Supplementary of: Social-Implicit: Rethinking Trajectory Prediction Evaluation and The Effectiveness of Implicit Maximum Likelihood Estimation

Abduallah Mohamed[1], Deyao Zhu[2], and Warren Vu[1]
Mohamed Elhoseiny[2,*] Christian Claudel[1,*]

[1] The University of Texas at Austin
{abduallah.mohamed, warren.vu, christian.claudel}@utexas.edu
[2] KAUST
{deyao.zhu, mohamed.elhoseiny}@kaust.edu.sa
[3] * Equal advising

A Interactive Demo

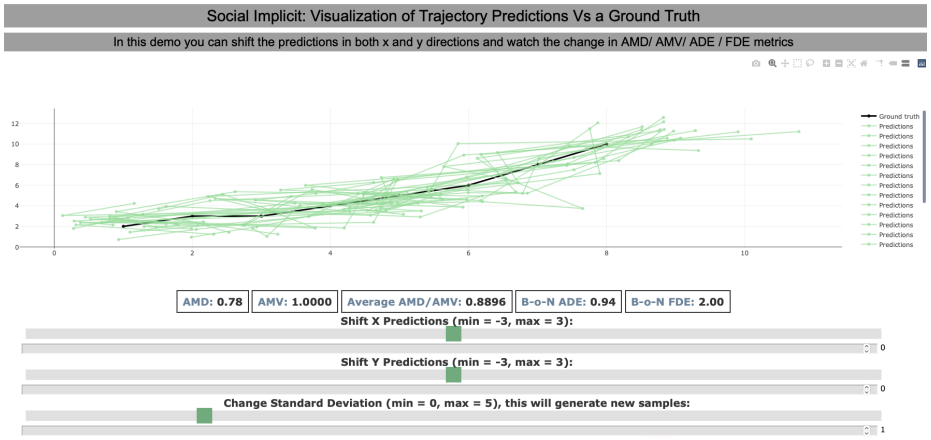

Fig. 1: Social-Implicit Interactive Demo. This demos shows the changes in the metrics in regards of the generated distribution.

We introduce an interactive demo that shows the change of ADE, FDE, AMD and AMV when the generated distribution changes or shifts. The demo URL: https://www.abduallahmohamed.com/social-implicit-amdamv-adefde-demo. By using this demo, one can see the direct effect of changing the distribution and how the ADE/FDE metrics are inadequate to evaluate the predicted quality. For

example, when the shift is huge in one of the x or y directions, the ADE/FDE will stay constant.

B Qualitative Analysis

Figures 4 and 5 show cases where our model performs well or where it might be under-performing. Starting from Figures 4, in the first row and the second row, we see a pedestrian turning left in the past and going straight in the future. S-GAN in the first case and S-GAN, S-STGCNN, ExpertTraj in the second case think confidently that the pedestrian will turn right in the future. But the pedestrian actually goes straightly, which is correctly predicted by our method and Trajectron++. In the third case, S-GAN and S-STGCNN give us a too slow prediction and ExpertTraj gives us a too fast and overturning prediction. In contrast, the predicted distribution of our method and Trajectron++ covers the ground truth well. In the last case, ExpertTraj performs well by placing the predicted concentration on the ground truth. Ours has a wrong prediction following the original trend of the observed motion. In the first row and second row of the second Figure 5, the pedestrian has a sudden turn in the middle of the future. Although all the methods fail to predict this turning, the predicted distribution generated by our method covers the ground truth the best. The third row shows a pedestrian not moving. All the methods give us a close-to-no-movement prediction here. Among them, the movement of ExpertTraj is the smallest. Ours are second-smallest. The last row shows a pedestrian going straight but switching the lane in the middle of the future. Our method and Trajectron++ cover the ground truth trajectory well, while S-STGCNN misses the new lane and ExpertTraj generates a no-existing turn. Overall, though ADE/FDE metrics were stating that Trajectron++ and ExpertTraj are state of art methods, we showed several cases that show the density away from the ground truth. Thus, the ADE/FDE gives an inadequate sense of models' accuracy, unlike the AMD/AMV metrics, which quantifies the whole generated distribution. This correlates with the results in Table[1] and our analysis of the experiments section where our model was performing the best on the ADM/AMV metrics. We also show multi-agent interaction in Fig 3. ExpertTraj is over-confident missing the ground-truth, S-STGCNN have wide variance with collision, Trajectron++ have ground-truth close to predicted distribution tail, while ours have the right balance.

C Evaluation of Deterministic Models

We trained Social-STGCNN [5] as a deterministic model on the ETH/UCY datasets. Instead of predicting a Gaussian distribution, it predicts the trajectory directory. The training used MSE as a loss function. We wanted to test two assumptions for evaluating a deterministic model. The first one is to train it multiple times and use ensemble to find the mean and variance per predicted trajectory. The other one adds up on the previous one by calculating the mean

and variance but fits a GMM and then samples from this GMM. In this experiment, we trained Social-STGCNN 3 times using different random seeds. Table 1 shows the results. The AMD and AMV of the first setting was reported. The KDE was not because there is no method to compute it from a mean and variance without sampling, unlike our metric AMD which has this ability by directly plugging in the mean and variance into the Mahalanobis distance equation. For the second setting, AMD, AMV and KDE were reported as we fitted the samples into a GMM fit then we sampled multiple samples. We only used 3 ensembles of Social-STGCNN to simulate a real-life situation, aka it is not feasible to train it 1000 times and create an ensemble out of it. We notice in Table 1 that the second settings exhibit a very large AMD and KDE, this is an indicator that the GMM fit did not converge because we only have 3 samples. Usually, we use 1000 samples to guarantee the GMM converges and thus the second settings is not feasible to be used as we need that many samples to fit the GMM model well. We notice in both first and second settings that the AMV values are the same. This was expected, as the AMV metric is an indicator of the spread. For the first setting, the AMD value seems reasonable for a deterministic model, as the work of [4] showed that most of motion predictions problem can be solved using a linear Kalman filter. This also supported by the enormous values of the AMV metric as a deterministic model does not have that much of a spread. We connect this with the results in the main paper on the ExpertTraj where the AMV values was on the same order of magnitude as the deterministic model we trained. In other terms, the ExpertTraj indeed behaves as a deterministic model because of the tight spread. We can notice this in some of the visual cases reported in Figures 4 and 5. For further analysis, we plot some samples generated from the ensemble of the deterministic model alongside the spread in Figure 2. We notice sometimes the spread of the predictions might be close to the ground truth as in the sample in the top left corner. Also, it can be completely off, as in the other samples. So we think that using an ensemble of a few versions of a deterministic model is a good approach to evaluate its performance using the AMD/AMV metric. Also, with the AMV metric as a target to optimize for, one can train the ensemble using methods that help encourage diversity [3].

Dataset	Ensemble			GMM Fit			General	
	AMD	AMV	KDE	AMD	AMV	KDE	ADE	FDE
eth	1.45	35.7	-	27.89	35.7	12.89	1.71	2.97
hotel	0.36	19.6	-	44.26	19.5	11.27	1.41	2.56
univ	0.62	170.1	-	24.62	169.9	13.30	1.17	2.13
zara1	1.18	28.0	-	28.95	28.0	13.86	1.71	3.18
zara2	1.03	96.9	-	12.54	96.8	8.36	1.16	2.10
Average	0.93	70.06	-	27.65	70.0	11.94	1.43	2.59

Table 1: Deterministic case experiment. We trained Social-STGCNN [5] as a deterministic model using different random seeds. The first setting reports the AMD/AMV using the mean and variance of the ensemble. The second setting reports the AMD/AMV/KDE using a GMM fit on the mean and variance of the ensemble. The ADE/FDE are the average through the ensembles.

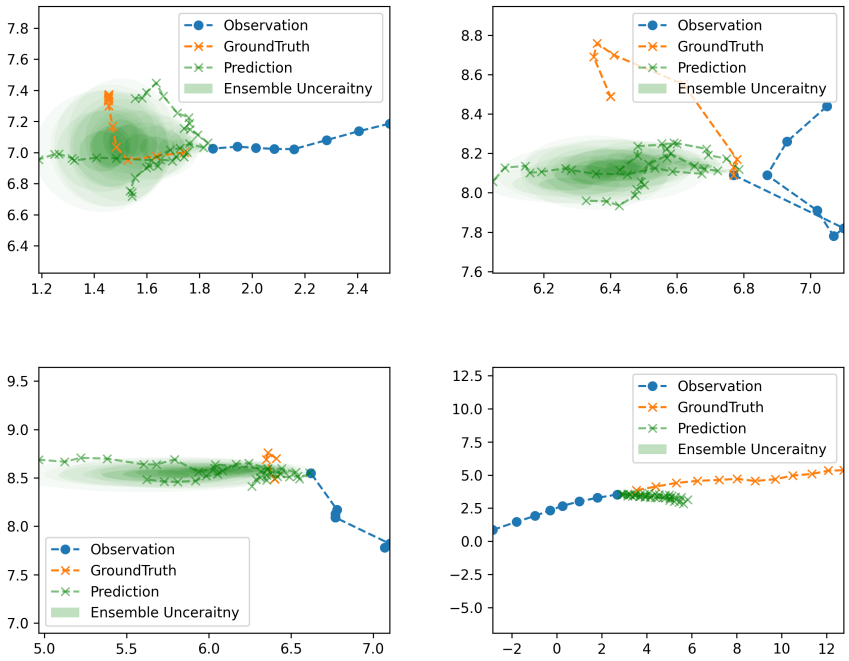

Fig. 2: Social-STGCNN deterministic version predictions.

D Evaluation on Stanford Drone Dataset (SDD)

Here, we test our model and metric on SDD [7] given our limited time. We follow the setting of a SOTA model DAG-Net [6]. Experimental results in Tab 2 show that our model outperforms DAG-Net.

E Social-Implicit Implementation Details

Social-Implicit comprises four zones, as discussed before. Table 3 shows more details about the zones. We notice each zone uses a different configuration of the random noise used to generate the samples. The slow zones use noise which has much lower variance than the faster zones. We also show the layer details of the Social-Cell in Table 4. Both local and global streams share the same design, except that the local stream uses Conv1D and the global stream uses Conv2D. We initialize the noise, global and local weights to zero. The noise weight is being multiplied by the sample generated from the random distribution and then added to the input tensor. The models were trained for 50 epochs with a learning rate

	ADE	FDE	AMD	AMV	KDE
STGAT [2]	0.58	1.11	-	-	-
Social-Ways [1]	0.62	1.16	-	-	-
DAG-Net [6]	0.53	1.04	3.17	0.247	1.76
Social-Implicit (ours)	**0.47**	**0.89**	**2.83**	**0.077**	3.89

Table 2: Results on SDD dataset.

= 1, then the learning rate drops to 0.1 after 45 epochs. The batch size was set to 128. We used SGD as an optimizer. We also used an augmentation technique for the trajectories similar to [8] to fight some imbalance in the datasets. We used random rotation by several degrees, reverse the trajectory, flip the x,y locations, jitter the location by a small value, increase the number of the nodes in the scene by combining it with another scene and changing the speed of the pedestrians. Implementation of the model and augmentation is available in the attached code.

Zone	Speed range	Noise
1	0-0.01 m/s	$\mathcal{N}(0, 0.05^2)$, if eth $\mathcal{N}(0, 0.175^2)$
2	0.01-0.1 m/s	$\mathcal{N}(0, 1^2)$ if eth $\mathcal{N}(1.5^2)$
3	0.1-1.2 m/s	$\mathcal{N}(0, 4^2)$
4	1.2 ms -	$\mathcal{N}(0, 8^2)$

Table 3: Social-Zones configurations. The speed range determines if an observed trajectory will be withing the zone or not. The random noise exhibits different variances depending on the zone.

Section	Layer Name	Configuration
Local Stream	Spatial CNN	Conv1D[P,P,3,1]
	Spatial Activation	ReLU
	Spatial ResCNN	Conv1D[P,P,1,0]
	Temporal CNN	Conv1D[T_o,T_p,3,1]
	Temproal ResCNN	Conv1D[T_o,T_p,1,0]
Global Stream	Noise Weight	1 Parameter
	Spatial CNN	Conv2D[P,P,3,1]
	Spatial Activation	ReLU
	Spatial ResCNN	Conv2D[P,P,1,0]
	Temporal CNN	Conv2D[T_o,T_p,3,1]
	Temproal ResCNN	Conv2D[T_o,T_p,1,0]
	Global Weight	1 Parameter
	Local Weight	1 Parameter

Table 4: Social-Cell configuration. A Conv1D or Conv2D with [x,x,x,x] = [input features, output features, kernel size, padding size]. The Res = Residual connection being added to the previous layer output. P is the dimension of the observed location. T_o and T_p is the number of observed and predicted time steps.

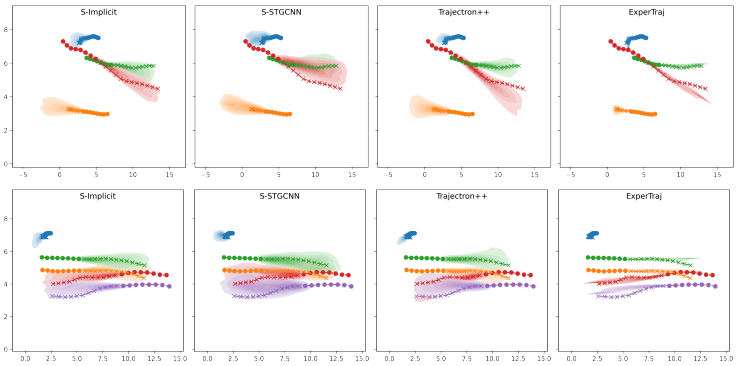

Fig. 3: Multi-pedestrian interaction cases on the ETH/UCY datasets.

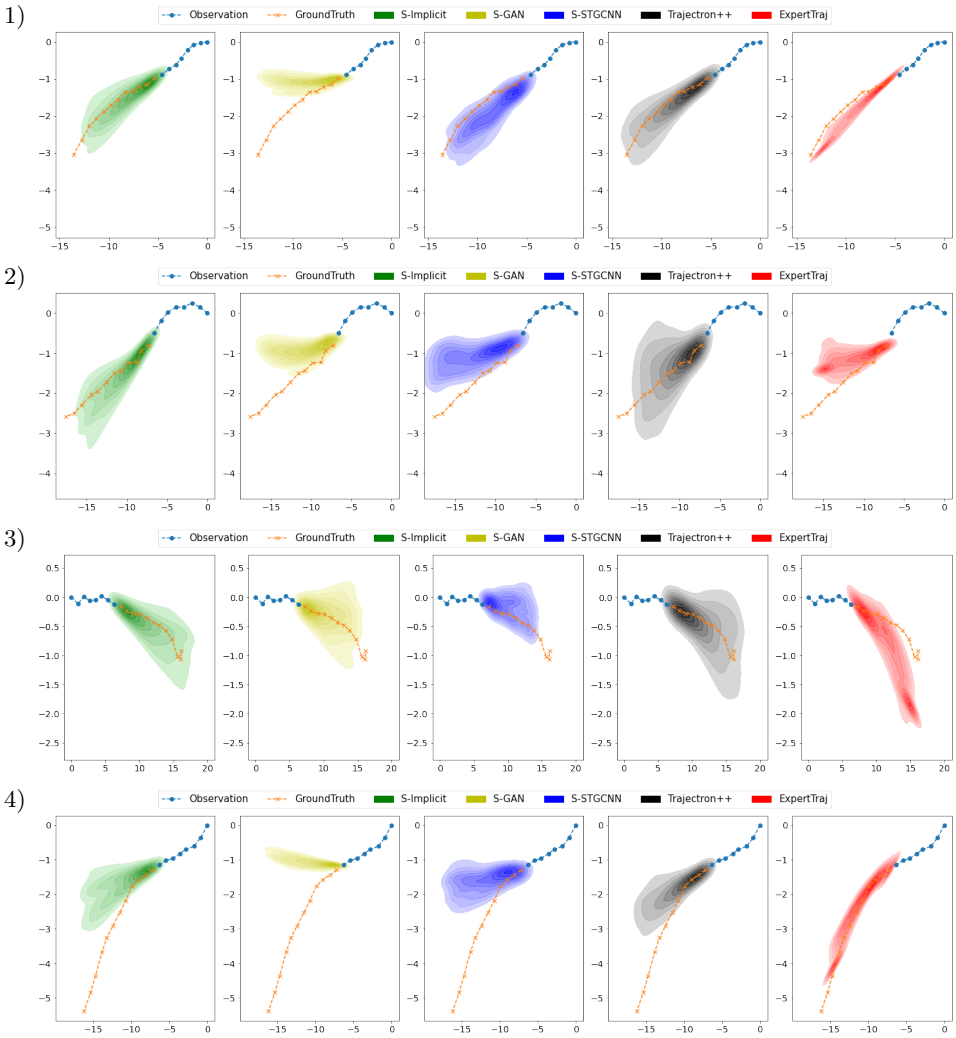

Fig. 4: Visualization of the predicted trajectories by several models on the ETH/UCY datasets.

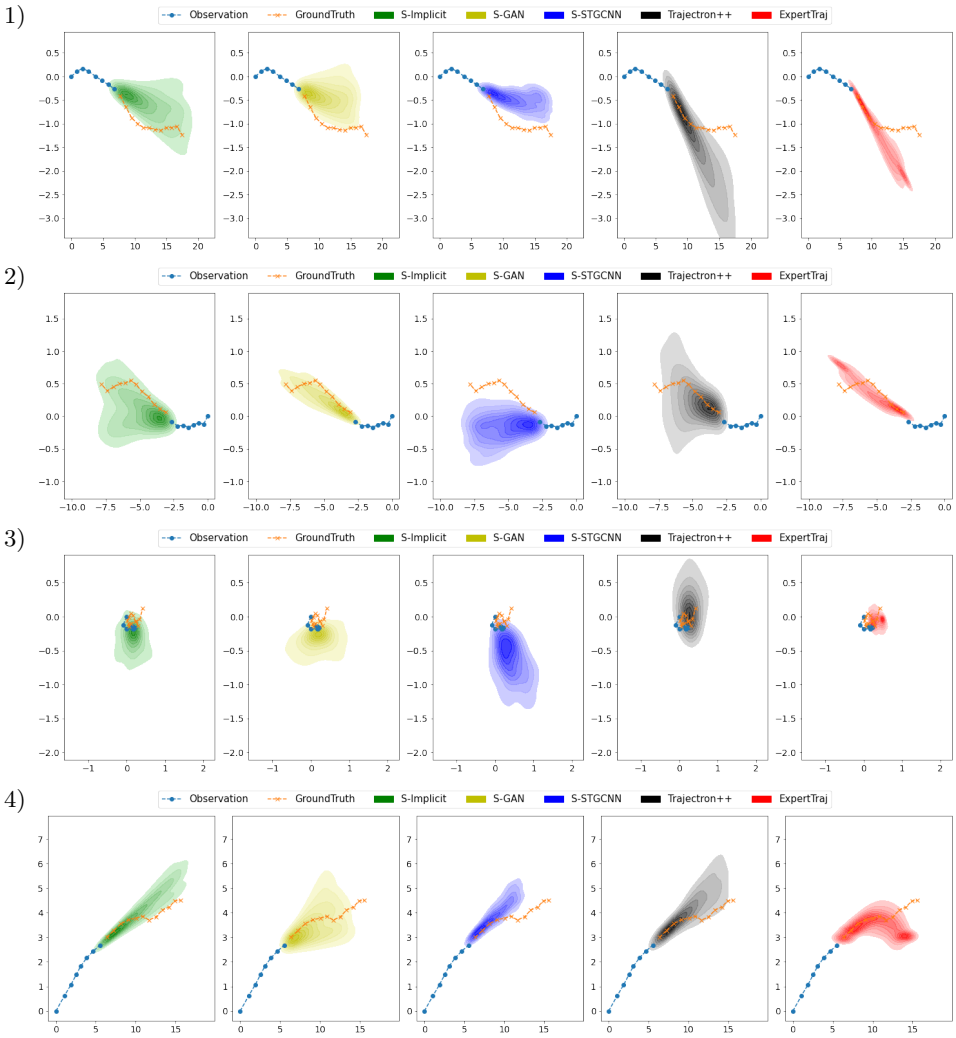

Fig. 5: Visualization of the predicted trajectories by several models on the ETH/UCY datasets.

References

1. Amirian, J., Hayet, J.B., Pettré, J.: Social ways: Learning multi-modal distributions of pedestrian trajectories with gans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019) 5
2. Huang, Y., Bi, H., Li, Z., Mao, T., Wang, Z.: Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6272–6281 (2019) 5
3. Liu, L., Wei, W., Chow, K.H., Loper, M., Gursoy, E., Truex, S., Wu, Y.: Deep neural network ensembles against deception: Ensemble diversity, accuracy and robustness. In: 2019 IEEE 16th international conference on mobile ad hoc and sensor systems (MASS). pp. 274–282. IEEE (2019) 3
4. Makansi, O., Cicek, Ö., Marrakchi, Y., Brox, T.: On exposing the challenging long tail in future prediction of traffic actors. arXiv preprint arXiv:2103.12474 (2021) 3
5. Mohamed, A., Qian, K., Elhoseiny, M., Claudel, C.: Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14424–14432 (2020) 2, 3
6. Monti, A., Bertugli, A., Calderara, S., Cucchiara, R.: Dag-net: Double attentive graph neural network for trajectory forecasting. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 2551–2558. IEEE (2021) 4, 5
7. Robicquet, A., Sadeghian, A., Alahi, A., Savarese, S.: Learning social etiquette: Human trajectory understanding in crowded scenes. In: European conference on computer vision. pp. 549–565. Springer (2016) 4
8. 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) 5