Online-Journals.org

iTDAF IETI Transactions on Data Analysis and Forecasting

iTDAF | eISSN: 2959-0442 | Vol. 3 No. 1 (2025) | OPEN ACCESS

https://doi.org/10.3991/itdaf.v3i1.53925

PAPER

Modeling Spatial-Temporal Social Interactions for Pedestrians Trajectory Prediction on Real and Synthetic Datasets

Md. Muktar Ali[1](✉), Md. Tahmeed Kowsher Hameem[1], Deen Islam[2], Md. Arafath Hossen Abir[1], Md. Moijeuddin Molla[1], F. M. Javed Mehedi Shamrat[3]

[1]Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Dhaka, Bangladesh

[2]Department of Computer Science and Engineering, Shanto Mariam University of Creative Technology, Permanent Campus, Dhaka, Bangladesh

[3]Department of Computer System and Technology, Universiti Malaya, Kuala Lumpur, Malaysia

muktar15-2298@diu.edu.bd

ABSTRACT

Humans moving in crowded spaces adapt their pace or alter their initial path. A variety of sociocultural factors and personal preferences influence these interactions. With the development of autonomous moving platforms that need to share their physical environment with humans, this task has become increasingly valuable. In this paper, we present a way to conveniently evaluate the ability of a model to predict social interactions between pedestrians. In this manner, we conduct experiments on two prediction models: Social long short-term memory (LSTM) and Vanilla LSTM models. We use real datasets by generating synthetic datasets for which we can define the impact of social interactions on the motion of pedestrians. These hand-tailored datasets exclude individual interactions and focus on the interactions between individuals. By comparing the models' performances on these datasets, we show that while the Social LSTM model can predict social interactions, the Vanilla LSTM model cannot. For our analysis, we introduce evaluation metrics that focus on the interactions between pedestrians, and these metrics go beyond the commonly used average and final displacement error for trajectory prediction. In particular, we analyze the prediction errors in regions of trajectories highly influenced by social interactions. Furthermore, we analyze the collision behavior of the model's predictions and classify trajectories concerning their degree of non-linearity. This allows us to compare the models' performances on motions differently influenced by social interactions.

KEYWORDS

multimedia trajectory prediction, human interaction, social force model, social long short-term memory (LSTM) model, vanilla LSTM model, crowded scenarios, encoder-decoder model

Ali, M.M., Hameem, M.T.K., Islam, D., Abir, M.A.H., Molla, M.M., Javed Mehedi Shamrat, F.M. (2025). Modeling Spatial-Temporal Social Interactions for Pedestrians Trajectory Prediction on Real and Synthetic Datasets. *IETI Transactions on Data Analysis and Forecasting (iTDAF)*, 3(1), pp. 49–69. https://doi.org/10.3991/itdaf.v3i1.53925

Article submitted 2024-12-17. Revision uploaded 2025-03-19. Final acceptance 2025-03-19.

1 INTRODUCTION

Navigating complex terrain has become associated with moving through crowded scenes while negotiating complex social interactions with multiple strangers. In recent years, the development of autonomous moving platforms such as self-driving cars or social robots has become important. As these technologies will share the same space as humans, predicting pedestrian trajectories will become safety-critical. Making such predictions is challenging due to the strong influence of social interactions on human motion [1]. In recent research, traditional techniques have primarily addressed the aspects of interpersonal and social acceptance by using hand-crafted functions to model social interactions between pedestrians [2], [3], and recent achievements in deep learning have motivated researchers to revisit the problem in a new data-driven fashion [4]–[6]. However, strictly depending on the nature of the data provided is a critical limitation. This dependence becomes a problem when evaluating data-driven models concerning a property not sufficiently represented in the data.

Group trajectory prediction models into two categories: traditional knowledge-driven models that use hand-crafted functions such as energy potentials to model the behavior of pedestrians, and data-driven models that aim to learn the motion behavior of pedestrians by processing a sufficient amount of data. In this paper, we investigate whether existing trajectory prediction models are capable of modeling social interactions between individuals. These interactions significantly influence the motions of pedestrians moving in crowded scenes. It is therefore crucial to understand them to reliably predict the future movements of pedestrians. However, as the degree of social interactions in commonly used datasets of real-world human motion is limited, we can only evaluate existing trajectory prediction models on a small subset of possible motion behaviors. To address this problem, we adopt the Social Force model [1] to generate synthetic datasets for which we can freely define the impact of social interactions on the motion of pedestrians. We then train data-driven models on the datasets generated and infer their ability to model social interactions from the difference in their performance between each dataset. In addition, we define methods to analyze the predictions of these models, evaluate the collision behavior of the model's predictions, and define a metric to measure the error in specific nonlinear regions. We classify trajectories by their degree of non-linearity and evaluate the performances of these classes. These metrics go beyond the classical L2 loss and allow us to conclude whether a model can learn social interactions or not.

2 LITERATURE REVIEW AND MOTIVATION

2.1 Knowledge-driven models

Human motion behavior can be studied from a crowd perspective or from an individual perspective. While the former macroscopic approach interprets a crowd of pedestrians as one continuous entity and focuses on collective movements [7]–[9], the latter microscopic perspective regards this crowd as a set of individuals. [10] In their work, the internal motivation of an individual to move with attractive and repulsive forces for the Social Force model. The influence of this

work has been huge, and it has often been revisited, for example in [11] and [12], to name only a few. Other microscopic models use different approaches to model the motions of pedestrians. For example, Bedau et al. [3] use cellular automata with a small rule set to model the collective motion behavior of pedestrians in bidirectional flow. Antonini et al. [2] use a discrete choice framework to model the short-term motion behavior of individuals with a choice set based on a combination of speed regimes and directions. Further knowledge-driven models aim to describe the motion of human crowds by using continuum dynamics. Respectively, Hughes introduces a model in [13] and [14] in which he defines an evolving potential function that guides pedestrians, represented by a density field, toward their destinations. Treuille et al. [15] revisit Hughes' work by transforming the continuous crowd field into a particle representation to create an efficient simulation for pedestrian dynamics.

While a variety of different approaches exist to describe the motion behavior of pedestrians with knowledge-driven models, they all resemble each other in one limitation. These models are based on a finite set of hand-crafted rules and functions that represent the current knowledge about human motion behavior. As state-of-the-art research in trajectory prediction shows [16], [17], we are yet far away from entirely understanding this complex behavior, including the personal motives and preferences of pedestrians and their interactions with each other. Therefore, data-driven methods have been used over the past few years to address the challenges of trajectory prediction. These models aim to learn human motion behavior by observing a sufficient amount of data and extracting key features from this data.

2.2 Data-driven models

Pedestrian trajectory prediction is highly sequential. It comprises iterative predicting of the next movement of a pedestrian based on previously made observations and predictions. With the advent of neural networks, the sequential nature of this task has motivated researchers to use recurrent neural networks (RNNs) and especially their variants, i.e., long short-term memory (LSTM) [18] and gated recurrent units (GRUs) [19], for the prediction. These networks are a rich class of dynamic models that have proven to be very successful for various sequence generation problems like speech recognition [20]–[22], machine translation [23], [24], or image/video captioning [25], [26], [27]. However, while RNNs are successful in learning and generalizing the properties of isolated sequences [28], they lack in jointly predicting multiple correlated sequences. Furthermore, they are deterministic models that predict only one single future trajectory for each pedestrian in a scene. For a specific scenario, there are multiple feasible paths a pedestrian might take. Deterministic models neglect the multi-modal character of human motion and tend to learn the average behavior of pedestrians. Contrarily, stochastic models aim to learn a distribution of future movements instead of a single trajectory. Lee et al. [29] introduce a stochastic model based on a recurrent conditional VAE that learns to generate multiple feasible samples of a trajectory and subsequently refines them. However, training a VAE can become difficult due to intractable probabilistic computations. Generative adversarial networks (GANs), where the training procedure is replaced by a minimax game between a generative model and a discriminative model, where proposed [30].

Gupta et al. [1] use this concept and propose a GAN-based model that predicts plausible trajectories by training adversarially against a recurrent discriminator. Other stochastic models [31], [32] use variants of GANs, i.e., BicycleGAN [33] and InfoGAN [34], that prevent the models from mode collapsing and, therefore, ensure diverse results.

2.3 Social interaction models

Pedestrians moving in crowded scenes socially interact with people in their vicinity. They adapt their movements according to the behavior of others. In this manner, various models and techniques have been developed that exploit social interactions between pedestrians. Alahi et al. [5] propose a social pooling layer that allows their LSTM-based model to jointly predict the trajectories of neighboring pedestrians. This pooling layer connects neighbors by aggregating the hidden states of their respective LSTM networks at every time step of the prediction process. Gupta et al. [1] use a multilayer perceptron (MLP), followed by max pooling to aggregate information across neighboring pedestrians in the generation step. Furthermore, they show that this technique performs comparatively well and is computationally more efficient than the proposed social pooling layer in [5]. Inspired by the latest successes with attention-based models [35], [36], Sadeghian et al. [37] present a social attention mechanism that aggregates information across different interactions between pedestrians and learns to extract information about the most influential neighbors. Amirian et al. [31] use a similar attention-based pooling scheme but additionally define hand-crafted interaction features that serve as a prior for the pooling process. Kosaraju et al. [32] use a graph attention network (GAT) [38] to learn the feature representations that encode the interactions between pedestrians.

2.4 Contribution

Overall, we see that numerous concepts and models address the problem of trajectory prediction. While some of these models focus on the social interactions between individuals [5], others look exclusively at the interactions between pedestrians and obstacles in their spatial environment [39], [40]. Moreover, many models consider both impacts [41]. One critical limitation of data-driven models is that they strictly depend on the data that is provided. This can become a problem when we want to evaluate models for a property not sufficiently represented in the data. Only [31] generates a hand-tailored dataset to study the problem of mode collapsing. This paper shows that the number of social interactions between pedestrians is limited in commonly used datasets. Hence, we design synthetic datasets that are specifically oriented to the evaluation of learning social interactions. While the Vanilla LSTM model predicts each trajectory independently of the motions of others, the Social LSTM model introduces a social pooling layer that aggregates information across neighboring pedestrians. To conclude the extent to which the social pooling layer allows us to predict these interactions, we compare the performances of both models on datasets with variable amounts of social interactions present. In datasets of real-world human motion, the level

of these interactions is inherently limited. We overcome this problem by presenting a method to generate synthetic datasets for which we can freely define this level. Furthermore, we propose evaluation metrics that focus on the interactions between pedestrians. These metrics go beyond the commonly used average and final displacement error for pedestrian trajectory prediction and ensure a rich and detailed analysis.

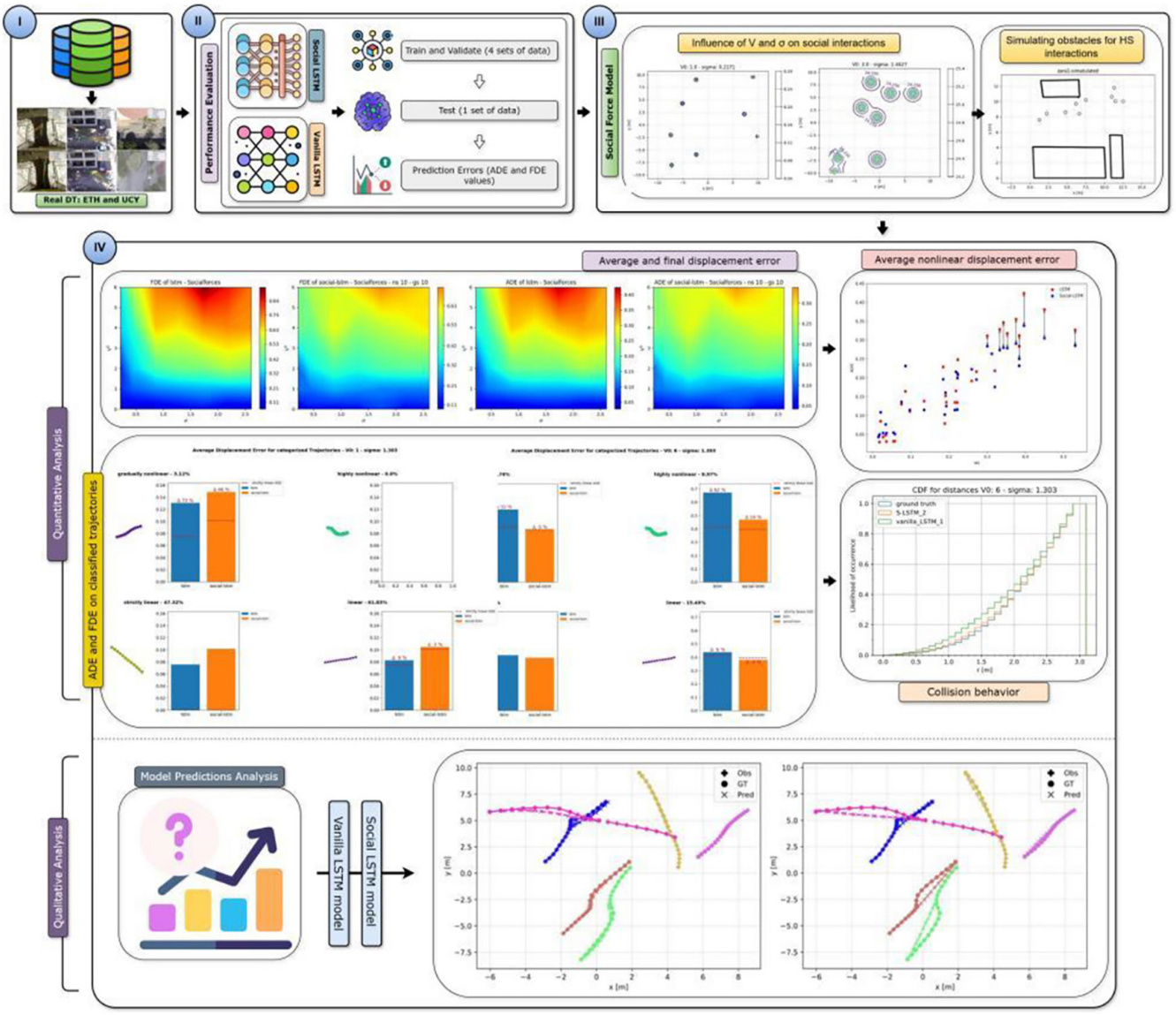

Fig. 1. The entire workflow of the proposed model for pedestrian's trajectory

3 EXPERIMENT

We introduce various evaluation metrics that go beyond the classical average and final displacement error for a detailed analysis. These metrics show quantitatively that the Social LSTM model outperforms the Vanilla LSTM model in predicting trajectories influenced by social interactions.

3.1 Real datasets

ETH [42] and UCY [43] datasets of real-world human trajectories and discuss their suitability for evaluating data-driven models that aim to learn social interactions. The Social Force model can be adapted to generate hand-tailored synthetic datasets that focus on pedestrian interactions and demonstrate how we can address human-space interactions when generating synthetic data. These datasets cover complex motion behaviors like group crossings or collision avoidance. The datasets are labeled manually at a frame rate of 2.5 fps to obtain real-world coordinates every 0.4 seconds. They cover four different scenarios with five sets of data. In particular, the ETH dataset comprises two sets of data, denoted as ETH and Hotel, with 750 pedestrians present. The UCY dataset includes three different sets of data, i.e., Zara01, Zara02, and Univ, containing 786 pedestrians in total. A more detailed overview of the properties of each dataset can be found in Table 1.

Table 1. Overview of the properties of the ETH and UCY datasets

	Dataset	Number of Frames	Number of Pedestrians
ETH	ETH	1934	360
	HOTEL	1807	390
UCY	UNIV	540	434
	ZARA01	902	148
	ZARA02	1052	204

We can observe that the Univ dataset includes the highest number of pedestrians while comprising the smallest number of frames. Therefore, we expect crowded scene behavior for this dataset. In this paper, we predict a trajectory by observing 3.2 seconds (8 time steps) of motion and forecasting the next 4.8 seconds (12 time steps). The complete trajectory of a pedestrian thus consists of 8 seconds (20 time steps). In the given datasets, we regularly find pedestrians who are only present for a shorter period or leave the scene for a moment and return later. This results in trajectories that consist of less than 20 time steps or miss information at certain points. There are several approaches to solve this problem, such as padding or interpolating the data. However, this can introduce non-negligible errors that produce misleading results. Therefore, we only evaluate the performance of a model on complete trajectories that consist of at least 20 consecutive time steps. This results in a significant reduction of usable trajectories in a given dataset.

In Table 2, we show that the number of suitable trajectories in each dataset is significantly less than the total number of trajectories present. In particular, each trajectory that consists of at least 20 consecutive time steps is counted as a suitable trajectory. The total number of trajectories also includes trajectories that consist of less than 20 time steps. As an example, a trajectory that consists of 16 consecutive time steps is represented in the total number of trajectories but does not count as a suitable trajectory. A trajectory that consists of 40 consecutive time steps represents two suitable trajectories since we can split it into two paths, each comprising 20 consecutive time steps.

Table 2. Comparison between the total number of trajectories in the datasets and the number of suitable trajectories (ST) for the prediction task

Dataset	Total Trajectories	No. of ST	% of ST
ETH	649	297	45.76%
HOTEL	511	145	28.38%
UNIV	1302	903	69.35%
ZARA01	320	178	55.63%
ZARA02	572	374	65.38%

Performance evaluation on real datasets. The Vanilla LSTM model and Social LSTM model on these datasets, we train and validate the models on four sets of data and test them on the remaining set. We use the following two evaluation metrics:

- **Average displacement error:** evaluates the average Euclidean distance between prediction $\hat{y}_i^t$ and ground truth y_i^t over all time steps t of the prediction process for all suitable trajectories N in a dataset:

$$ADE = \frac{1}{N}\frac{1}{T_{pred} - (T_{obs}+1)}\sum_{i=1}^{N}\sum_{t=T_{obs}+1}^{T_{pred}} \| y_i^t - \hat{y}_i^t \| \quad (1)$$

- **Final displacement error:** Evaluates the Euclidean distance between the prediction of the final position $\hat{y}_i^{T_{pred}}$ and the respective ground truth data $y_i^{T_{pred}}$ for all suitable trajectories N in a dataset:

$$FDE = \frac{1}{N}\sum_{i=1}^{N} \| y_i^{T_{pred}} - \hat{y}_i^{T_{pred}} \| \quad (2)$$

To obtain an adequate comparison between the two models, we use similar configurations for both models. In particular, we use an embedding dimension of 32 for the spatial coordinates before using them as input to the LSTM cells of the models. We define the dimension of the hidden state tensor of the encoder to be 64 and the decoder to be 32. In addition, a layer with ReLU non-linearity between these states to match the respective dimensions. For the Social LSTM model, we set the neighborhood size *ns* to be 10 meters and use a grid size of 10 meters, such that each cell C_k spans one square meter. Furthermore, we use a layer with ReLU non-linearity on top of the calculated hidden state tensor H_i^T before concatenating it with the hidden state information of the previous LSTM cell $H_i^T - 1$. The hyper-parameters are chosen for each dataset separately, based on the evaluation of the corresponding validation set. The evaluation results for both models are shown in Table 3.

Table 3. Prediction errors of the models for the ETH and UCY datasets

	Dataset	Vanilla LSTM	Social LSTM
ETH	ETH	1.22/2.60	1.22/2.54
	HOTEL	0.37/0.71	0.28/0.55
UCY	UNIV	0.61/1.22	0.69/1.42
	ZARA01	0.47/0.99	0.56/1.15
	ZARA02	0.36/0.75	0.45/0.92
	AVG	0.61/1.25	0.64/1.32

The ADE and FDE values are divided, where the first value indicates the ADE. All values represent meters. The last row denotes the average values over all datasets.

As can be seen, the Vanilla LSTM model outperforms the Social LSTM model on the UCY datasets (Zara01, Zara02, and Univ). While both models have the same average displacement error for the ETH dataset, the Social LSTM model performs slightly better on this dataset than the FDE. The Social LSTM model outperforms the Vanilla LSTM model for the hotel dataset. We focus on the capacity of a model to predict social interactions. Therefore, the number of social interactions present in each dataset is crucial. As shown in [42], the ETH and UCY datasets cover complex motion behavior, including social interactions between multiple pedestrians. However, the evaluation of these datasets, depicted in Table 3, does not indicate whether one of the analyzed models can learn social interactions. In particular, we would assume a model that is not capable of modeling social interactions to perform poorly on datasets with many social interactions present. Yet, we do not know the specific number of social interactions present in the given datasets. To address this problem, we generate hand-tailored datasets for which we can define the intensity and range of the social interactions between pedestrians. This allows us to compare the performance of the models on datasets with variable amounts of social interactions present. By generating synthetic datasets, we also overcome the above-mentioned problem of limited suitable trajectories. Generating synthetic datasets allows us to create an almost unlimited amount of data and, therefore, an almost unlimited amount of suitable trajectories.

3.2 Synthetic datasets

We generate synthetic datasets that are focused on social interactions between pedestrians. The social force model uses attractive and repulsive forces to describe the motion behavior of pedestrians. The motion of a pedestrian is composed of four components: an acceleration term towards the destination, a repulsive force modeling human-human interactions, a repulsive force describing human-space interactions, and an attractive force towards specific regions of interest. By manipulating the repulsive force between pedestrians, we can conveniently define the influence of social interactions in the modeled data. As we focus on these interactions, we neglect the impact of obstacles or attractive effects in a scene.

$$\overrightarrow{F_\alpha}(t) : \overrightarrow{F_\alpha^0}\left(\vec{v}_\alpha, v_\alpha^0 \overrightarrow{e_\alpha}\right) + \sum_\beta \overrightarrow{F_{\alpha\beta}}\left(\overrightarrow{e_\alpha}, \overrightarrow{r_\alpha} - \overrightarrow{r_\beta}\right) \tag{3}$$

The closer two pedestrians get to each other, the stronger they interact. This effect, we choose the repulsive potential to be a monotonically decreasing exponential function:

$$V_{\alpha\beta}\left(\| \overrightarrow{r_{\alpha\beta}} \|\right) = V^0 e^{\frac{-\| \vec{r}_{\alpha\beta} \|}{\nu}} \tag{4}$$

where V^0 and σ are free parameters and $\| \rightarrow r_{\alpha\beta} \|$ denotes the distance between two pedestrians α and β. Note that by varying the values of V^0 and σ, we can control the shape of this potential and, the impact of social interactions on the motion of pedestrians in the dataset.

Hence, these parameters allow us to efficiently determine the impact of social interactions on the motion behavior of pedestrians. As an example, Figure 2 depicts scenes from two datasets with different values for V^0 and σ, in which the values of the repulsive potential are visualized.

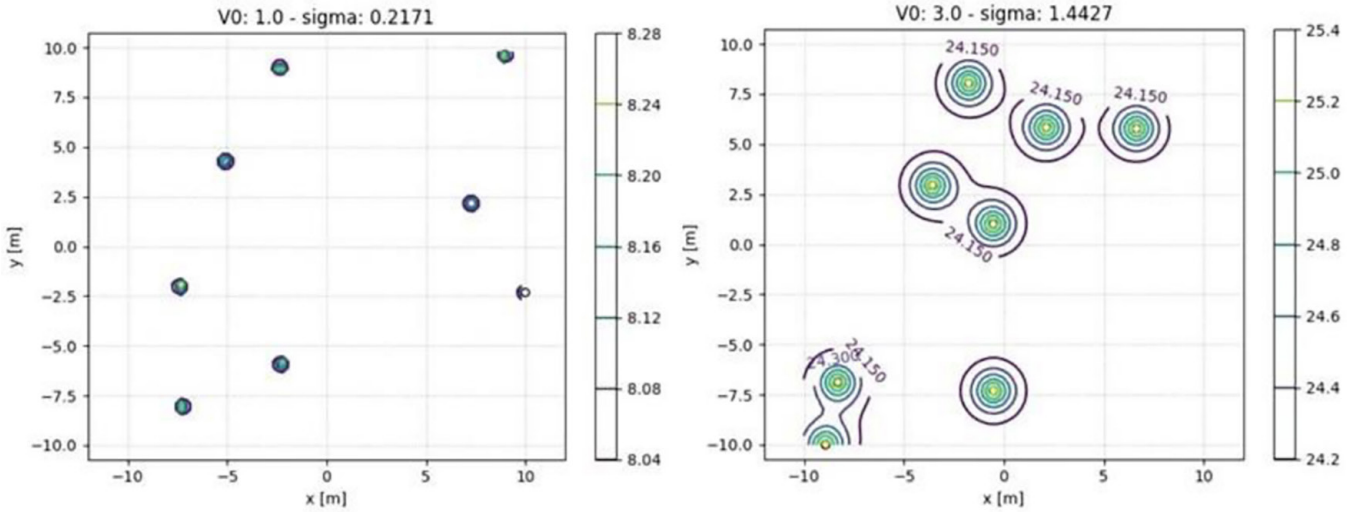

Fig. 2. Scenes from two different datasets simulated by our adaption of the Social Force model

While the left scene represents a dataset with low values V0 = 1 and $\sigma = 0.2171$, the right scene is from a dataset with comparatively high values V0 = 3 and $\sigma = 1.4427$. The visualized repulsive potentials differ in magnitude and range, respectively.

Simulating obstacles for human-space interactions. The impact of obstacles in the datasets generated, we need to consider the repulsive effects between pedestrians and obstacles in the resultant force of the Social Force model:

$$\vec{F}_\alpha(t): \quad \vec{F}_\alpha^0\left(\vec{v}_\alpha, v_\alpha^0 \overrightarrow{e_\alpha}\right) + \sum_\beta \overrightarrow{F_{\alpha\beta}}\left(\overrightarrow{e_\alpha}, \overrightarrow{r_\alpha} - \overrightarrow{r_\beta}\right) + \sum_B \overrightarrow{F_{\alpha B}}\left(\overrightarrow{e_\alpha}, \overrightarrow{r_\alpha} - \overrightarrow{r_B}\right) \tag{5}$$

Similar to the repulsive effects between pedestrians, the respective forces are modeled by the gradient of a monotonically decreasing repulsive potential $U_{\alpha\beta}(\|\rightarrow r_{\alpha\beta}\|)$. Based on the choice of this potential, the impact of human-space interactions on pedestrians is defined. The data generated to model a specific situation or property should be as realistic as possible.

To model scenarios of real-world datasets, we implement a method that reads segmented scene context information and embeds it in the process of data generation. In particular, the segmentation of a real-world image is processed by approximating the boundary boxes of specific obstacle classes. These approximations are then used to define the borders of obstacles B. An example of this process is depicted in Figure 3. As can be seen, the scene context from the Zara01 scenario on the left is segmented and then embedded in the process of generating a corresponding synthetic dataset, as depicted on the right. This allows for the generation of synthetic datasets based on real-world scenarios.

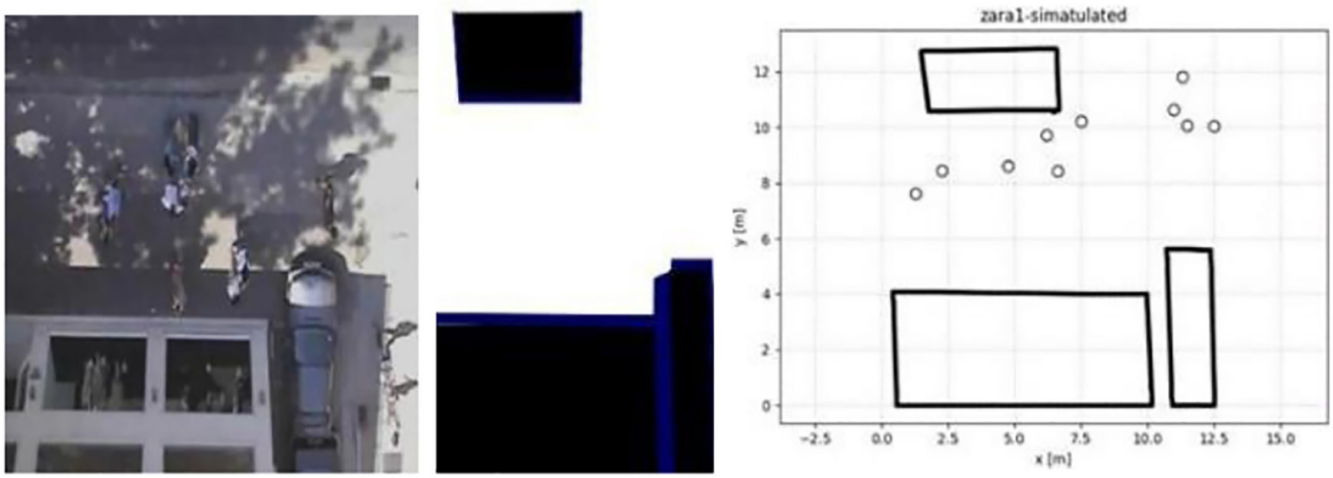

Fig. 3. Scenes from two different datasets simulated by our adaption of the Social Force model

While the left scene represents a dataset with low values V0 = 1 and σ = 0.2171, the right scene is from a dataset with comparatively high values V0 = 3 and σ = 1.4427. The visualized repulsive potentials differ in magnitude and range, respectively.

3.3 Evaluation metrics

We train and evaluate the introduced models on synthetic datasets that focus on the interactions between pedestrians. In particular, we generate datasets that in total represent 35 different combinations of the values for V^0 and σ. With this, we obtain datasets that gradually differ in the amount of social interactions present. A detailed overview of these combinations can be found in Table 4. Note that the impact of social interactions on pedestrians increases with rising values of V^0 and σ. Each dataset represents a scenario, in which a predefined constant number of pedestrians N_{const} move in a square of 20×20 meters. By varying this number, we can define the density of pedestrians in a scene. For each pair of V^0 and σ, we generate three sets of data: a set on which we train the models, a validation set, and a test set on which we evaluate the models. An overview of the configurations of these sets is given in Table 5.

Table 4. Detailed overview of the different combinations of the values for V^0 and σ

V^0	σ						
6	0.2171	0.4343	0.8686	1.303	1.7371	2.171	2.6058
4	0.2171	0.4343	0.8686	1.303	1.7371	2.171	2.6058
2	0.2171	0.4343	0.8686	1.303	1.7371	2.171	2.6058
1	0.2171	0.4343	0.8686	1.303	1.7371	2.171	2.6058
0	0.2171	0.4343	0.8686	1.303	1.7371	2.171	2.6058

For each pair of these values, we generate datasets on which we evaluate the models introduced.

3.4 Quantitative evaluation

The quantitative evaluation of the Vanilla LSTM model and the Social LSTM model by analyzing their overall performance on synthetic datasets with varying amounts of social interactions present. For this, we evaluate the average and final displacement error on 35 different datasets. Each dataset is generated according to the method described, using a high number of pedestrians in each scene $N_{const} = 14$ and a unique combination of the values for V^0 and σ, as defined in Table 4. For the Vanilla LSTM model and the Social LSTM model, we use the same configurations as for the evaluation of datasets of real-world human motion. To conveniently characterize the datasets generated for the impact of social interactions present, we classify the future trajectories Y of a dataset classification scheme and calculate the following weighted sum ws:

$$ws = w_1 \cdot \frac{|L|}{|L \cup GNL \cup HNL|} + w_1 \cdot \frac{|GNL|}{|L \cup GNL \cup HNL|} + w_1 \cdot \frac{|HNL|}{|L \cup GNL \cup HNL|} \tag{6}$$

Table 5. General configurations of the different sets of data

Phase	No. Frames	Space of Square [m]
Train	9000 (1h)	20×20
Val	1800 (12 min)	20×20
Test	1800 (12 min)	20×20

The time between two frames is 0.4 seconds. As the training set comprises 9000 frames, it represents one hour of motion behavior. The validation and test set are limited to 12 minutes of motion.

This weighted sum represents the distribution of classified trajectories in a dataset. It therefore gives a first impression of the influence of social interactions on the motion of the pedestrians in a dataset. As an example, a value of $ws = 0$ indicates that all successfully classified trajectories of a dataset are linear. Higher values for ws imply that there are also gradually or highly nonlinear trajectories in the dataset. In particular, the closer the value of ws gets to one, the higher the amount of highly nonlinear trajectories in the dataset. As trajectories that are assigned to the other class do not represent an interpretable trajectory class, we neglect them for the calculation of ws. The weighted sum for each dataset is depicted in Figure 4. We can see a clear trend that with rising values of V^0 and σ, the value of ws and, therefore, the amount of nonlinear trajectories in the dataset increases. This confirms that the created datasets differ in their amount of social interactions. While datasets with high values of V^0 and σ comprise many social interactions that cause trajectories to become highly nonlinear, datasets with low values for V^0 and σ barely include these interactions, such that the motion of the pedestrians is predominantly linear. This is reflected in the values of ws.

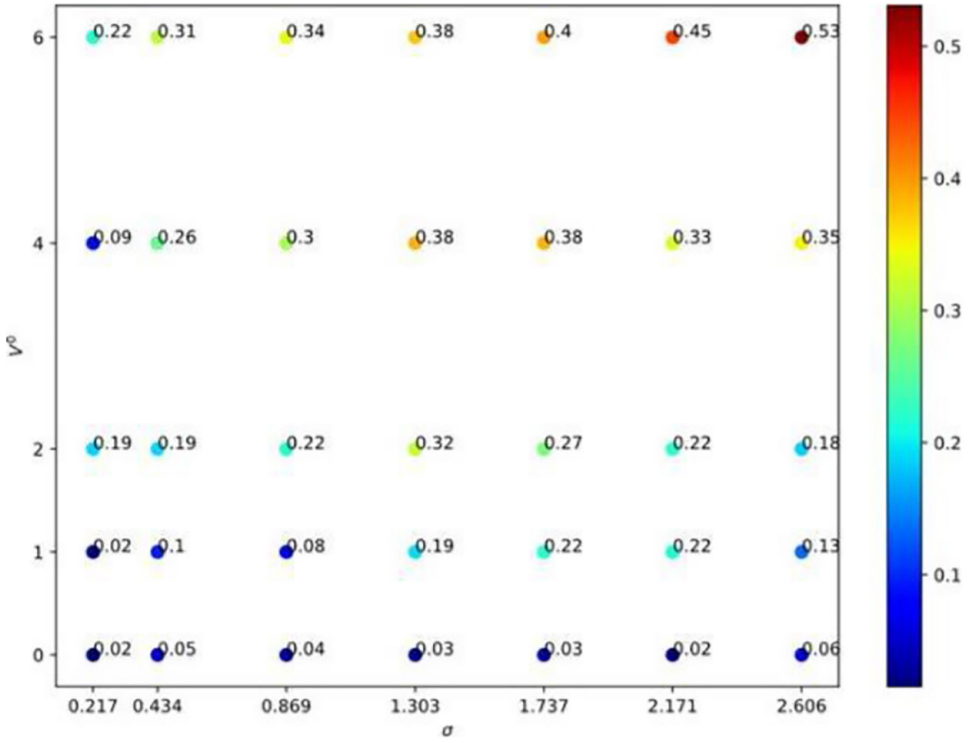

Fig. 4. Weighted sums calculated for each generated dataset according to Equation 6

For increasing values of V^0 and σ, we observe rising values of *ws*. This shows that the number of gradual and highly nonlinear trajectories in a dataset increases with increasing values of V^0 and σ.

Average and final displacement error. The average and final displacement error of the Vanilla LSTM model and the Social LSTM model are calculated for all datasets. The respective results are depicted in Figure 5. While the Vanilla LSTM model performs comparatively well on datasets with weak or spatially limited social interactions, the performance significantly decreases on datasets with strong interactions. As we analyze the results of the model's average displacement error in Figure 5, we can particularly see that with rising values of V^0 and σ, the ADE rapidly increases. This trend is also reflected in the final displacement error of the model, as illustrated in Figure 5. Note that the final displacement error is in general higher than the average displacement error of the model.

In comparison, we illustrate the ADE of the Social LSTM model in Figure 5. While we can observe a similar trend towards higher average displacement errors for a rising impact of social interactions in the datasets, we note that the decrease in the performance is not as significant as for the Vanilla LSTM model. In particular, while both models perform approximately equivalently on datasets with low social interactions, the Social LSTM model outperforms the Vanilla LSTM model on datasets with strong and far-ranging interactions. This difference is also shown in the final displacement error of the two models. We can see that the FDE of the Social LSTM model is significantly lower than the FDE of the Vanilla LSTM model for datasets with strong interactions between pedestrians.

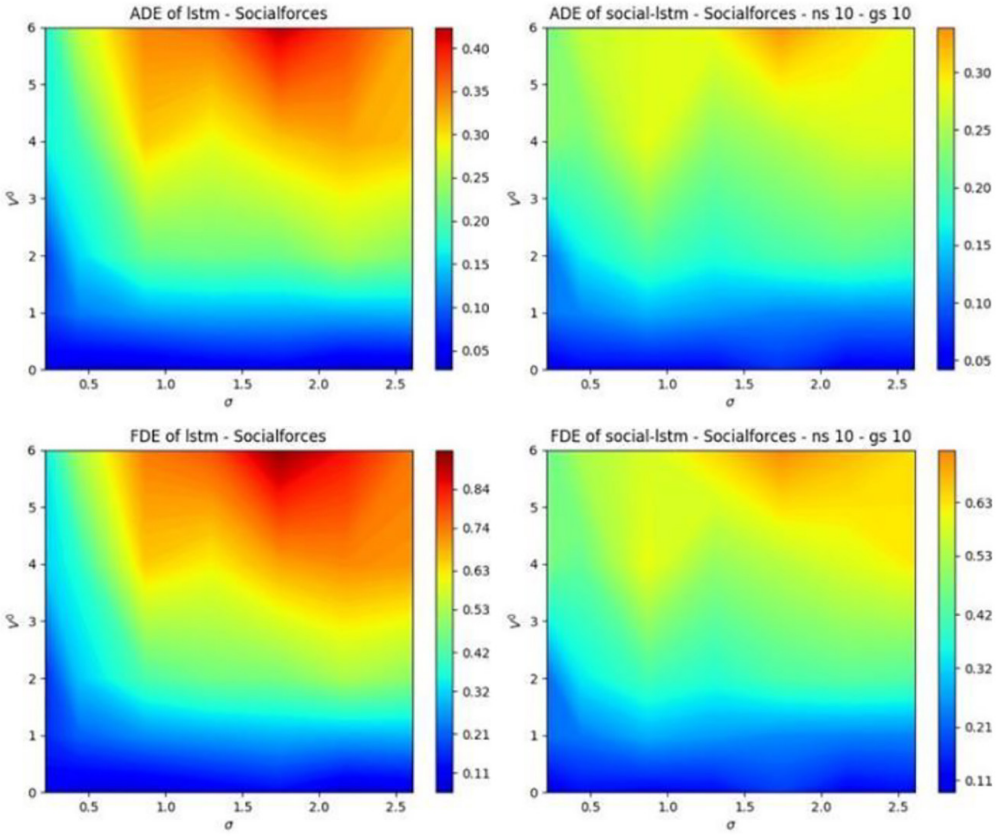

Fig. 5. Scenes from two different datasets simulated by our adaption of the Social Force model

Average non-linear displacement error. The overall average and final displacement error of the Social LSTM model outperforms the Vanilla LSTM model on datasets with strong social interactions. To gain more insights on how well these models predict social interactions. When pedestrians interact with each other, they adapt their motions according to the behavior of others. Therefore, social interactions between pedestrians cause their respective trajectories to become nonlinear. The average nonlinear displacement error evaluates the error of a model in specific nonlinear regions of the trajectories. To identify these regions, a threshold value, td, is defined. The corresponding error is obtained by calculating the ADE for the identified regions, excluding the points of the trajectories for which the curvature is below the threshold.

Table 6. Values for the ADE and FDE of the Vanilla LSTM model on 35 different datasets, generated with a constant number of pedestrians in each scene of $N_{const} = 14$

Vanilla LSTM Model							
V^0	σ						
	0.2171	0.4343	0.8686	1.303	1.7371	2.171	2.6058
6	0.17/0.33	0.26/0.53	0.35/0.74	0.36/0.78	0.42/0.94	0.38/0.85	0.33/0.74
4	0.16/0.31	0.19/0.37	0.31/0.66	0.28/0.62	0.31/0.70	0.33/0.72	0.32/0.71
2	0.08/0.15	0.15/0.30	0.21/0.42	0.22/0.46	0.22/0.47	0.25/0.53	0.23/0.49
1	0.08/0.16	0.11/ 0.21	0.13/0.25	0.13/0.27	0.13/0.29	0.13/0.28	0.14/0.29
0	0.04/0.11	0.03/0.07	0.03/0.06	0.04/0.09	0.05/0.12	0.03/0.06	0.03/0.07

The values of the ADE and FDE are separated by a slash, where the first value indicates the ADE. All values represent meters. The performance of the Vanilla LSTM model is compared with the performance of the Social LSTM model in Table 7.

Table 7. Values for the ADE and FDE of the Social LSTM model on 35 different datasets, generated with a constant number of pedestrians in each scene of $N_{const} = 14$

Social LSTM Model							
V^0	σ						
	0.2171	0.4343	0.8686	1.303	1.7371	2.171	2.6058
6	0.22/0.45	0.26/0.53	0.28/0.58	0.29/0.65	0.34/0.73	0.30/0.68	0.29/0.64
4	0.23/0.47	0.22/0.44	0.28/0.60	0.23/0.50	0.25/0.54	0.27/0.59	0.28/0.62
2	0.10/0.21	0.16/0.34	0.20/0.41	0.18/0.37	0.20/0.42	0.21/0.44	0.20/0.43
1	0.11/0.22	0.11/0.23	0.14/0.29	0.13/0.26	0.12/0.25	0.11/0.24	0.11/0.24
0	0.04/0.09	0.05/0.13	0.05/0.11	0.05/0.12	0.08/0.17	0.05/0.11	0.05/0.13

The values of the ADE and FDE are separated by a slash, where the first value indicates the ADE. All values represent meters. The performance of the Social LSTM model is compared with the performance of the Vanilla LSTM model in Table 6.

Collision behavior. Pedestrians rarely collide with each other, as they normally respect personal space and keep a certain distance from strangers. Models that are capable of learning social interactions should be able to predict trajectories that reproduce this avoidance behavior. We define a collision to have occurred if the distance between two pedestrians is less than $r_{coll} = 1.0$ meters. In this manner, we analyze two crowded datasets with strong social interactions that differ only in the range of the respective forces. The specific parameters of these datasets can be found in Table 8.

Table 8. Average and final displacement error of the Vanilla LSTM model and the Social LSTM model for dataset B

Dataset	N_{const}	V^0	σ
Dataset A	20	6	1.303
Dataset C	20	6	2.6058

We compare the results of both models and the ground truth data in Table 9. The given values represent the number of collisions between pedestrians during the prediction process in the respective datasets. We observe that for both datasets, the Social LSTM model predicts collision behavior that is close to that of the ground truth data. Contrarily, we see that the Vanilla LSTM model predicts almost twice as many collisions as given in the datasets.

We plot the cumulative distribution function (CDF) for the distances between the pedestrians of dataset A in Figure 6. To focus on how pedestrians are influenced by social interactions, we exclusively consider the interval $r = [0, 3]$. The respective distributions of the Vanilla LSTM model and the Social LSTM model reflect the results in Table 9. We observe that the CDF of the Vanilla LSTM model deviates from the distribution of the ground truth data. In particular, we can see that the pedestrians predicted by the model are in general closer to each other, compared to the ground truth data. In comparison, the CDF of the Social LSTM model almost completely fits the distribution of the actual data.

Table 9. Amount of collisions between the pedestrians during the prediction process

Dataset	LSTM	S-LSTM	Ground Truth
Dataset A	12.3%	7.9%	6.8%
Dataset C	3.0%	1.6%	1.5%

While the Social LSTM model predicts a frequency of collisions that is close to the ground truth data, the Vanilla LSTM model predicts collisions more frequently than they occur in the actual data.

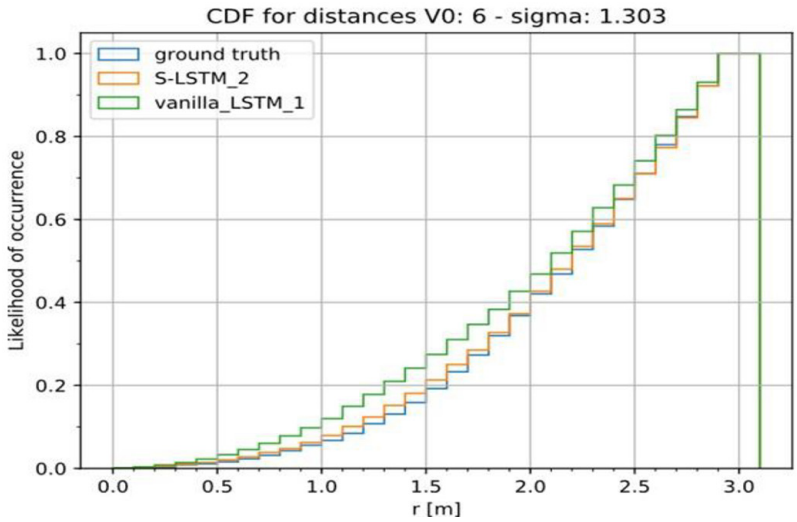

Fig. 6. Cumulative distribution function for the Euclidean distance between the pedestrians in dataset A for the ground truth data and both models

Note: The CDF of the Vanilla LSTM model deviates from the CDF of the ground truth data.

We exclusively consider the interval $r = [0, 3]$. The respective distributions of the Vanilla LSTM model and the Social LSTM model reflect the results in Table 9. We observe that the CDF of the Vanilla LSTM model deviates from the distribution of the ground truth data. In particular, we can see that the pedestrians predicted by the model are in general closer to each other compared to the ground truth data. In comparison, the CDF of the Social LSTM model almost completely fits the distribution of the actual data.

3.5 Qualitative evaluation

The quantitative evaluation in equation 6 shows that the Social LSTM model outperforms the Vanilla LSTM model on motion behavior that is significantly influenced by the social interactions between pedestrians. To gain more insight into the actual prediction behavior of these models, we qualitatively compare the predictions of the models on social scenes where pedestrians interact with each other. To qualitatively evaluate the collision avoidance capacity of the models, we present an example scenario in Figure 7, in which two pedestrians pass each other and avoid close contact by altering their paths respectively.

As can be seen, the Vanilla LSTM model is not able to reproduce the avoidance behavior of the pedestrians. Instead, it predicts two predominantly

linear trajectories that are not influenced by the motion of neighboring individuals. As the model does not take the motion behavior of neighboring pedestrians into account, it fails to predict the actual interactions between the pedestrians. In comparison, we see that the Social LSTM model predicts two trajectories that avoid a possible collision. The model considers the motion of neighboring pedestrians and replicates the corresponding avoidance behavior. These observations match the results of the quantitative evaluation of the models' collision behavior.

Figure 8 illustrates the prediction results of the Vanilla LSTM model and the Social LSTM model for an example scenario of a dataset generated with strong social interactions. In particular, we define $V^0 = 6$, $\sigma = 2.171$, and $N_{const} = 8$. In the center of the scene, we observe two pedestrians strongly interacting with each other. Similar to the prediction behavior depicted in Figure 7, we see that the Vanilla LSTM does not consider the impact of neighboring pedestrians on the scene. Therefore, it mainly predicts linear trajectories, which in the case of the two pedestrians in the centre would lead to a collision. Contrarily, we observe that the Social LSTM model predicts a motion behavior that considers the movements of neighboring individuals. For the two pedestrians in the center, it predicts trajectories that reflect the social interactions between the pedestrians and avoid a possible collision. Both models often fail to predict accurate trajectories because they lack information about the destination of the pedestrians. As a consequence, the models seem to fail in predicting suitable trajectories, although they consider social interactions between pedestrians during the prediction process.

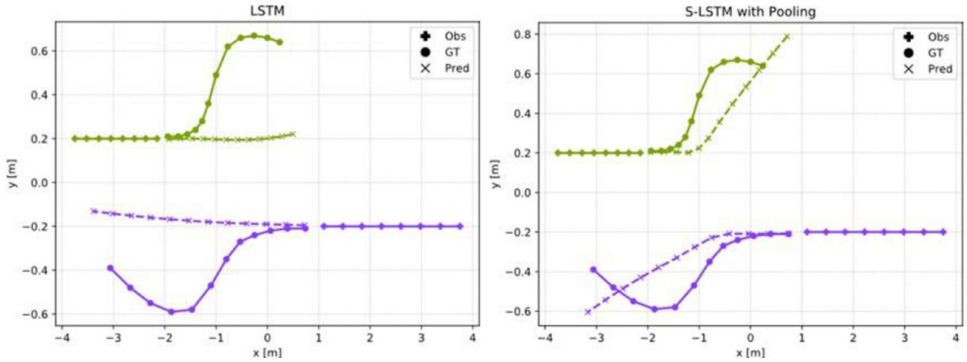

Fig. 7. Example scenario of two pedestrians passing each other and altering their paths to avoid close contact

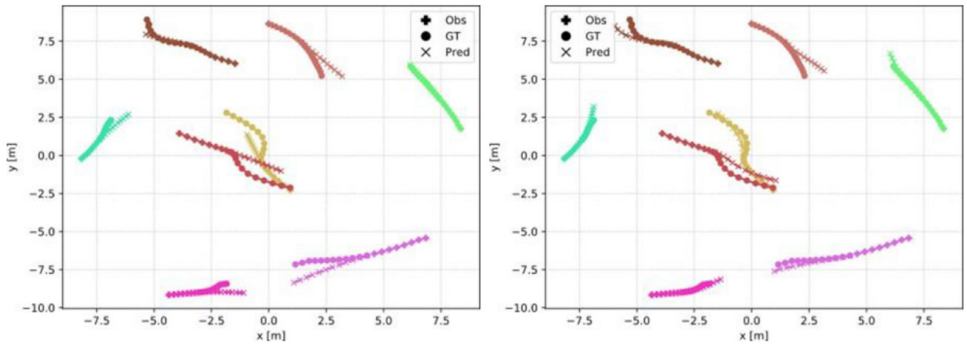

Fig. 8. Comparison between the predicted trajectories of the Vanilla LSTM (a) and the Social LSTM model (b) for a dataset with strong social interactions between the pedestrians

The Social LSTM model predicts trajectories that reflect social interactions and avoid collisions.

To focus on the performance of these models on social interactions, we analyze how the prediction behavior changes when we provide the models with information about the final position $y \rightarrow i^{Tpred}$ of each pedestrian i. An example of the respective results for a dataset with strong and far-ranging social interactions ($V^0 = 6; \sigma = 2.6058$) is illustrated in Figure 9. We can see that both models accurately predict the final location of each pedestrian. However, similar to previous results, we observe that the Vanilla LSTM model mainly predicts linear trajectories toward the final position of each pedestrian and fails to model social interactions between individuals. In comparison, the Social LSTM model can accurately predict the observed motions.

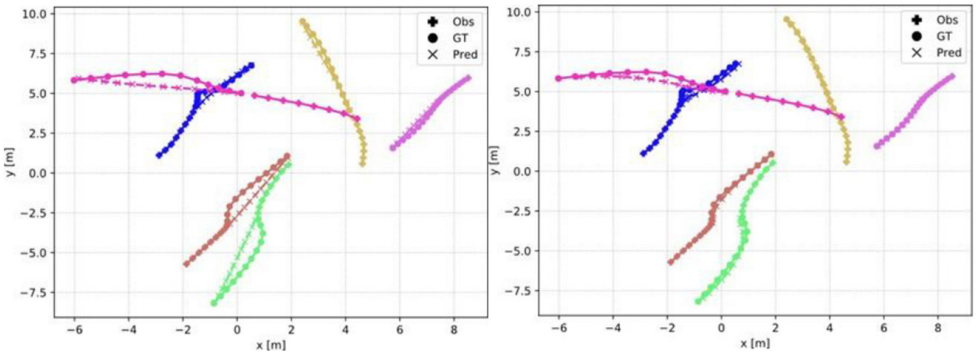

Fig. 9. Predictions of the Vanilla LSTM model (a) and the Social LSTM model (b) for a dataset with strong and far-ranging social interactions between pedestrians

Both models are provided with information about the destination of each pedestrian in the scene. While the Vanilla LSTM model neglects social interactions between pedestrians, the Social LSTM model accurately predicts these interactions.

4 CONCLUSIONS AND FUTURE WORK

We generated datasets for the impact of social interactions on the motion of pedestrians. By analyzing the performances of the Vanilla LSTM model and the Social LSTM model on these datasets, we evaluate their ability to predict social interactions. While both models perform similarly on datasets with weak social interactions, the Social LSTM model outperforms the Vanilla LSTM model on datasets with strong social interactions. Furthermore, the Social LSTM model performs significantly better than the Vanilla LSTM model on trajectories that are gradually or highly influenced by social interactions. This becomes particularly apparent for datasets in which social interactions have a strong impact on the motion of pedestrians. Additionally, we analyzed the collision behavior within the datasets and compared it with the predicted collision behavior of both models. We saw that while the Vanilla LSTM model predicts collisions between pedestrians much more frequently than they occur in the actual data, the Social LSTM model can reproduce the collision behavior of the dataset. A qualitative evaluation of the trajectories predicted by both models supports these results. In particular, the Vanilla LSTM model fails to predict the avoidance behavior of two pedestrians passing each other.

Given these results, the Social LSTM model is capable of predicting social interactions between pedestrians. In comparison, the Vanilla LSTM model is not

sensitive to neighboring individuals and, thus, not able to predict these interactions. For future studies, we suggest evaluating additional trajectory prediction models on the datasets generated, using the evaluation metrics proposed in this paper and the presented results as baselines. In particular, it might be interesting to investigate how other models that exploit social interactions perform compared to the Social LSTM model. Once several models are compared in this manner, the results could be used to identify the architectural requirements of a model that can precisely predict social interactions. Furthermore, we suggest using the methods introduced to generate hand-tailored datasets that simulate human-space interactions to evaluate and compare trajectory prediction models that aim to predict interactions between pedestrians and obstacles.

5 REFERENCES

[1] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," *arXiv preprint arXiv:1803.10892*, 2018. https://doi.org/10.48550/arXiv.1803.10892

[2] G. Antonini, M. Bierlaire, and M. Weber, "Discrete choice models for pedestrian walking behavior," *Transportation Research Part B: Methodological*, vol. 40, no. 8, pp. 667–687, 2006. https://doi.org/10.1016/j.trb.2005.09.006

[3] V. J. Blue and J. L. Adler, "Cellular automata model of emergent collective bi-directional pedestrian dynamics," in *Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life*, M. A. Bedau, J. S. McCaskill, N. H. Packard, and S. Rasmussen, Eds., 2000, pp. 437–445. https://doi.org/10.7551/mitpress/1432.003.0061

[4] H. Dirk and M. Péter, "Social force model for pedestrian dynamics," *Physical Review*, vol. 51, no. 5, pp. 5282–4286, 1995. https://doi.org/10.1103/PhysRevE.51.4282

[5] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded space," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 961–971. https://doi.org/10.1109/CVPR.2016.110

[6] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. D. Reid, H. Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal trajectory forecasting using Bicycle-GAN and graph attention networks," *arXiv preprint arXiv:1907.03395*, 2019. https://doi.org/10.48550/arXiv.1907.03395

[7] A. Sadeghian, V. Kosaraju, N. Hirose, H. Rezatofighi, and S. Savarese, "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints," *arXiv preprint arXiv:1806.01482,* 2019. https://doi.org/10.48550/arXiv.1806.01482

[8] D. Helbing, "A fluid dynamic model for the movement of pedestrians," *Complex Systems*, vol. 6, pp. 391–415, 1992. https://doi.org/10.48550/arXiv.cond-mat/9805213

[9] D. Helbing, "Physikalische modellierung des dynamischen verhaltens von fußgängern (physical modeling of the dynamic behavior of pedestrians)," 1990. Available at SSRN: https://doi.org/10.2139/ssrn.2413177

[10] A. Treuille, S. Cooper, and Z. Popović, "Continuum crowds," in *ACM SIGGRAPH 2006 Papers (SIGGRAPH '06)*, New York, NY, USA: Association for Computing Machinery, 1995, pp. 1160–1168. https://doi.org/10.1145/1179352.1142008

[11] L. Leal-Taixé, G. Pons-Moll, and B. Rosenhahn, "Everybody needs somebody: Modeling social and grouping behavior on a linear programming multiple people tracker," in *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, 2011, pp. 120–127. https://doi.org/10.1109/ICCVW.2011.6130233

[12] M. Luber, J. A. Stork, G. D. Tipaldi, and K. O. Arras, "People tracking with human motion predictions from social forces," in *IEEE International Conference on Robotics and Automation*, Anchorage, AK, USA, 2010, pp. 464–469. https://doi.org/10.1109/ROBOT.2010.5509779

[13] R. Raghavendra, A. D. Bue, M. Cristani, and V. Murino, "Abnormal crowd behavior detection by social force optimization," in *Human Behavior Understanding (HBU 2011)*, in Lecture Notes in Computer Science, A.A. Salah and B. Lepri, Eds., vol. 7065. Heidelberg, Berlin: Springer, 2011, pp. 134–145. https://doi.org/10.1007/978-3-642-25446-8_15

[14] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *Computer Vision – ECCV 2016. ECCV 2016*. in Lecture Notes in Computer Science(), B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., vol. 9912. Switzerland, Charm: Springer, 2016, pp. 549–565. https://doi.org/10.1007/978-3-319-46484-8_33

[15] A. Treuille, S. Cooper, and Z. Popović, "Continuum crowds," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1160–1168, 2006. https://doi.org/10.1145/1141911.1142008

[16] J. Li, F. Yang, M. Tomizuka, and C. Choi, "Evolvegraph: Heterogeneous multi-agent multimodal trajectory prediction with evolving interaction graphs," *arXiv preprint arXiv:2003.13924*, 2020. https://doi.org/10. 48550/arXiv.2003.13924

[17] M. Lisotto, P. Coscia, and L. Ballan, "Social and scene-aware trajectory prediction in crowded spaces," *arXiv preprint arXiv:1909.08840*, 2019. https://doi.org/10.48550/arXiv.1909.08840

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Compute*, vol. 9, no. 8, pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[19] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014. https://doi.org/10.48550/arXiv.1412.3555

[20] J. Chorowski, D. Bahdanau, K. Cho, and Y. Bengio, "End-to-end continuous speech recognition using attention-based recurrent," *arXiv preprint arXiv:1412.1602*, 2014. https://doi.org/10.48550/arXiv.1412.1602

[21] Y. Zhang, M. Pezeshki, P. Brakel, S. Zhang, C. L. Y. Bengio, and A. Courville, "Towards end-to-end speech recognition with recurrent neural networks," *arXiv preprint arXiv:1701.02720*, 2017. https://doi.org/10.48550/arXiv.1701.02720

[22] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013. https://doi.org/10.48550/arXiv.1312.6114

[23] V. Kosaraju and A. Sadeghian, "Social-BiGAT: Multimodal trajectory forecasting using Bicycle-GAN and graph attention networks," *arXiv preprint arXiv:1907.03395*, 2019. https://doi.org/10.48550/arXiv.1907.03395

[24] I. Sutskever and O. Vinyals, "Sequence to sequence learning with neural networks," *arXiv preprint arXiv:1409.3215*, 2014. https://doi.org/10.48550/arXiv.1409.3215

[25] L. Gao, Z. Guo, and H. Zhang, "Video captioning with attention based lstm and semantic consistency," *IEEE Transactions on Multimedia*, vol. 19, no. 9, pp. 2045–2055, 2017. https://doi.org/10.1109/TMM.2017.2729019

[26] S. Venugopalan, L. A. Hendricks, R. Mooney, and K. Saenko, "Improving LSTM-based video description with linguistic knowledge mined from text," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1961–1966. https://doi.org/10.48550/arXiv.1604.01729

[27] C. Wang and H. Yang, "Image captioning with deep bidirectional LSTMs," in *Proceedings of the 24th ACM International Conference on Multimedia*, 2016, pp. 988–997. https://doi.org/10.1145/2964284.2964299

[28] A. Graves, "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850201*, 2013. https://doi.org/10.48550/arXiv.1308.0850

[29] N. Lee, W. Choi, and P. Vernaza, "DESIRE: Distant future prediction in dynamic scenes with interacting agents," *arXiv preprint arXiv:1704.04394*, 2017. https://doi.org/10.48550/arXiv.1704.04394

[30] J. Goodfellow *et al.*, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014. https://doi.org/10.48550/arXiv.1406.2661

[31] J. Amirian, J.-B. Hayet, and J. Pettre, "Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs," *arXiv preprint arXiv:1904.09507*, 2019. https://doi.org/10.48550/arXiv.1904.09507

[32] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. D. Reid, H. Rezatofighi, and S. Savarese, "Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 137–146. https://doi.org/10.48550/arXiv.1409.3215

[33] J. Zhu *et al.*, "Toward multimodal image-to-image translation," *arXiv preprint arXiv:1711.11586*, 2017. https://doi.org/10.48550/arXiv.1711.11586

[34] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," *arXiv preprint arXiv:1606.03657*, 2016. https://doi.org/10.48550/arXiv.1606.03657

[35] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," *arXiv preprint arXiv:1406.6247*, 2014. https://doi.org/10.48550/arXiv.1406.6247

[36] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," *arXiv preprint arXiv:1502.03044,* 2015. https://doi.org/10.48550/arXiv.1502.03044

[37] A. Sadeghian, V. Kosaraju, N. Hirose, H. Rezatofighi, and S. Savarese, "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints," *arXiv preprint arXiv:1806.01482*, 2018. https://doi.org/10.48550/arXiv.1806.01482

[38] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention network," *arXiv preprint arXiv:1710.10903*, 2017. https://doi.org/10.48550/arXiv.1710.10903

[39] B. T. Morris and M. M. Trivedi, "Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2287–2301, 2011. https://doi.org/10.1109/TPAMI.2011.64

[40] A. Sadeghian, F. Legros, M. Voisin, R. Vesel, A. Alahi, and S. Savarese, "CAR-Net: Clairvoyant attentive recurrent network," *arXiv preprint arXiv:1711.10061*, 2018. https:-doi.org/10.48550/arXiv.1711.10061

[41] P. V. C. B. C. P. H. S. T. N. Lee, W. Choi and M. Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," *arXiv preprint arXiv:1704.04394*, 2017. https://doi.org/10.48550/arXiv.1704.04394

[42] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *2009 IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 261–268. https://doi.org/10.1109/ICCV.2009.5459260

[43] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 655–664, 2007. https://doi.org/10.1111/j.1467-8659.2007.01089.x

6 AUTHORS

Md. Muktar Ali is with the Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Birulia, Savar, Dhaka-1216, Bangladesh (E-mail: muktar15-2298@diu.edu.bd).

Md. Tahmeed Kowsher Hameem is with the Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Birulia, Savar, Dhaka-1216, Bangladesh (E-mail: tahmeedhameem1@gmail.com).

Deen Islam is with the Department of Computer Science and Engineering, Shanto Mariam University of Creative Technology, Permanent Campus Plot 06, Uttara, Dhaka-1230, Bangladesh (E-mail: deenislam.dataanalyst.cse@gmail.com).

Md. Arafath Hossen Abir is with the Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Birulia, Savar, Dhaka-1216, Bangladesh (E-mail: arafathkhanabir@gmail.com).

Md. Moijeuddin Molla is with the Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Birulia, Savar, Dhaka-1216, Bangladesh (E-mail: 2016moize@gmail.com).

F. M. Javed Mehedi Shamrat is with the Department of Computer System and Technology, Universiti Malaya, Kuala Lumpur 50603, Malaysia (E-mail: javedmehedicom@gmail.com).