with a convolutional model of semantic interactions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2019
13. Kosaraju, V., Sadeghian, A., Martín-Martín, R., Reid, I., Rezatofighi, H., Savarese, S.: Social-BiGAT: multimodal trajectory forecasting using bicycle-GAN and graph attention networks. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 137–146 (2019)

14. Lee, N., Wongun, C., Paul, V., Choy, C.B., Torr, P.H.S., Manmohan, C.: DESIRE: distant future prediction in dynamic scenes with interacting agents. In: Proceedings of the IEEE Computer Vision and Pattern Recognition, pp. 2165–2174, July 2017
15. Lerner, A., Chrysanthou, Y., Lischinski, D.: Crowds by example. *Comput. Graph. Forum* **26**(3), 655–664 (2010)
16. Li, J., Ma, H., Tomizuka, M.: Conditional generative neural system for probabilistic trajectory prediction. arXiv preprint [arXiv:1905.01631](https://arxiv.org/abs/1905.01631) (2019)
17. Li, Y., et al.: Deep learning for lidar point clouds in autonomous driving: a review. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 3412–3432 (2020)
18. Liang, J., Jiang, L., Murphy, K., Yu, T., Hauptmann, A.: The garden of forking paths: towards multi-future trajectory prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 10508–10518 (2020)
19. Liang, J., Jiang, L., Niebles, J.C., Hauptmann, A.G., Fei-Fei, L.: Peeking into the future: predicting future person activities and locations in videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5725–5734 (2019)
20. Mangalam, K., et al.: It is not the journey but the destination: endpoint conditioned trajectory prediction. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12347, pp. 759–776. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58536-5_45
21. Mantini, P., Shah, S.K.: Multiple people tracking using contextual trajectory forecasting. In: IEEE Symposium on Technologies for Homeland Security, pp. 1–6 (2016)
22. Mayank, B., Alex, K., Abhijit, O.: ChauffeurNet: learning to drive by imitating the best and synthesizing the worst. In: Proceedings of Robotics: Science and Systems, June 2019
23. Mohamed, A., Qian, K., Elhoseiny, M., Claudel, C.: Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 14424–14432 (2020)
24. Pellegrini, S., Ess, A., Schindler, K., Gool, L.J.V.: You’ll never walk alone: modeling social behavior for multi-target tracking. In: IEEE International Conference on Computer Vision, pp. 261–268 (2009)
25. Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezatofighi, H., Savarese, S.: SoPhie: an attentive GAN for predicting paths compliant to social and physical constraints. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1349–1358 (2019)
26. Salzmann, T., Ivanovic, B., Chakravarty, P., Pavone, M.: Trajectron++: dynamically-feasible trajectory forecasting with heterogeneous data. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12363, pp. 683–700. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58523-5_40
27. Sun, J., Jiang, Q., Lu, C.: Recursive social behavior graph for trajectory prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 660–669 (2020)
28. Tang, Y.C., Salakhutdinov, R.: Multiple futures prediction. In: Proceedings of the Advances in Neural Information Processing Systems (2019)
29. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
30. Vemula, A., Muelling, K., Oh, J.: Social attention: modeling attention in human crowds. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 1–7 (2018)
31. Xu, Y., Ren, D., Li, M., Chen, Y., Fan, M., Xia, H.: Tra2Tra: trajectory-to-trajectory prediction with a global social spatial-temporal attentive neural network. *IEEE Robot. Autom. Lett.* **6**(2), 1574–1581 (2021)

32. Xu, Y., Yang, J., Du, S.: CF-LSTM: cascaded feature-based long short-term networks for predicting pedestrian trajectory. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 12541–12548 (2020)
33. Yu, C., Ma, X., Ren, J., Zhao, H., Yi, S.: Spatio-temporal graph transformer networks for pedestrian trajectory prediction. arXiv preprint [arXiv:2005.08514](https://arxiv.org/abs/2005.08514) (2020)
34. Zhang, P., Ouyang, W., Zhang, P., Xue, J., Zheng, N.: SR-LSTM: state refinement for LSTM towards pedestrian trajectory prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12085–12094 (2019)
35. Zhu, Y., Qian, D., Ren, D., Xia, H.: StarNET: pedestrian trajectory prediction using deep neural network in star topology. arXiv preprint [arXiv:1906.01797](https://arxiv.org/abs/1906.01797) (2019)