Human motion trajectory prediction using the Social Force Model for real-time and low computational cost applications

Óscar Gil*, and Alberto Sanfeliu

Institut de Robòtica i Informàtica Industrial, CSIC-UPC
{ogil,sanfeliu}@iri.upc.edu

Abstract. Human motion trajectory prediction is a very important functionality for human-robot collaboration, specifically in accompanying, guiding, or approaching tasks, but also in social robotics, self-driving vehicles, or security systems. In this paper, a novel trajectory prediction model, Social Force Generative Adversarial Network (SoFGAN), is proposed. SoFGAN uses a Generative Adversarial Network (GAN) and Social Force Model (SFM) to generate different plausible people trajectories reducing collisions in a scene. Furthermore, a Conditional Variational Autoencoder (CVAE) module is added to emphasize the destination learning. We show that our method is more accurate in making predictions in UCY or BIWI datasets than most of the current state-of-the-art models and also reduces collisions in comparison to other approaches. Through real-life experiments, we demonstrate that the model can be used in real-time without GPU's to perform good quality predictions with a low computational cost.

Keywords: Human Motion Prediction, Social Force Model, Generative Adversarial Network, Conditional Variational Autoencoder

1 Introduction

Several studies [2, 15] about Theory of Mind and mirror neurons, emphasize prediction as an essential tool for humans to increase their performance in social interactions through an anticipative behavior. A person can build a model about the internal mental state of people via social interactions to predict future actions.

In particular, human motion prediction is a very broad field with a large number of different categories that depend on multiples factors like the person task, the person model or the person body parts. In the case of human motion trajectory prediction, a very exhaustive taxonomy based on the model approach and the contextual cues have been presented in [18].

Human motion trajectory prediction is a complex task very difficult to understand, due to the very different strategies people use to avoid collisions; the variety of social interactions; the relative nature of consider something an obstacle or

* Work supported under the European project CANOPIES with grant number H2020-ICT-2020-2-101016906 and JST Moonshot R & D Grant Number JPMJMS2011-85.

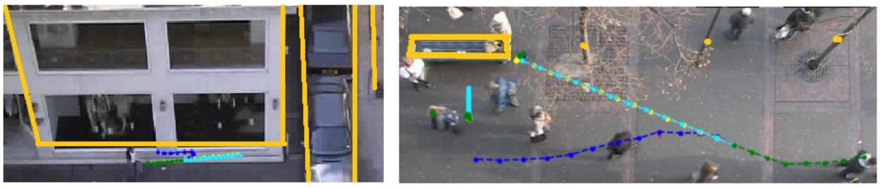


Fig. 1: **Complex cases.** The left picture shows a person who randomly changes the movement direction because is waiting for someone. The right picture shows a bench that can be an obstacle for some pedestrians or a goal for others.

a goal; and the sudden changes in the movements due to internal unpredictable stimulus (refer to Fig.1).

The stimuli for pedestrian motion can be internal or external. The internal stimuli are very difficult to infer because they are related to the particular person's thoughts. The external stimuli are related to the environment, but the response of the stimulus is related to the person's mental state as well. On the whole, any response of the person depends on external and internal stimuli and there aren't handwritten rules or laws to explain all the cases.

Due to this complex dependence, this work uses the environment information through the Social Force Model (SFM) [6] and people features like the velocities and the resultant forces. This information is used to generate a set of possible paths. The advantage of generating a set of paths is that the multimodal behavior of pedestrians can be handled.

The remainder of this paper is organized as follows. In section 2, the related work is introduced. In Section 3, the SFM used to encode the environment information is described. In Section 4, a description of the complete approach, that combines a Generative Adversarial Network (GAN) and a Conditional Variational Autoencoder (CVAE) is introduced. Section 5 provides an analysis of the metrics to evaluate the models and make state-of-the-art comparison as fair as possible. In Section 6, the evaluation by the usual methodology in different datasets is performed and the real-life experiment results are analyzed. Finally, in section 7, the conclusions are provided.

2 Related Work

Nowadays, multiple approaches to human motion prediction have been developed. These models are normally data-driven models to forecast the next skeleton 3D movements in concrete tasks like, for example, walking, eating, smoking, or running [13, 8]. Commonly, these data-driven methods formulate the problem as a sequence-to-sequence task where the data is processed by a Recurrent Neural Network (RNN) or a Long Short-Term Memory (LSTM). These network architectures are chosen due to their ability to encode temporal information.

In human motion trajectory prediction, as a part of human motion prediction, there are a lot of elements in common like social interaction or multimodal behavior. By contrast, there are fewer dependencies on the task, and the person model can be simplified, for example by considering a point or 2D circle.

Physics-based approaches, like Constant Velocity Model (CVM) [21] or Extended Kalman Filter (EKF), are simpler than other methods, but can give a good performance, for example, in linear trajectories. These approaches can be obstacle-aware methods like the Social Force Model [6], which can be combined with goal estimation [3] for better performance.

Actually, data-driven models are very common in trajectory prediction. Most of these approaches utilize LSTM architectures. In this paragraph, we discuss models that do not take into account static obstacles or collisions. A first method is Social LSTM [1], which uses an encoder-decoder structure with LSTM cells and a pooling mechanism to encode the people that is close to a person. The MX-LSTM model [5] considers not only people positions (tracklets), but gaze direction (vislets) too. A transformer is proposed in the AgentFormer model [22], which learns simultaneously the social and time dimensions. Social GAN [4] uses a GAN and a pooling mechanism to generate a multimodal distribution of the trajectories. In PECNet [12], a CVAE is included to obtain an accurate estimation of the goals.

Furthermore, obstacle-aware models like Next [9] and SoPhie [19] extract features from the image of the scene using a convolutional neural network (CNN) and then, a RNN is utilized to obtain predictions. Other works, like Trajectron++ [20], use a CNN to incorporate the map information and take into account the dynamics through an RNN. NSP model [23] uses the Social Force Model (SFM) and a Neural Differential Equation for the motion prediction. In [11], the scene features are used to obtain waypoints along the trajectory to improve the results in a long prediction horizon.

3 Social Force Model

The Social Force Model of Helbing and Molnar [6] for pedestrian dynamics allows to simulate social interactions as forces. In a scene with a set of pedestrians P and obstacles O, it is considered that a pedestrian $p \in P$ moves towards a goal with the following attractive force:

$$\mathbf{f_p^{goal}} = k(\mathbf{v_p^0} - \mathbf{v_p}) \tag{1}$$

where $\mathbf{v_p}$ is the current velocity of the pedestrian and k^{-1} is the relaxation time to achieve the desire velocity pointing towards the goal, $\mathbf{v_p^0}$.

To consider the influence of pedestrians and obstacles, avoid collisions and respect social distances, a repulsive force is defined as:

$$\mathbf{f_{z,p}^{int}} = A_z e^{(d_z - d_{z,p})/B_z} \mathbf{\hat{d}_{z,p}} \tag{2}$$

where z can be a pedestrian or obstacle. A_z, B_z and d_z are parameters that can be adjusted. $d_{z,p}$, is the Euclidean distance between z and the pedestrian p. $\mathbf{\hat{d}_{z,p}}$ is the unitary vector in the line between p and z positions, pointing to p.

Zanlungo et al. [24] consider collision prediction into the interaction force between pedestrians $p, q \in P$. In this case the repulsive force for pedestrian p is:

$$\mathbf{f^{int}_{q,p}}(\{\mathbf{v_{q,p}}\}, \{\mathbf{d_{q,p}}\}, \mathbf{v_p}) = A_q \frac{v_p}{t_p} e^{-d_{q,p}/B_q} \frac{\mathbf{d'_{q,p}(t_p)}}{d'_{q,p}(t_p)} \tag{3}$$

where $\{\mathbf{v_{q,p}}\}$ is the set of relative velocities between p and other pedestrians. $\{\mathbf{d_{q,p}}\}$ is the set of vectors with all relative distances between p and other pedestrians, pointing to p. t_p=$\min_q\{t_{q,p}\}$, where $t_{q,p}$ is the time in which p is at the minimum distance from q and $\mathbf{d'_{q,p}(t_p)}$ is the relative position of p regarding q, in t=$t_{q,p}$. When the angle between $\mathbf{v_{q,p}}$ and $\mathbf{d_{q,p}}$ complies with $|\theta_{p,q}| > \pi/4$, then $t_{q,p} = \infty$. Bearing in mind these forces, the resulting force for a pedestrian p is:

$$\mathbf{F_p} = \mathbf{f^{goal}_p} + \sum_{q \in P} \mathbf{f^{int}_{q,p}} + \sum_{o \in O} \mathbf{f^{int}_{o,p}} \tag{4}$$

This model has been generalized to robots by means of the Extended Social Force Model (ESFM). This generalization is very useful for planning in collaborative tasks that involve social-aware navigation [17, 16].

4 Social Force GAN Model

In this section, we describe approach which considers SFM, a GAN and a CVAE.

4.1 Problem Formulation

Trajectory prediction in an environment can be considered as a prediction of multiple positions of points along different discrete times (timesteps). In order to make predictions, an observation horizon, T_{obs} and a prediction horizon of timesteps, T_{pred}, are established in a trajectory $\mathbf{X_i}$, for a pedestrian i, with the correspondent ground truth positions $(x^t_i, y^t_i) \in \mathbf{X_i}$ associated in each horizon:

$$\mathbf{X^{obs}_i} = \{(x^t_i, y^t_i) | t = 1, 2, \ldots, T_{obs}\} \tag{5}$$

$$\mathbf{X^{pred}_i} = \{(x^t_i, y^t_i) | t = T_{obs} + 1, \ldots, T_{obs} + T_{pred}\} \tag{6}$$

The main objective is to forecast the future positions of the trajectory:

$$\mathbf{Y_i} = \{(x^t_i, y^t_i) | t = T_{obs} + 1, \ldots, T_{obs} + T_{pred}\} \tag{7}$$

as close as possible to the ground truth positions in the prediction horizon, $\mathbf{X^{pred}_i}$. To achieve this goal, the previous positions of the trajectory, $\mathbf{X^{obs}_i}$, can be used as information. Other cues as the gaze, the potential goals, the neighbors' positions or the environment map can also be used.

4.2 Social Force Representation

SFM is used in this work to obtain an environment representation R, as a part of the inputs for the Social Force GAN model (SoFGAN). Given a pedestrian $p \in P$, the environment representation for this pedestrian is calculated using (2), to calculate repulsive forces because of static obstacles, and using (3) to calculate forces due to other pedestrians.

These 2 types of repulsive forces could be added separately or in 1 unique force but, in the last case, the resultant forces would not give a complete description of the environment, because there are a lot of different combinations of static obstacles or pedestrians that can give the same resultant forces.

Therefore, to avoid this problem, M angle bins centered in pedestrian p are considered to divide the space. Using this method, two resultant forces are calculated for each angle bin. One for the static obstacles and another for the pedestrians. The two sets of forces are used as the Social Force Representation:

$$R_p = \{\mathbf{F^{bin}_{ped}}, \mathbf{F^{bin}_{obst}}\} \tag{8}$$

where each set of forces is:

$$\mathbf{F^{bin}_{ped}} = \{\sum_{q \in P_i} \mathbf{f^{int}_{q,p}}\}_{i=1}^{M} \tag{9}$$

$$\mathbf{F^{bin}_{obst}} = \{\sum_{o \in O_i} \mathbf{f^{int}_{o,p}}\}_{i=1}^{M} \tag{10}$$

O_i and P_i are the subsets of obstacles and pedestrians whose forces applied in p are into the angle bin i and comply with $\bigcup_{i=1}^{M} O_i = O$ and $\bigcup_{i=1}^{M} P_i = P$. An example is shown in Fig. 2.

The forces of the static obstacles are calculated modeling obstacles in the environment as polygons and choosing always the nearest point of the polygon to p as the obstacle point to calculate the force. The homography matrix of the images is used to obtain the Cartesian coordinates of the polygons. Taking into account this representation, the inputs of the SoFGAN model for all timesteps in the observation horizon are as follows:

- $\mathbf{f^{goal}_p}$: The attractive force calculated using (1) for each pedestrian.
- R_p: The Social Force Representation.
- $\mathbf{X^{obs}_{rel}}$: The relative positions of each pedestrian in the observation horizon and considering the first relative position zero.
- $\mathbf{F^{obs}_{total}}$: The necessary forces at each timestep to generate the trajectory starting at the first position.
- $\mathbf{X^{obs}_r}$: The relative positions of each pedestrian in the observation horizon taking as reference the first position in the trajectory.
- $\mathbf{g^{GT}_r}$: The final ground truth position for each trajectory taking as reference the first position in the trajectory. Only used during training.

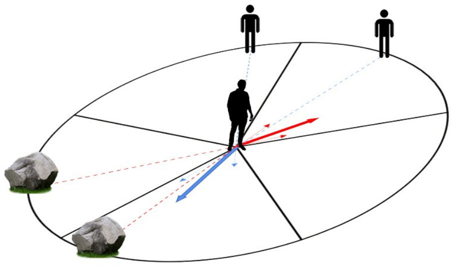


Fig. 2: **Social Force Representation.** In this example 5 angle bins are considered. The thick vectors are the resultant forces of the thin vectors in each angle bin. There are 2 blue forces of 2 pedestrians and 2 red forces of 2 obstacles applied in a pedestrian located in the center.

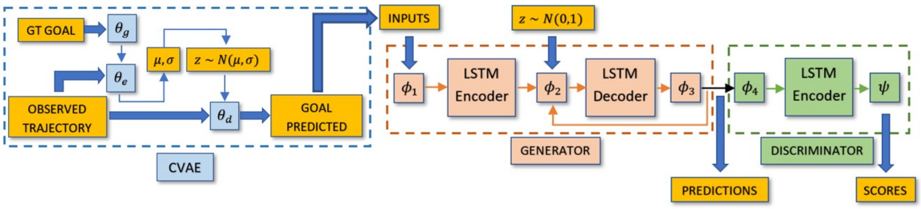


Fig. 3: **Social Force GAN with CVAE architecture.** The GAN generator provides the predictions using the predicted goals given by the CVAE module.

The forces to generate the trajectory, $\mathbf{F}^{\mathbf{obs}}_{\mathbf{total}}$, are calculated using the relative positions and the timestep value (to compute their velocities). Considering $\Delta t = 1$ and a unitary mass, the force is calculated as a difference of the straight sections in the trajectory at each timestep. For the attractive force, the goal is in the last position of the observation horizon.

4.3 SoFGAN model description

The model can be summarized as:

$$\{\mathbf{Y}_{\mathbf{rel}}, \mathbf{F}^{\mathbf{pred}}_{\mathbf{total}}\} = SoFGAN(\mathbf{X}^{\mathbf{obs}}_{\mathbf{rel}}, \mathbf{f}^{\mathbf{goal}}_{\mathbf{p}}, R_p, \mathbf{F}^{\mathbf{obs}}_{\mathbf{total}}, \mathbf{X}^{\mathbf{obs}}_{\mathbf{r}}, \mathbf{g}^{\mathbf{GT}}_{\mathbf{r}}) \quad (11)$$

where $\mathbf{Y}_{\mathbf{rel}}$ is the set of the predicted relative positions from which the absolute positions for all pedestrians $\{\mathbf{Y_i}\}_{i=1}^{N}$ can be calculated. $\mathbf{F}^{\mathbf{pred}}_{\mathbf{total}}$ are the necessary forces to generate at each timestep a new position, to form a second predicted trajectory. Therefore, SoFGAN can also be considered as a force predictor.

The architecture of the model is basically a GAN, like in [4] and a CVAE module. The complete model is shown in Fig. 3, where ϕ_1, ϕ_2, ϕ_3 and ϕ_4 are linear transformations. ψ, θ_g, θ_e, θ_d are MLP's.

The CVAE module is used to reduce the main source of the prediction errors, the last position prediction. For that reason, it is used to estimate the goal trajectories. The inputs for the CVAE are $\mathbf{X_r^{obs}}$ and $\mathbf{g_r^{GT}}$. The second one is only used during training. During the test, θ_g and θ_e are not used. The only necessary input for the CVAE in the test phase is the observed trajectory because $\mathbf{z}$ is sampled from $N(\mathbf{0}, \alpha\mathbf{I})$, a normal distribution with mean zero and a fixed variance α. If the CVAE module is not used, the GAN inputs are $\mathbf{X_{rel}^{obs}}$, $\mathbf{f_p^{goal}}$, R_p and $\mathbf{F_{total}^{obs}}$. In case of use the CVAE module, the attractive force is substituted by the CVAE predicted goals, $\mathbf{g_r^{pred}}$, because it improves the model results.

The scores obtained in the discriminator are used as labels to train the model. The losses used to train the GAN and the CVAE in this approach are the adversarial loss, the variety loss [4] and the CVAE loss. The CVAE loss is composed by the Kullback-Leibler (KL) divergence and the variety loss applied to the last predicted position:

$$L_{adv} = \min_G \max_D [\mathbb{E}_{x\sim p_d} log D_{\theta_d}(x) + \mathbb{E}_{z\sim p(z)} log(1 - D_{\theta_d}(G_{\theta_g}(z)))] \tag{12}$$

$$L_{var} = \min_k \|\mathbf{Y_{rel}^k} - \mathbf{X_{rel}^{k,pred}}\| \tag{13}$$

$$L_{CVAE} = \lambda_2 D_{KL}(N(\boldsymbol{\mu}, \boldsymbol{\sigma})\|N(\mathbf{0}, \mathbf{I})) + \lambda_3 \min_k \|\mathbf{g_r^{pred}} - \mathbf{g_r^{GT}}\| \tag{14}$$

$$L_{total} = \mathbb{E}_{p\in P}[L_{adv} + \lambda_1 L_{var} + L_{CVAE}] \tag{15}$$

where k is the number of generator samples.

5 Metrics

In this work, as in [22] and [20], k trajectory samples are used to compute the minimum Average Displacement Error (mADE) and the minimum Final Displacement Error (mFDE) separately over the sampled scenes that contain different number of pedestrians.

Another metric to measure the social-awareness of the model is the average % of colliding pedestrians per frame in a dataset, $\%c$. This value is used in different works like [19], although they do not give a detailed explanation about how to obtain this metric. In this work, a collision is detected when 2 pedestrians get closer than 0.1 m in a frame, but not between successive frames. The $\%c$ of collisions is then computed as follows:

$$\%c = \sum_{j=1}^{k} \frac{1}{k} \left(\sum_{i=1}^{N_f} \frac{1}{N_f} (\frac{100 N_{ij}^c}{N_{ij}}) \right) \tag{16}$$

where N_f is the number of predicted timesteps in the test set, k is the number of generator samples, N_{ij}^c is the number of colliding pedestrians in timestep or frame i in the sample j. N_{ij} is the total number of pedestrians in the timestep i and sample j.

In this work, as usual, the observation horizon is $T_{obs} = 3.2s$, which corresponds to 8 timesteps. The prediction horizon is $T_{pred} = 4.8s$, which corresponds to 12 timesteps.

6 Experiments

The SoFGAN model has been evaluated through the BIWI [14], UCY [7] datasets using the leave-one-out cross validation technique as in [1], [4] and [19]. Moreover, the model has been implemented in ROS to predict people in real-time with a low computational cost.

6.1 Evaluation Results

Model	**ETH**	**HOTEL**	**ZARA1**	**ZARA2**	**UNIV**	**AVG**
SGAN [4]	0.81/1.52	0.72/1.61	0.34/0.69	0.42/0.84	0.60/1.26	0.58/1.18
SoPhie [19]	0.70/1.43	0.76/1.67	0.30/0.63	0.38/0.78	0.54/1.24	0.54/1.15
Next [9]	0.73/1.65	0.30/0.59	0.38/0.81	0.31/0.68	0.60/1.27	0.46/1.00
CVM-20 [21]	0.96/2.09	0.29/0.54	0.52/1.03	0.34/0.70	0.61/1.26	0.54/1.12
PECNet [12]	0.54/0.87	0.18/0.24	0.22/0.39	0.17/0.30	0.35/0.60	0.29/0.48
Trajectron++ [20]	0.57/1.05	0.16/0.26	0.22/0.41	0.16/0.31	0.28/0.56	0.28/0.52
AgentFormer [22]	0.45/0.75	0.14/0.22	0.18/0.30	0.14/0.24	0.25/0.45	0.23/0.39
Y-NET [11]	0.28/0.33	0.10/0.14	0.17/0.27	0.13/0.22	0.24/0.41	0.18/0.27
NSP [23]	**0.25/0.24**	**0.09/0.13**	**0.16/0.27**	**0.12/0.20**	**0.21/0.38**	**0.17/0.24**
SoFGAN	0.44/0.68	0.13/0.18	0.20/0.35	0.17/0.31	0.25/0.46	0.24/0.40
fSoFGAN	0.48/0.79	0.15/0.22	0.22/0.42	0.20/0.38	0.28/0.53	0.27/0.47

Table 1: **mADE/mFDE results.** Comparison between SoFGAN and other models for 20 samples. Numbers in bold type are the best results.

The results of the evaluation in terms of $mADE$ and $mFDE$, for BIWI and UCY datasets, are shown in Table 1, where SoFGAN is the model with a CVAE trained using data augmentation. To improve the performance, the 20 samples are selected through a k-means clustering of 1000 samples, as in [11]. fSoFGAN evaluate the $mADE$ and $mFDE$ of the generated trajectories using the total forces. For SGAN [4], SoPhie [19] and Next [9] the paper results in [9] are exposed. CMV-20 use 20 samples and it has been implemented because the evaluation in [21] is different.

The Trajectron++ results are different from the paper because an error in the velocity and acceleration estimation has been corrected. The Y-NET and NSP results have been obtained from the papers directly because the datasets and hyperparameters to reproduce the evaluation are not public. Only the AgentFormer model is the one that has been tested obtaining slightly better results than the SoFGAN model.

Table 2 shows the $\%c$ for different models where GT is the ground truth. It is important to underscore that the results of SGAN and SoPhie $\%c$ are the ones that appear in their corresponding papers, although we do not know if their $\%c$ calculation method is the same as our method. The same occurs with NSP and the Y-NET $\%c$. Nevertheless, the improvement compared with CVM-20 is very significant. Although SoFGAN does not provide the best results in terms of mADE, mFDE and %c, it outperforms most of state of the art methods. The SoFGAN and NSP models show that the use of forces can be very useful to encode the environment information and improve the predictions.

Model	**ETH**	**HOTEL**	**ZARA1**	**ZARA2**	**UNIV**	**AVG**
GT	0.000	0.000	0.000	0.000	0.056	0.011
SGAN [4]	2.509	1.752	1.749	2.020	**0.559**	1.717
SoPhie [19]	1.757	1.936	1.027	1.464	0.621	1.361
CVM-20 [21]	1.764	1.430	2.680	2.163	4.172	2.442
Y-NET [11]	**0.000**	**0.000**	0.820	1.310	1.510	0.730
NSP [23]	**0.000**	**0.000**	**0.000**	**0.660**	1.480	**0.430**
SoFGAN	0.250	0.500	0.707	1.226	3.688	1.274

Table 2: **%c across BIWI and UCY datasets for 20 samples.**

An ablation study has been performed to demonstrate the benefits of the CVAE and the data augmentation. M(w/o1) is the model without the CVAE. M(w/o2) is the M(w/o1) model without data augmentation. The average mADE and mFDE are shown in Table 3 for the BIWI and UCY datasets.

Model	**SoFGAN**	**M(w/o1)**	**M(w/o2)**
AVG	**0.24/0.40**	0.25/0.44	0.29/0.52

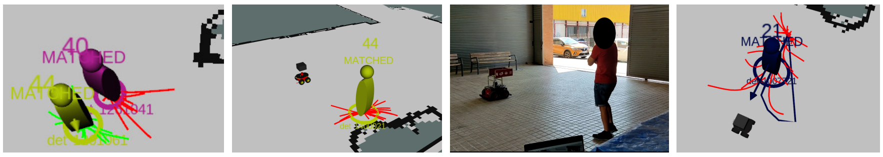
Table 3: **mADE/mFDE results of the ablation study.**

6.2 Real-life Experiments

Through a ROS implementation in the Helena robot, real outdoor experiments have been performed with people. Helena is a transporter robot with a RS-LiDAR-16 and a Pioneer P3-DX. Although the Helena prediction is not computed, Helena is taken into account as an agent to compute the forces.

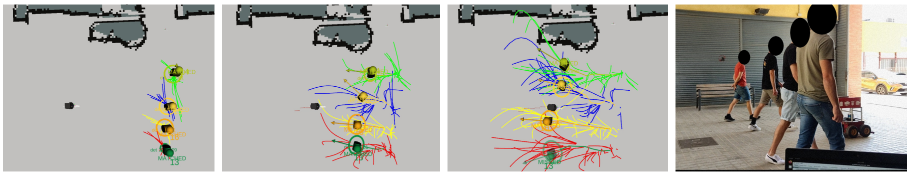
The first experiment is to evaluate the predictions when the Helena robot is not moving. In Fig. 4, from left to right, in the first image the interaction between 2 people in a conversation is shown. The model does not sample trajectories between them. In the second image, the effect of the obstacle behind the person cause that most of predictions appear in different directions. In the last image, the person walks towards Helena. In this case, the predictions take into account the robot and try to avoid collisions.

The second experiment, shown in Fig. 5, evaluates the interaction between 4 people and the interaction between the group and the robot. From left to



(a) RViz (b) RViz (c) Real World (d) RViz

Fig. 4: **Predictions when Helena is not moving.** The predictions are the red and green lines in the ground. The visualization is obtained using RViz.



(a) RViz: t (b) RViz: t+1 (c) RViz: t + 2 (d) Real World

Fig. 5: **Predictions during an encounter between 4 people and Helena.** The predictions are the colored lines in the ground.

right, in the first image, the group is not moving and the predictions are short and avoid collisions between them. In the second image, the group moves and the predictions try to avoid collisions with the robot. In the third image, the people change their velocity directions due to the robot and the blue and green predictions close to obstacles try to avoid a collision.

6.3 Implementation Details

For this work we have used the Adam optimizer, with a learning rate of 0.0005 for both generator and discriminator. For the variety loss, the λ_1 weight is 0.5 and the number of samples is 20. When the CVAE module is added, λ_1 is the same, λ_2 is 1 and λ_3 is 0.5. The CVAE μ parameter for the probability distribution has 16 components and $\boldsymbol{\sigma} = \alpha\mathbf{I}$ is a 16×16 matrix. For testing μ is zero and α is 3. The parameters for μ and σ have been chosen between other combinations to obtain a good performance. For noise, the dimension is 32 and the dimensions for hidden states are 64 for each encoder and 128 for the decoder. The embedding dimension is 64 and the batch size is 256.

The LSTM encoder and decoder have one layer. The MLP, ψ, used in the discriminator, has dimensions (64, 1024, 1). The CVAE MLP's, θ_g, θ_e, θ_d have these dimensions successively: (2, 8, 16, 16), (32, 8, 50, 32) and (32, 1024, 512, 1024, 2). All the layers use ReLu as activation function and batch normalization, except for the last layer of θ_d.

To calculate the Social Force Representation, four angle bins have been considered. All the models have been trained using Pytorch in a Tesla K40 GPU.

The real experiments have been performed using a CPU and ROS Melodic. The Lidar measures are provided to the Spencer tracker [10] to detect moving obstacles. The static obstacles are considered using the 2-D occupancy grid of the environment map. This map is used for robot localization and navigation in ROS. The model can compute the predictions in less than 100 ms.

7 Conclusions

In this work, we have presented a new human trajectory predictor model denominated Social Force Generative Adversarial Network (SoFGAN). This new trajectory predictor, is based on Social Force Model (SFM) and Generative Adversarial Network (GAN). One of the main advantages of this new trajectory predictor, is that includes the social forces of the environment and the moving pedestrians, to improve the human prediction in the next timesteps. Additionally, we have included a CVAE in order to learn the trajectory goal distribution resulting in an improvement of the model. The experimental results on standard datasets of complex human motions, show that our predictor gets good results in comparison to the best state of the art methods that we have tested. Moreover, we also obtain good results in the percent of colliding pedestrians per frame, $\%c$, in ETH/UCY datasets. Unlike other authors, we demonstrate that our model can be used in real time applications with a low computational cost to perform human-like predictions.

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. Brown, E., Brüne, M.: The role of prediction in social neuroscience. Frontiers in Human Neuroscience 6, 147 (05 2012)
3. Ferrer, G., Sanfeliu, A.: Bayesian human motion intentionality prediction in urban environments. Pattern Recognition Letters 44, 134–140 (2014)
4. 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)
5. Hasan, I., Setti, F., Tsesmelis, T., Del Bue, A., Galasso, F., Cristani, M.: Mx-lstm: mixing tracklets and vislets to jointly forecast trajectories and head poses. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6067–6076 (2018)
6. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Physical review E 51(5), 4282 (1995)
7. Lerner, A., Chrysanthou, Y., Lischinski, D.: Crowds by example. In: Computer graphics forum. vol. 26, pp. 655–664. Wiley Online Library (2007)
8. Li, Y., Li, K., Wang, X., Da Xu, R.Y.: Exploring temporal consistency for human pose estimation in videos. Pattern Recognition 103, 107258 (2020)

9. 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)
10. Linder, T., Breuers, S., Leibe, B., Arras, K.O.: On multi-modal people tracking from mobile platforms in very crowded and dynamic environments. In: 2016 IEEE International Conference on Robotics and Automation (ICRA). pp. 5512–5519 (2016)
11. Mangalam, K., An, Y., Girase, H., Malik, J.: From goals, waypoints & paths to long term human trajectory forecasting. In: Proc. International Conference on Computer Vision (ICCV) (Oct 2021)
12. Mangalam, K., Girase, H., Agarwal, S., Lee, K.H., Adeli, E., Malik, J., Gaidon, A.: It is not the journey but the destination: Endpoint conditioned trajectory prediction. In: European Conference on Computer Vision. pp. 759–776. Springer (2020)
13. Martinez, J., Black, M.J., Romero, J.: On human motion prediction using recurrent neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4674–4683 (2017)
14. Pellegrini, S., Ess, A., Schindler, K., Van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 261–268. IEEE (2009)
15. Pezzulo, G., Candidi, M., Dindo, H., Barca, L.: Action simulation in the human brain: twelve questions. New Ideas in Psychology 31(3), 270–290 (12 2013)
16. Repiso, E., Garrell, A., Sanfeliu, A.: Robot approaching and engaging people in a human-robot companion framework. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 8200–8205. IEEE (2018)
17. Repiso, E., Garrell, A., Sanfeliu, A.: Adaptive side-by-side social robot navigation to approach and interact with people. International Journal of Social Robotics pp. 1–22 (2019)
18. Rudenko, A., Palmieri, L., Herman, M., Kitani, K.M., Gavrila, D.M., Arras, K.O.: Human motion trajectory prediction: A survey. The International Journal of Robotics Research 39(8), 895–935 (2020)
19. 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)
20. Salzmann, T., Ivanovic, B., Chakravarty, P., Pavone, M.: Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16. pp. 683–700. Springer (2020)
21. Schöller, C., Aravantinos, V., Lay, F., Knoll, A.: What the constant velocity model can teach us about pedestrian motion prediction. IEEE Robotics and Automation Letters 5(2), 1696–1703 (2020)
22. Yuan, Y., Weng, X., Ou, Y., Kitani, K.: Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
23. Yue, J., Manocha, D., Wang, H.: Human trajectory prediction via neural social physics. In: Proceedings of the European Conference on Computer Vision (ECCV) (2022)
24. Zanlungo, F., Ikeda, T., Kanda, T.: Social force model with explicit collision prediction. EPL (Europhysics Letters) 93(6), 68005 (2011)