HAL open science

TTC-SLSTM: Human trajectory prediction using time-to-collision interaction energy

Huu-Tu Dang, Raphael Korbmacher, Antoine Tordeux, Benoit Gaudou, Nicolas Verstaevel

To cite this version:

Huu-Tu Dang, Raphael Korbmacher, Antoine Tordeux, Benoit Gaudou, Nicolas Verstaevel. TTC-SLSTM: Human trajectory prediction using time-to-collision interaction energy. 15th IEEE International Conference on Knowledge and Systems Engineering (KSE 2023), Academy of Cryptography Techniques (ACT); VNU University of Engineering and Technology (VNU-UET); IEEE, Oct 2023, Hanoi, Vietnam. pp.1-6, 10.1109/KSE59128.2023.10299443 . hal-04251961

HAL Id: hal-04251961

https://hal.science/hal-04251961

Submitted on 20 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TTC-SLSTM: Human trajectory prediction using time-to-collision interaction energy

Huu-Tu Dang
UMR 5505 IRIT
Université Toulouse Capitole
Toulouse, France
huu-tu.dang@ut-capitole.fr

Raphael Korbmacher
Department for traffic safety and reliability
University of Wuppertal
Wuppertal, Germany
korbmacher@uni-wuppertal.de

Antoine Tordeux
Department for traffic safety and reliability
University of Wuppertal
Wuppertal, Germany
tordeux@uni-wuppertal.de

Benoit Gaudou
UMR 5505 IRIT
Université Toulouse Capitole
Toulouse, France
benoit.gaudou@ut-capitole.fr

Nicolas Verstaevel
UMR 5505 IRIT
Université Toulouse Capitole
Toulouse, France
nicolas.verstaevel@ut-capitole.fr

***Abstract*—Recent data-driven approaches using neural networks have shown promising results for pedestrian trajectory prediction. These algorithms outperform the knowledge-based and physics-based models in terms of distance error. However, it has been observed that the neural networks produce too many collisions and pedestrian overlaps, leading to unrealistic predictions. To address this problem, we propose in this contribution a hybrid extension of the Social-LSTM data-driven approach by introducing a collision loss in the training. The collision loss is provided by an interaction energy based on the time-to-collision with the neighbors. The predictions are evaluated and compared to the Social-LSTM model using both distance-error metrics and collision metrics. The results show that collision and pedestrian overlap in the predicted trajectories decreases exponentially as the collision loss weight in the training increases, while the displacement error remains approximately constant. These preliminary results make the proposed hybrid algorithm a promising approach for realistic pedestrian prediction, especially in high-density situations.**

***Index Terms*—Pedestrian trajectory prediction, Social-LSTM, time-to-collision, interaction energy, hybrid approach**

I. Introduction

In an increasingly urbanised world, the modelling of pedestrian dynamics is necessary for mobility planning in towns and cities, or for the design of large infrastructures or the organisation of events with large audiences, among other applications. These issues are of great societal interest, in terms of mobility, safety and environment. The modelling of pedestrian dynamics is by its nature multidisciplinary. The last decade has seen the rapid emergence of alternative approaches solely driven by data to predicting pedestrian dynamics [3], [4]. Numerous recent empirical analyses have shown that deep learning algorithms using displacement-based error metrics and neural networks, more specifically long short-term memory (LSTM) and generative adversarial networks (GAN), can predict pedestrian trajectories more accurately than traditional physical models in terms of prediction displacement error. LSTM networks have emerged as the most widely employed architecture in this area, primarily due to their capability to adeptly handle sequential data, such as trajectories. The first breakthrough came with the publication of the Social-LSTM from Alahi et al. [6]. It presents a pooling layer, called social pooling, that gathers the hidden states of nearby pedestrians. Further studies have extended these initial algorithms by incorporating elements such as scene information [9], attention mechanisms [10], graph neural networks [11], [12], and heterogeneity among pedestrians [13]. Besides the use of LSTM algorithms, researchers applied other supervised learning algorithms such as generative adversarial networks, like the Social-GAN for predicting pedestrian trajectories. Due to the architecture of the Social-GAN, which contains a generator and a discriminator, it allows to predict a distribution of potential future trajectories. An even more recent trend in modelling consists of hybrid extended approaches in which further physical constraints or knowledge on pedestrian behavior is incorporated into the data-based algorithms. Such advanced hybrid approaches not only allow accurate trajectory prediction in terms of displacement error, but can also limit unrealistic collisions and pedestrian overlap, especially in high-density situations, that data-driven algorithms based only on displacement error metrics tend to describe [4], [7].

In this contribution, we propose to extend the Social-LSTM (SLSTM) data-driven approach [6] by introducing a collision loss in the training, to tackle the problem of unrealistic collisions behavior. The collision metric is provided by an interaction energy based on the times to collision (TTC) with the neighbors, successfully introduced to model pedestrian

The authors acknowledge the Franco-German research project MADRAS funded in France by the Agence Nationale de la Recherche (ANR), grant number ANR-20-CE92-0033, and in Germany by the Deutsche Forschungsgemeinschaft (DFG), grant number 446168800.

dynamics a few years ago [8] (hence the name TTC-SLSTM). A factor $\lambda \geq 0$ weights the collision loss in the training. The results show that collision and pedestrian overlap in the predicted trajectories decrease exponentially for increasing λ, while the displacement error remains approximately constant. These preliminary results make the proposed TTC-SLSTM algorithms a promising hybrid approach for realistic pedestrian prediction, especially in high-density situations.

The rest of the contribution is structured as follows. The next section covers the methodological aspects of the study, including definitions of the SLSTM and TTC-SLSTM algorithms and the evaluation metrics. The empirical analysis using real pedestrian trajectories is described in section III, including a comparison of the approaches. Section IV summarises the results and outlines future research directions.

II. Methodology

This section presents the methodology for the proposed TTC-SLSTM for pedestrian trajectory prediction. The problem of pedestrian trajectory prediction is first formulated, followed by the TTC-SLSTM algorithm description and evaluation metrics for assessing the predictions.

A. Problem formulation

The primary objective is to accurately predict the future positions of pedestrians based on their historical positions and the contextual information of their surroundings. This problem can be formulated as a sequence-to-sequence prediction task, in which the input consists of a series of observed pedestrian positions, while the output comprises the predicted future positions. Given a scene with N pedestrians, their observed positions at time instants $1 \leq t \leq T_{obs}$ is represented as $p(t) = \{p_1^t, p_2^t, ..., p_N^t\}$. Here, $p_i^t = (x_i^t, y_i^t)$ denotes the xy-coordinate of the ith pedestrian, and T_{obs} represents the total number of observation time steps. The future trajectories of the pedestrians are predicted within the time interval $T_{obs} + 1 \leq t \leq T_{obs} + T_{pred}$, where T_{pred} is the number of prediction time steps. The predicted position of the i-th pedestrian at time step t is denoted as $\hat{p}_i^t = (\hat{x}_i^t, \hat{y}_i^t)$.

B. Algorithms

The three key components contributing to the TTC-SLSTM are introduced including time-to-collision, interaction energy, and SLSTM.

1) Time-to-Collision: The Time-to-Collision (TTC) metric estimates how long it would take for two pedestrians to collide with each other if they continue to move at their current velocities [8]. Let r_i represent the radius of the pedestrian body i, and its velocity can be represented as $v_i = (v_{x_i}, v_{y_i})$. The relative position and velocity between the pedestrian i and j can be denoted by $p_{ij} = (x_i - x_j, y_i - y_j)$ and $v_{ij} = (v_{x_i} - v_{x_j}, v_{y_i} - v_{y_j})$, respectively. A collision between pedestrian i and pedestrian j occurs if a ray, originating from (x_i, y_i) and extending in the direction of v_{ij}, intersects the circle centered at (x_j, y_j) with a radius of $r_i + r_j$ at some time τ_{ij} in the future. This condition can be mathematically represented as $||p_{ij} + v_{ij}.t||^2 < (r_i + r_j)^2$ where $||.||$ denotes Euclidean norm. Solving this quadratic inequality for t yields τ_{ij} as the smallest positive root:

$$\tau_{ij} = \frac{-p_{ij} \cdot v_{ij} - \sqrt{(p_{ij} \cdot v_{ij})^2 - ||v_{ij}||^2(||p_{ij}||^2 - (r_i + r_j)^2)}}{||v_{ij}||^2} \tag{1}$$

In case of a collision, $\tau_{ij} = 0$, whereas τ_{ij} is assigned to a large positive number when no collision occurs.

2) Interaction energy: The concept of interaction energy in the field of pedestrian dynamics, as described by Karamouzas et al. [8], is a fundamental component in understanding social interactions among pedestrians. The analysis of a large collection of human motion datasets uncovers a simple power-law interaction based on the estimated time to a potential future collision rather than merely depending on the physical distance between pedestrians. The interaction energy between the pedestrian i and j is formulated as follows:

$$E_{ij} = E(\tau_{ij}) = \frac{k}{\tau_{ij}^2} e^{-\tau_{ij}/\tau_0} \tag{2}$$

where k represents a constant that normalizes the interaction energy, and τ_0 signifies the upper limit for the time range of interaction. The short-range (power-law) interaction has been demonstrated to simulate pedestrian interactions across a wide range of situations effectively and can be employed to reproduce many well-known crowd phenomena [8].

3) Social-LSTM: LSTM networks have demonstrated effective performance in sequence learning tasks. One such task, the prediction of pedestrian trajectories, presents the challenges, that the trajectory of a pedestrian can be significantly influenced by the trajectories of surrounding pedestrians. The number of these neighboring influences can fluctuate widely, especially in densely crowded environments [7]. Alahi et al. [6] proposed a solution to that challenge by adding a "social" pooling layer. This added layer accounts for the spatial and temporal effects of neighbors on an individual's movement, reflecting the inherently social nature of human motion. Given the hidden-state dimension D, and the neighborhood size N_o a tensor H_i^t for the ith trajectory is created with the size $N_o \times N_o \times D$ according to equation 3:

$$H_i^t(m, n, :) = \sum_{j \in N_i} 1_{mn}[x_j^t - x_i^t, y_j^t - y_i^t] h_j^{t-1}. \tag{3}$$

In this equation, h_j^{t-1} denotes the hidden state of the SLSTM corresponding to the jth person at time $t - 1$. The term $1_{mn}[x, y]$ serves as an indicator function to ascertain whether a point (x, y) resides within the (m, n) cell of the grid. Lastly, N_i signifies the set of local neighbors influencing person i within a specific distance. The negative log-likelihood loss (NLL) for the i-th trajectory is minimized to learn the parameters of the SLSTM model.

In the original Social-LSTM model, the training loss function calculates the negative log-likelihood loss for the predicted coordinates $(\hat{x}, \hat{y})$ at each time step, taking into account the estimated standard deviation, mean, and correlation coefficient of the bivariate Gaussian distribution. This loss function provides a measure of the discrepancy between the predicted probability distribution of the pedestrian trajectories and the observed data. Minimizing this value during the training process allows the model to generate more accurate predictions that adhere closely to the actual trajectories.

4) TTC-SLSTM: While the loss function of the SLSTM has shown promising results, it is important to recognize potential limitations: trajectories predicted with low loss values can still manifest unrealistic behavior, particularly high collision rates. This paradox arises because the loss function primarily focuses on the overall distance accuracy of the prediction without explicitly considering the physical plausibility of these trajectories. More precisely, the model can generate trajectories that are statistically accurate (i.e. low distance error) but physically implausible (i.e. result in high collision rates) as it does not inherently account for real-world constraints such as the need for pedestrians to avoid collisions.

To address this limitation, the loss function is augmented with an additional term that captures the interaction energy to neighboring agents (L_i for the ith trajectory):

$$\mathrm{L}_i = \mathrm{NLL}_i + \lambda \frac{1}{T_{pred}} \sum_{t=1}^{T_{pred}} \sum_{j \neq i} \tanh(E_{ij}) \tag{4}$$

where λ is the weight. The $\tanh$ function is applied to the interaction energy to normalize excessively large values for small τ. Fig.1 graphically illustrates the function of $\tanh(E(\tau))$, with $k = 1.5, \tau_0 = 3$ which are the calibrated parameters [8].

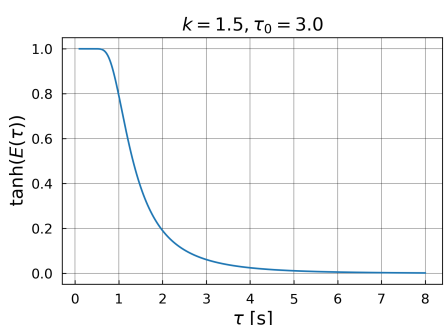


Fig. 1: $\tanh$ function of the interaction energy $E(\tau)$ with $k = 1.5$ and $\tau_0 = 3.0$.

The second term in the new loss function describes the total sum of the $tanh$ function applied to the interaction energy with the neighboring pedestrians. A penalty is added for predicted trajectories with small τ values relative to the neighbors. Consequently, the neural network then learns to minimize both the negative log-likelihood loss and energy term in the proposed training loss function.

C. Evaluation metrics

Two types of evaluation metrics are used including distance-error metrics and collision metrics. The evaluation metrics are typically applied to a *primary pedestrian* in each prediction scene, where the primary pedestrian refers to the specific individual whose future trajectory is the primary focus of the prediction.

The **Average Displacement Error (ADE)** and **Final Displacement Error (FDE)** are commonly used metrics for evaluating the accuracy of pedestrian trajectory predictions [4]. ADE is defined as the average Euclidean distance between the ground truth and prediction of the primary pedestrian over all predicted time steps. In contrast, FDE measures the Euclidean distance between the final predicted position and the corresponding ground truth position of the primary pedestrian. Lower values of ADE and FDE indicate a more accurate predicted trajectory to the ground truth.

$$\begin{aligned} \textbf{ADE} &= \frac{1}{MT_{pred}} \sum_{i=1}^{M} \sum_{t=1}^{T_{pred}} \sqrt{(\hat{x}_i^t - x_i^t)^2 + (\hat{y}_i^t - y_i^t)^2} \\ \textbf{FDE} &= \frac{1}{M} \sum_{i=1}^{M} \sqrt{(\hat{x}_i^{T_{pred}} - x_i^{T_{pred}})^2 + (\hat{y}_i^{T_{pred}} - y_i^{T_{pred}})^2} \end{aligned} \tag{5}$$

where M is the total number of predicted primary pedestrians.

In recent literature, an increasing focus has been placed on the importance of incorporating collision metrics in human trajectory prediction [4], [7]. These metrics offer a deeper understanding of the realism of the predicted trajectories, particularly in high-density scenarios where pedestrians typically adhere to social norms of avoiding collisions. In this study, the **Prediction Collision (Col-I)** and **Groundtruth Collision (Col-II)** [4] are utilized to evaluate collisions in trajectory predictions. The Col-I metric quantifies the proportion of collisions of predicted trajectories between the primary pedestrian and neighbors in the predicted future scene. In contrast, the Col-II metric computes the percentage of collisions between the prediction of the primary pedestrian and the groundtruth of neighboring pedestrians in the future scene.

$$\begin{aligned} \textbf{Col-I} &= \frac{1}{M} \sum_{i=1}^{M} \min(1, \sum_{j \neq i} \sum_{t=1}^{T_{pred}} [\|\hat{p}_i^t, \hat{p}_j^t\| < r_i + r_j]) \\ \textbf{Col-II} &= \frac{1}{M} \sum_{i=1}^{M} \min(1, \sum_{j \neq i} \sum_{t=1}^{T_{pred}} [\|\hat{p}_i^t, p_j^t\| < r_i + r_j]) \end{aligned} \tag{6}$$

where $[\cdot]$ denotes the Iverson bracket, which assigns $[P] = 1$ if the statement P is true and 0 otherwise. $\|\hat{p}_i^t, \hat{p}_j^t\|$ is the Euclidean distance of the predicted position of the primary pedestrian i and the pedestrian j in the corresponding scene at time step t.

It is important to note that the results of Col-I and Col-II metrics are significantly impacted by the selection of the pedestrian radius. Recent studies have utilized a radius of

$0.1\ m$ [4], however, this value may be insufficient to represent a pedestrian adequately. Based on the heuristic of Moussaïd et al. [16] for estimating the size of pedestrians, a pedestrian radius of $0.2\ m$ is selected.

In this contribution, we introduce an additional collision evaluation metric: the **average interaction energy (AE)** in the predicted trajectories, which calculates the average interaction energy in the predicted trajectory of the primary pedestrian i to others in the corresponding scene. Although there are similarities among Col-I, Col-II, and AE metrics, as all three evaluate collisions in the predictions, Col-I and Col-II do not adequately quantify the extent of collisions occurring in the predicted scene. For instance, a predicted scene with one collision and another with ten collisions may have the same Col-I value, whereas the AE metric provides a continuous value that increases with the number of collisions. To prevent extreme values of interaction energy when the TTC is very close to or equal to 0, a small constant $\epsilon = 0.01$ is added into the interaction energy equation:

$$\begin{aligned} \hat{E}(\tau) &= \frac{k}{\tau^2 + \epsilon} e^{-\tau/\tau_0} \\ \mathbf{AE} &= \frac{1}{MT_{pred}} \sum_{i=1}^{M} \sum_{t=1}^{T_{pred}} \sum_{j \neq i} \hat{E}^t(\tau_{ij}) \end{aligned} \tag{7}$$

III. Experiments

This section presents the experiments conducted on the widely used datasets, along with the results obtained using distance-error metrics (**ADE** and **FDE**) as well as collision metrics (**Col-I**, **Col-II**, and **AE**). Additionally, a comparative analysis between the SLSTM and TTC-SLSTM models with varying values of λ is provided.

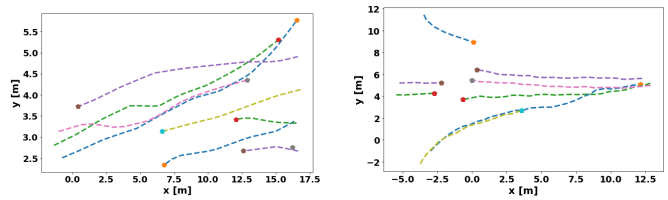
A. Datasets

Two publicly available datasets are used to train the neural network and evaluate its predictions, which include ETH [17] and UCY [18]. These datasets were collected in outdoor experiments and exhibit diverse pedestrian traffic with unidirectional, bidirectional, and multidirectional flows.

The ETH dataset comprises 750 trajectories split between the ETH and HOTEL sub-datasets. The 786 trajectories of the UCY dataset are divided into ZARA01, ZARA02, and UCY. Fig. 2 presents some examples of trajectories in these datasets, where each line corresponds to one trajectory. By averaging instantaneous densities over time, the densities of these sub-datasets were calculated to be 0.13 to 0.38 pedestrians per square meter.

B. Implementation details

The implementation employs the commonly accepted configurations of related contributions [3], with a prediction length of $4.8s$ and an observation length of $3.6s$. Given the datasets' framerate of 0.4 frames per second, this corresponds to using 9 observations to generate 12 predictions. The ADAM optimizer with a learning rate of 0.001 is selected.



(a) ZARA1. (b) ETH.

Fig. 2: Trajectory examples from the datasets.

Training is carried out for 15 epochs with a batch size of 8, using the proposed loss function described previously. For validation and testing, a hold-out validation strategy is adopted by allocating 15% of the dataset for each validation and testing, while the remaining data serves as the training set. All computations are performed using the PyTorch framework [1] on a M1 MacBook Pro with 32 GB of memory.

C. Results

The influence of the interaction energy term on trajectory predictions is investigated by training models with different values of λ $(0.1, 0.25, 0.5, 1.0, 2.0)$ on both the ETH and UCY datasets simultaneously. The results of the TTC-SLSTMs are compared with the original SLSTM model, which corresponds to $\lambda = 0.0$. For each model, the average performance is computed based on a total of 12 trainings, with the confidence interval of 95%. Table I provides the average of the prediction results over the five sub-datasets by using five different evaluation metrics: ADE, FDE, Col-I, Col-II, and AE.

*1) **ADE** and **FDE**:* Fig. 3a,b shows the trend of average ADE and FDE results of the different models. Generally, the ADE and FDE results follow similar patterns across various values of λ. When compared to the baseline SLSTM model, improved ADE and FDE results can be seen for λ values within the range of 0.25 to 1.0. The best performance is achieved at $\lambda = 0.5$, with the ADE of 0.555 and the FDE of 1.135, which correspond to improvements of approximately 0.9% and 1.0%, respectively, over the baseline SLSTM model. However, as λ increases to 2.0, both ADE and FDE significantly rise, indicating that excessively high λ values can negatively impact distance-based accuracy.

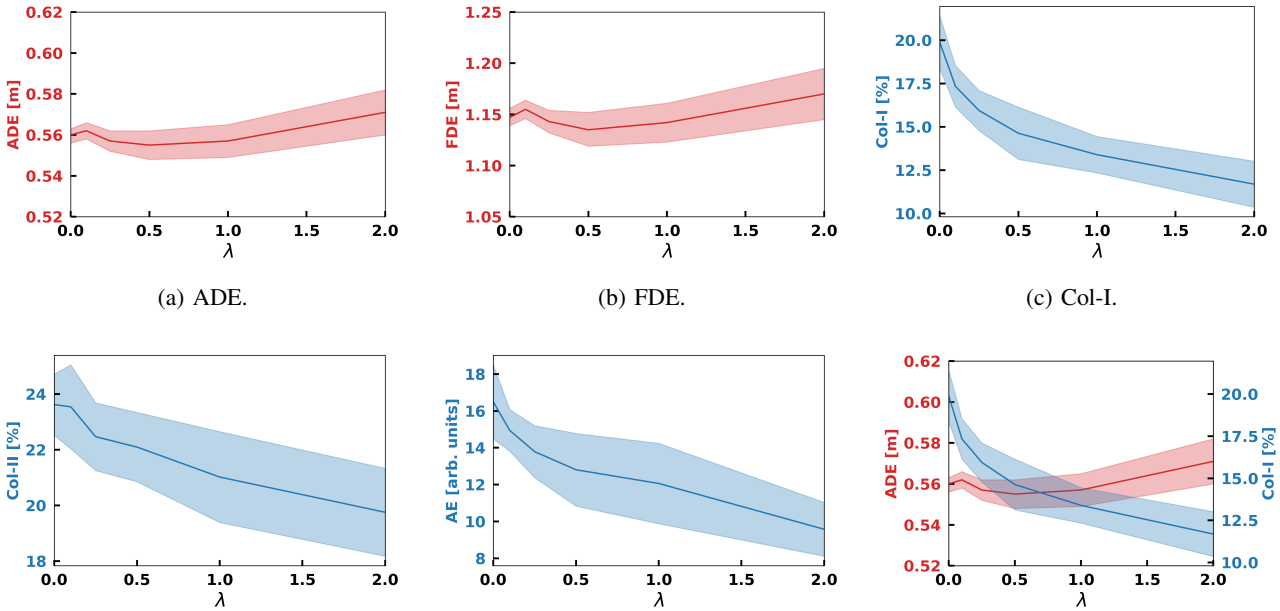
*2) **Col-I**, **Col-II**, and **AE**:* On the other hand, the TTC-SLSTM models overperform the baseline SLSTM model for all three metrics: Col-I, Col-II, and AE, at any value of $\lambda > 0$. The Col-I, Col-II, and AE results decrease consistently when λ increases (as seen in Fig. 3c,d,e) and $\lambda = 2.0$ performs best at almost sub-datasets for Col-I, Col-II, and AE (as shown in Table I).

For both Col-I and AE results, the average values decrease exponentially as increasing λ. With Col-I, the TTC-SLSTM shows an exponential decrease from 19.888 (λ = 0.0) to 11.689 (λ = 2.0), resulting in an approximate 41.2% improvement.

[1] http://pytorch.org

TABLE I: Prediction results of different models.

Metric	Data	SLSTM	TTC-SLSTM				
			$\lambda = 0.1$	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 2.0$
ADE	ETH [17]	0.677	0.681	**0.676**	0.678	0.689	0.706
	HOTEL [17]	0.484	0.484	0.477	0.473	0.473	**0.461**
	ZARA01 [18]	0.500	0.505	0.493	0.486	**0.481**	0.491
	ZARA02 [18]	**0.438**	**0.438**	0.440	0.439	0.439	0.469
	UCY [18]	0.699	0.702	**0.697**	0.698	0.704	0.728
	Average	0.560	0.562	0.557	**0.555**	0.557	0.571
FDE	ETH [17]	1.302	1.309	**1.296**	1.299	1.323	1.371
	HOTEL [17]	0.882	0.890	0.878	0.865	0.875	**0.862**
	ZARA01 [18]	1.049	1.065	1.042	1.017	**1.001**	1.007
	ZARA02 [18]	0.949	0.952	0.956	**0.948**	**0.948**	1.004
	UCY [18]	1.555	1.558	**1.543**	1.548	1.562	1.605
	Average	1.147	1.155	1.143	**1.135**	1.142	1.170
Col-I	ETH [17]	24.568	22.828	21.661	20.218	19.855	**15.505**
	HOTEL [17]	12.145	11.430	10.130	7.379	**4.048**	4.522
	ZARA01 [18]	17.381	9.285	5.972	5.476	4.998	**3.811**
	ZARA02 [18]	17.978	16.822	16.752	16.126	16.281	**14.814**
	UCY [18]	27.367	26.327	25.206	23.957	21.779	**19.791**
	Average	19.888	17.338	15.944	14.631	13.392	**11.689**
Col-II	ETH [17]	28.555	29.999	29.566	28.914	28.914	**26.525**
	HOTEL [17]	17.857	16.668	14.287	13.336	**11.667**	11.668
	ZARA01 [18]	30.237	29.763	28.311	25.713	21.666	**17.143**
	ZARA02 [18]	19.290	**19.289**	19.612	20.060	20.678	19.752
	UCY [18]	22.158	21.970	**20.555**	22.442	22.159	23.673
	Average	23.619	23.538	22.466	22.093	21.017	**19.752**
AE	ETH [17]	20.605	17.509	17.544	13.631	13.081	**9.377**
	HOTEL [17]	11.357	11.168	9.954	9.543	6.035	**3.640**
	ZARA01 [18]	10.164	6.387	3.925	**3.496**	**3.496**	6.502
	ZARA02 [18]	23.186	23.456	23.036	24.387	23.943	**16.307**
	UCY [18]	17.126	16.177	14.492	12.982	13.768	**12.077**
	Average	16.488	14.939	13.790	12.808	12.065	**9.581**



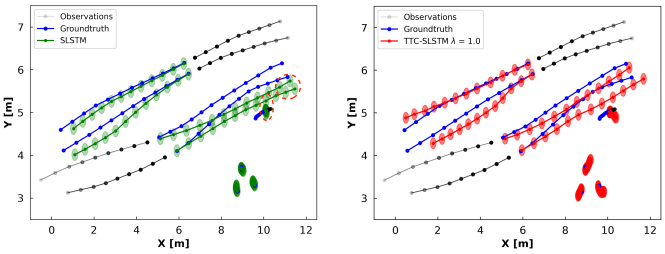
(a) ADE. (b) FDE. (c) Col-I.

(d) Col-II. (e) AE. (f) ADE vs Col-I.

Fig. 3: Average results of evaluation metrics over different values of λ.

Likewise, the AE results exhibit approximately a roughly 41.7% improvements from 16.448 (λ = 0.0) to 9.581 (λ = 2.0). For Col-II, the average results also show a significant improvement of around 16.4% from 23.619 (λ = 0.0) to 19.752 (λ = 2.0). These results suggest that a higher value of λ enhances performance for these collision metrics. However, there is always a trade-off between the optimization of distance-error metrics (ADE, FDE) and collision metrics (Col-I, Col-II, AE). Here, the best value for λ is determined as 1.0, which results in improved ADE and Col-I metrics (see Fig. 3f).

Fig. 4 visually compares the predictions generated by SLSTM and TTC-SLSTM ($\lambda = 1.0$) models. The depicted scene contains eight pedestrians, with four in motion while the others remain stationary. The prediction by the SLSTM model exhibits collisions in the predicted trajectories as highlighted by the red dashed circle in Fig. 4a. In contrast, the TTC-SLSTM successfully addresses this issue by producing collision-free trajectory predictions (see Fig. 4b).



(a) SLSTM. (b) TTC-SLSTM $\lambda = 1.0$.

Fig. 4: Prediction examples of SLSTM and TTC-SLSTM.

IV. Conclusion

Recent studies have increasingly focused on developing neural networks for pedestrian trajectory prediction. However, existing algorithms primarily minimize distance-error loss functions. This contribution proposes a novel training loss function that enhances the neural networks' ability to make more realistic predictions. The proposed loss function incorporates not only the distance-error term but also the collision loss, which is derived from an interaction energy based on time-to-collision with neighboring pedestrians. The proposed approach is evaluated on widely used datasets. The results show that collision and pedestrian overlap in the predicted trajectories decreases exponentially as the collision loss weight in the training increases, while the displacement error remains approximately constant. However, excessively high collision weight values can negatively impact distance-based accuracy despite still improving collision metrics. This demonstrates the ability of the neural networks to make physically plausible trajectory predictions with the proposed loss function. Our future work will explore incorporating the effect of additional behaviors, such as grouping and leader-following, into pedestrian trajectory prediction.

References

[1] Bellomo, N., Piccoli, B. & Tosin, A. Modeling crowd dynamics from a complex system viewpoint. *Mathematical Models And Methods In Applied Sciences*. **22**, 1230004 (2012)

[2] Adrian, J., Seyfried, A. & Sieben, A. Crowds in front of bottlenecks at entrances from the perspective of physics and social psychology. *Journal Of The Royal Society Interface*. **17**, 20190871 (2020)

[3] Korbmacher, R. & Tordeux, A. Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-based Approaches. *IEEE Transactions On Intelligent Transportation Systems*. (2022)

[4] Kothari, P., Kreiss, S. & Alahi, A. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Transactions On Intelligent Transportation Systems*. (2021)

[5] Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S. & Alahi, A. Social gan: Socially acceptable trajectories with generative adversarial networks. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 2255-2264 (2018)

[6] 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).

[7] Korbmacher, R., Dang-Huu, T., Tordeux, A., Verstaevel, N. & Gaudou, B. Differences in pedestrian trajectory predictions for high-and low-density situations. *14th International Conference On Traffic And Granular Flow (TGF) 2022*. (2022)

[8] Karamouzas, I., Skinner, B. & Guy, S. Universal power law governing pedestrian interactions. *Physical Review Letters*. **113**, 238701 (2014)

[9] Xue, H., Huynh, D. & Reynolds, M. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction. *2018 IEEE Winter Conference On Applications Of Computer Vision (WACV)*. pp. 1186-1194 (2018)

[10] Haddad, S., Wu, M., Wei, H. & Lam, S. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model. *ArXiv Preprint ArXiv:1902.05437*. (2019)

[11] Huang, Y., Bi, H., Li, Z., Mao, T. & Wang, Z. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. *Proceedings Of The IEEE/CVF International Conference On Computer Vision*. pp. 6272-6281 (2019)

[12] Monti, A., Bertugli, A., Calderara, S. & Cucchiara, R. Dag-net: Double attentive graph neural network for trajectory forecasting. *2020 25th International Conference On Pattern Recognition (ICPR)*. pp. 2551-2558 (2021)

[13] Lai, W., Xia, Z., Lin, H., Hsu, L., Shuai, H., Jhuo, I. & Cheng, W. Trajectory prediction in heterogeneous environment via attended ecology embedding. *Proceedings Of The 28th ACM International Conference On Multimedia*. pp. 202-210 (2020)

[14] Antonucci, A., Papini, G. P. R., Palopoli, L., & Fontanelli, D. . Generating reliable and efficient predictions of human motion: A promising encounter between physics and neural networks. (2020)

[15] Kothari, P., Sifringer, B., & Alahi, A. Interpretable social anchors for human trajectory forecasting in crowds. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 15556-15566 (2021).

[16] Moussaïd, M., Helbing, D. & Theraulaz, G. How simple rules determine pedestrian behavior and crowd disasters. *Proc. Of The National Academy Of Sciences*. **108**, 6884-6888 (2011)

[17] Pellegrini, S., Ess, A., Schindler, K. & Van Gool, L. You'll never walk alone: Modeling social behavior for multi-target tracking. *2009 IEEE 12th International Conference On Computer Vision*. pp. 261-268 (2009)

[18] Lerner, A., Chrysanthou, Y. & Lischinski, D. Crowds by example. *Computer Graphics Forum*. **26** pp. 655-664 (2007)