An Improved Discriminator for GAN-Based Trajectory Prediction Models

Renhao Huang, Yang Song, Maurice Pagnucco
School of Computer Science and Engineering, University of New South Wales, NSW 2052, Australia
renhao.huang@student.unsw.edu.au, {yang.song1, morri}@unsw.edu.au

***Abstract*—Pedestrian trajectory prediction is an important component in autonomous systems, such as self-driving cars and social robots. It aims to accurately predict or plan future paths for pedestrians according to their movement histories. Recent studies have shown promising progress and most of them use some advanced encoder-decoder structures with Generative Adversarial Networks (GANs) to generate a distribution of multiple plausible paths of an agent. However, GAN-based models suffer from hard-training problems and training Recurrent Neural Networks (RNNs) is especially difficult. In this paper, we propose a discriminator that shares its encoder with the generator to reduce the training difficulty. We incorporate this discriminator into two successful stochastic models designed for pedestrian trajectory prediction. Our experimental results demonstrate that the new discriminator outperforms the baseline structures in general on multiple datasets.**

I. Introduction

Predicting pedestrians' future trajectories is challenging especially when their destinations are unknown. Many factors may increase the complexity of trajectories such as human-human interactions, surrounding environment and individual decisions. Nevertheless, pedestrians tend to follow social rules to adjust their paths to avoid collisions with their surrounding pedestrians or obstacles. These rules form social forces [1] that navigate pedestrians and can give some clues to predict future movement. Earlier methods statistically handcraft social features or build Long-Short Term Memory (LSTM) models [2] to predict future paths.

Stochastic models have recently been developed to generate multiple plausible routes for each pedestrian, most of which follow the encoder-decoder structure as shown in Fig. 1. Among the stochastic models, Social GAN [3], SoPhie [4], Social Ways [5] and Social BiGAT [6] incorporate the GAN architecture [7], which has a discriminator to distinguish whether a path is real or fake to assist with predicting a distribution conditioned on observed paths. In particular, Social GAN and SoPhie use the LSTM-based GANs, Social Ways uses InfoGAN [8] and Social BiGAT integrates BicycleGAN [9]. These GAN-based models achieve excellent performance on multiple datasets. There are also some non-GAN based models [10], [11] that remove the discriminator but incorporate graphs for further improvement. For example, SGAT [10] uses the Graph Attention Network [12] in the encoder to capture the interaction between pedestrians.

In this study, we focus on GAN-based trajectory prediction methods. Although GAN-based models can generate more accurate distributions theoretically, it has been demonstrated that a suitable choice of the discriminator is essential to generate good distributions and an FSGAN model has recently been proposed [13]. This finding also concurs with our empirical studies. To this end, we propose an improved discriminator that shares the trained encoder with the generator. This approach aims to simplify the training of the discriminator to stabilise its performance. We incorporate our improved discriminator into two highly effective stochastic models (Social GAN and a variant of SGAT) and demonstrate generally improved performance over the original models and also FSGAN on a number of public datasets.

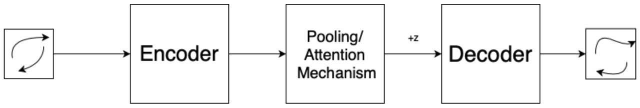

Fig. 1. Encoder-Decoder based model structure. Encoder for feature extraction; decoder for path generation; and pooling/attention module to integrate spatial information.

II. Problem Analysis and Solution

In our approach, we adopt the GAN-based model used in Social GAN with RNNs in the encoders and decoders. Our objective is to predict the future trajectory $\hat{Y}_i$ of each person i close to the ground truth Y_i based on the observed inputs $X_i = \{(x_i^t, y_i^t), t = 1..t_{obs}\}$, where x and y are the coordinates and t is the time step. We define the variety loss L_{L2} and adversarial loss $L_{GAN} = L_D + L_G$ as:

$$\begin{aligned} L_{L2} &= \min_k ||Y_i - \hat{Y_i}^{(k)}||_2 \\ L_D &= -E_{x \sim p}[LogD(x)] - E_{z \sim p}[1 - LogD(G(z))] \\ L_G &= -E_{z \sim p}[LogD(G(z))] \end{aligned} \tag{1}$$

Social GAN follows the training process that repeatedly updates the discriminator using L_D then the generator using $\lambda L_{L2} + (1-\lambda) L_G$. Initially, L_{L2} dominates the total losses (10 times of L_G) to create new trajectories with convincing direction and speed. Then, with the convergence of L_{L2}, L_G starts to control the generated trajectories, which results in a limited span of distribution. This phenomenon is indicated in [13] as well. Although further training helps spread the trajectories to a suitable distribution, the generated trajectories tend to divert considerably from the ground truths. Therefore, the prediction performance in terms of ADE and FDE at around 100 epochs is actually better than that at 200 epochs even though the distribution is more realistic qualitatively at 200 epochs, as shown in Fig. 4.

A plausible explanation is that, different from image tasks, temporal data generation is hard for GAN. If the decoder fails to generate accurate predictions at previous time steps, the future steps may follow such results to generate new inaccurate predictions. For stochastic models, both generator and discriminator have their own RNNs to encode the embedded human trajectories, but they aim to handle different tasks: the generator uses those features to generate future paths while the discriminator attempts to evaluate the generated path. Therefore, it is challenging to ensure that the RNNs in generator and discriminator are successfully trained and correlated with each other.

To address this problem, FSGAN presents a pure feed-forward discriminator to avoid training RNNs inside the discriminator. However, it is hard to find an optimal set of hyper-parameters to achieve expected performance. Therefore, we propose a simpler solution that the discriminator shares parameters of the encoder with the generator (Fig. 2): each time the discriminator needs to extract trajectory features, it uses the encoder pre-trained by the generator. Then, the generator can well train its encoder due to the L_{L2} correction. In addition, because the discriminator only needs to train its fully connected layers that follow a feedforward structure, the model complexity decreases and hence the training becomes more stable.

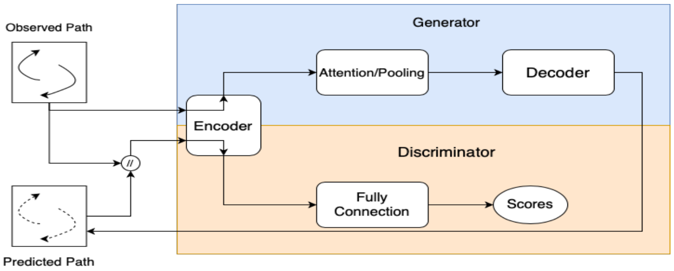

Fig. 2. The proposed architecture with shared encoder.

III. Experimental Results

To assess our method's performance, we substituted the discriminator in Social GAN (SGAN) and SGAT-GAN with our improved one, namely encoder-sharing (SE). SGAT-GAN was constructed by replacing the pooling module of SGAN with GAT to create a GAN-based variant of SGAT. We trained the models with batch size of 64 without changing the other hyper-parameters. Training was slightly accelerated because of the sharing of encoders between discriminator and generator. As shown in Fig. 3, the GAN model is unstable at the beginning because the encoder in generator has not completed training. After a number of epochs, both adversarial losses start to concentrate on an equilibrium point and the losses become more stable. Fig. 4 shows that although the ADE and FDE values of both models increase with more epochs, SE always outperforms the original version. We refer our explanation of increasing ADE and FDE to Section II. Also, the change of ADE and FDE with our SE model is more stable than the original SGAN through training. Final quantitative results in Table I were obtained with 200 epochs of training. It can be seen that our SE models have generally better results than the original SGAN and SGAT-GAN and also FSGAN.

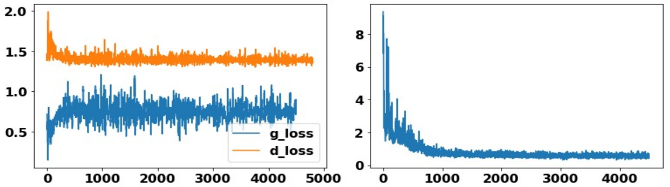

Fig. 3. Plot of L_{GAN} (left) and L_{L2} (right) vs Iterations on test UNIV using SGAN-SE.

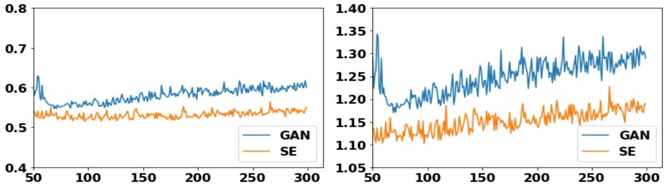

Fig. 4. Plot of ADE (left) and FDE (right) vs Epochs on test UNIV after 50 epochs of training with SGAN as the base model.

TABLE I

Quantitative results with two error metrics Average Displacement Error (ADE) and Final Displacement Error (FDE) for 12 future time steps and $k = 20$ trajectories.

	SGAN	SGAN-SE	FSGAN	SGAT-GAN	SGAT-SE
ETH	0.81/1.52	**0.63/1.13**	0.68/1.16	0.70/1.35	**0.61/1.06**
HOTEL	0.48/1.03	**0.39/0.79**	0.43/0.89	**0.37/0.67**	0.43/0.91
UNIV	0.60/1.26	**0.53/1.15**	0.54/**1.14**	0.59/1.23	**0.50/1.19**
ZARA1	**0.34/0.69**	**0.34**/0.70	0.35/0.71	0.35/**0.69**	**0.34**/0.70
ZARA2	0.42/0.84	**0.36/0.61**	0.32/0.67	**0.31/0.64**	0.36/0.78

IV. Conclusion and Future Work

This paper presents a method to solve the shortcoming of GAN-based pedestrian trajectory prediction models. We introduce an improved discriminator that shares the encoder with the generator to simplify its training complexity. We find our discriminator can stabilise and improve trajectory prediction performance. However, the success of non-GAN based stochastic models [10], [11] illustrates as well that currently there is insufficient evidence that trajectory prediction benefits significantly from GAN. A possible reason is that using ADE and FDE as evaluation metrics is not a perfect way to fully determine the prediction performance as factors such as human minds and the real destination cannot be predicted. These GAN based models can indeed generate multiple routes without following the ground truth but may be highly useful in real cases. Therefore, we suggest that there is still large potential in GAN-based models and improving GAN structures further or finding more suitable evaluation metrics is well worth exploration.

References

[1] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics." *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, vol. 51, no. 5, pp. 4282–4286, 1995.

[2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, F.-F. Li, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *CVPR*, 2016.

[3] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *CVPR*, 2018.

[4] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, "SoPhie: An attentive gan for predicting paths compliant to social and physical constraints," in *CVPR*, 2019.

[5] J. Amirian, J.-B. Hayet, and J. Pettré, "Social ways: Learning multimodal distributions of pedestrian trajectories with gans," in *CVPRW*, 2019.

[6] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. D. Reid, S. H. Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal trajectory forecasting using Bicycle-GAN and graph attention networks," in *NeurIPS*, 2019.

[7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *NeurIPS*, 2014.

[8] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *NeurIPS*, 2016.

[9] J. Zhu, R. Zhang, D. Pathak, and T. Darrell, "Toward multimodal image-to-image translation," in *NeurIPS*, 2017.

[10] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "STGAT: Modeling spatial-temporal interactions for human trajectory prediction," in *ICCV*, 2019.

[11] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *CVPR*, 2020.

[12] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *ICLR*, 2018.

[13] K. Parth and A. Alexandre, "Adversarial loss for human trajectory prediction," in *hEART*, 2019.