AG-GAN: An Attentive Group-Aware GAN for pedestrian trajectory prediction

Yue Song, Niccoló Bisagno, Syed Zohaib Hassan and Nicola Conci[1]

[1]University of Trento, Department of Information Engineering and Computer Science - DISI

Email: yue.song@studenti.unitn.it, niccolo.bisagno@unitn.it, syedzohaib.hassan@unitn.it, nicola.conci@unitn.it

Overview

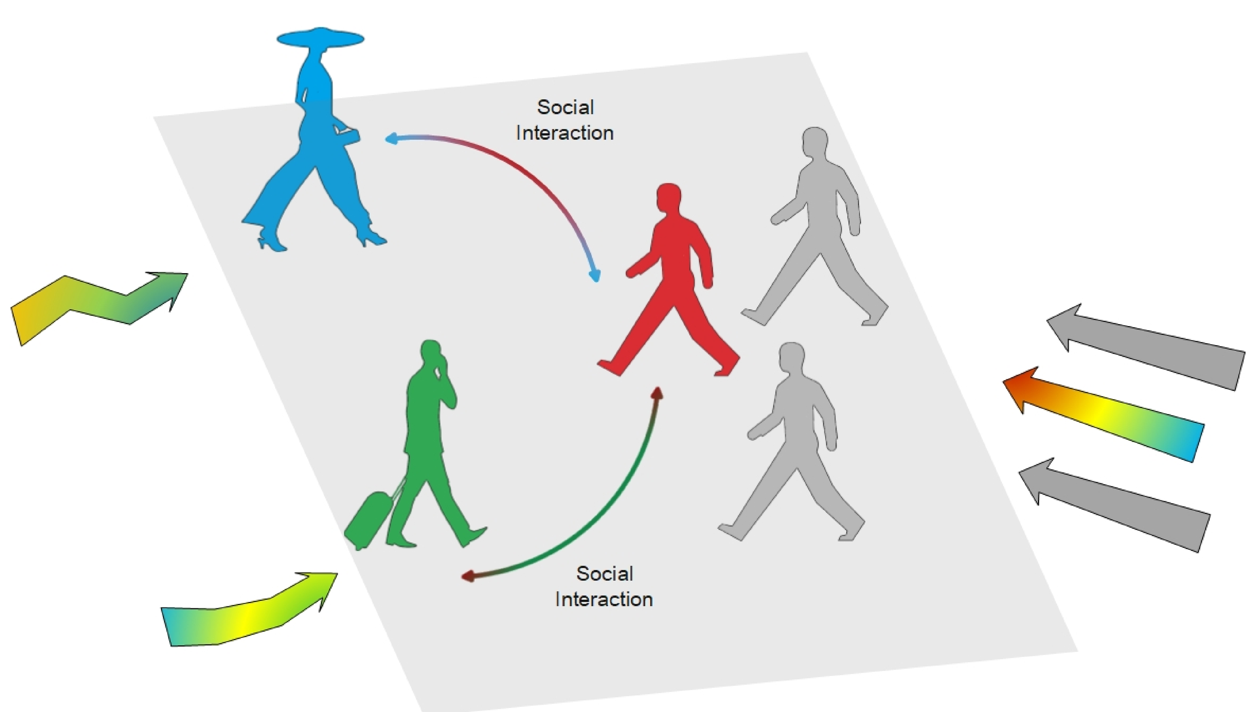

Figure 1:AG-GAN considers only social interactions from agents who are not moving coherently with the pedestrian of interest; an attention mechanism is used to exploit the past motion history and the interaction within the neighborhood.

Novelties

Understanding human behaviors in crowded scenarios requires analyzing not only the position of the subjects in space, but also the scene context. In our approach we address the following issues:

- **Interactions between socially-related pedestrians**. We exploit human-to-human interactions and social relationship to improve the trajectory prediction task.
- **Attentive exploitation of past trajectories**. We exploit the relevance of each portion of the past trajectories, like the presence of sharp turns, using an attention-based module.
- **Metrics for predicted paths.** We exploit parameters like trajectory similarity or collision rate for evaluation.

Method

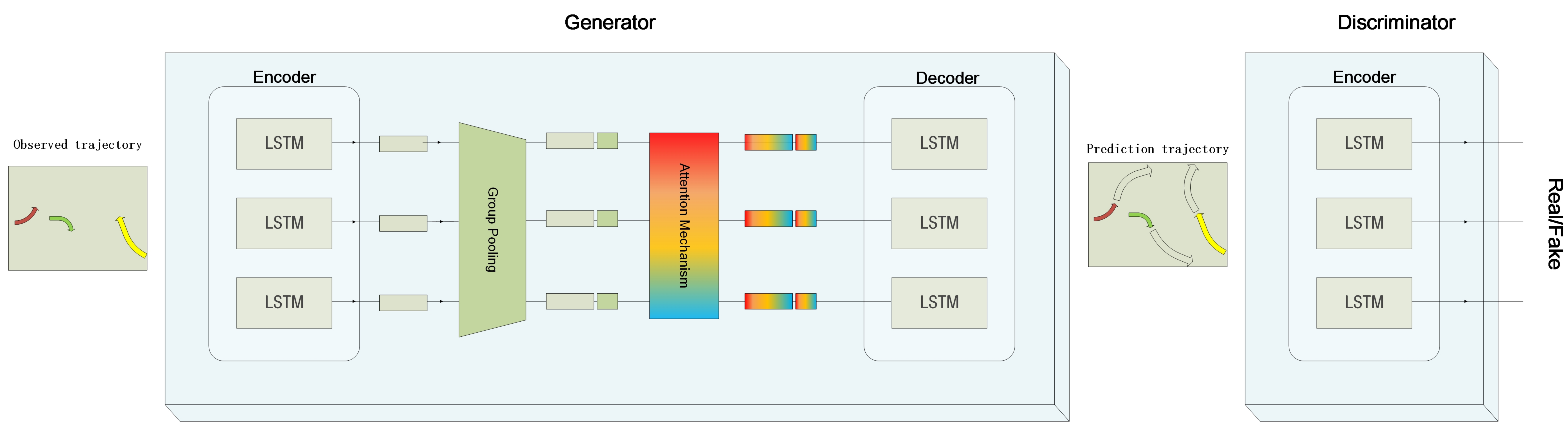

Figure 2:The architecture of the proposed AG-GAN. The Generator G takes past trajectories and encodes them into hidden states. The hidden states are pooled by a group pooling module with the information of selected neighborhood interaction. The attention mechanism helps the decoder focusing on relevant segments of the trajectories for future path generation. The Discriminator D is fed with both prediction path and ground truth.

Attention mechanism

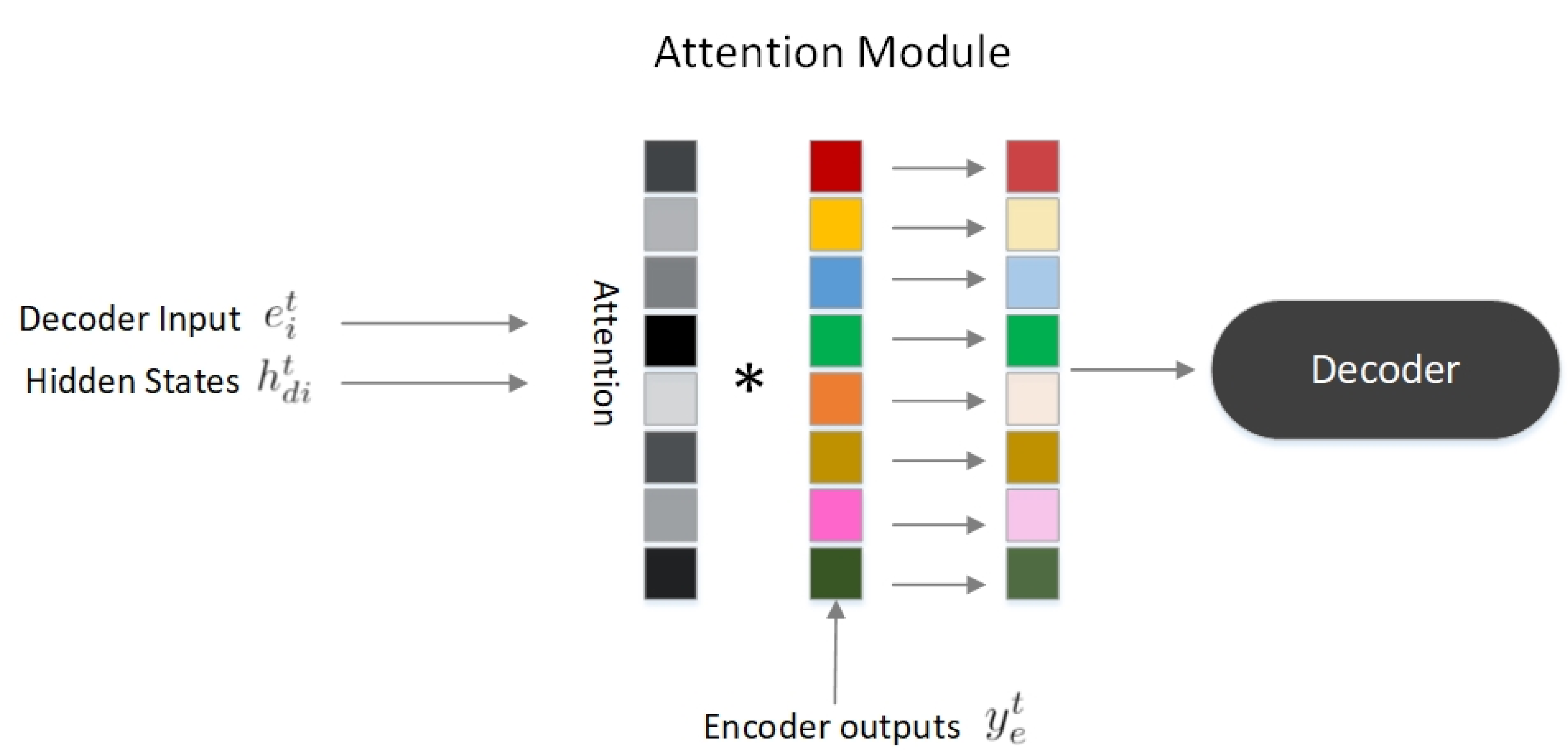

Figure 3:The Attention mechanism allows to focus on the most representative segments of past trajectories (e.g. sharp turns).

Group pooling

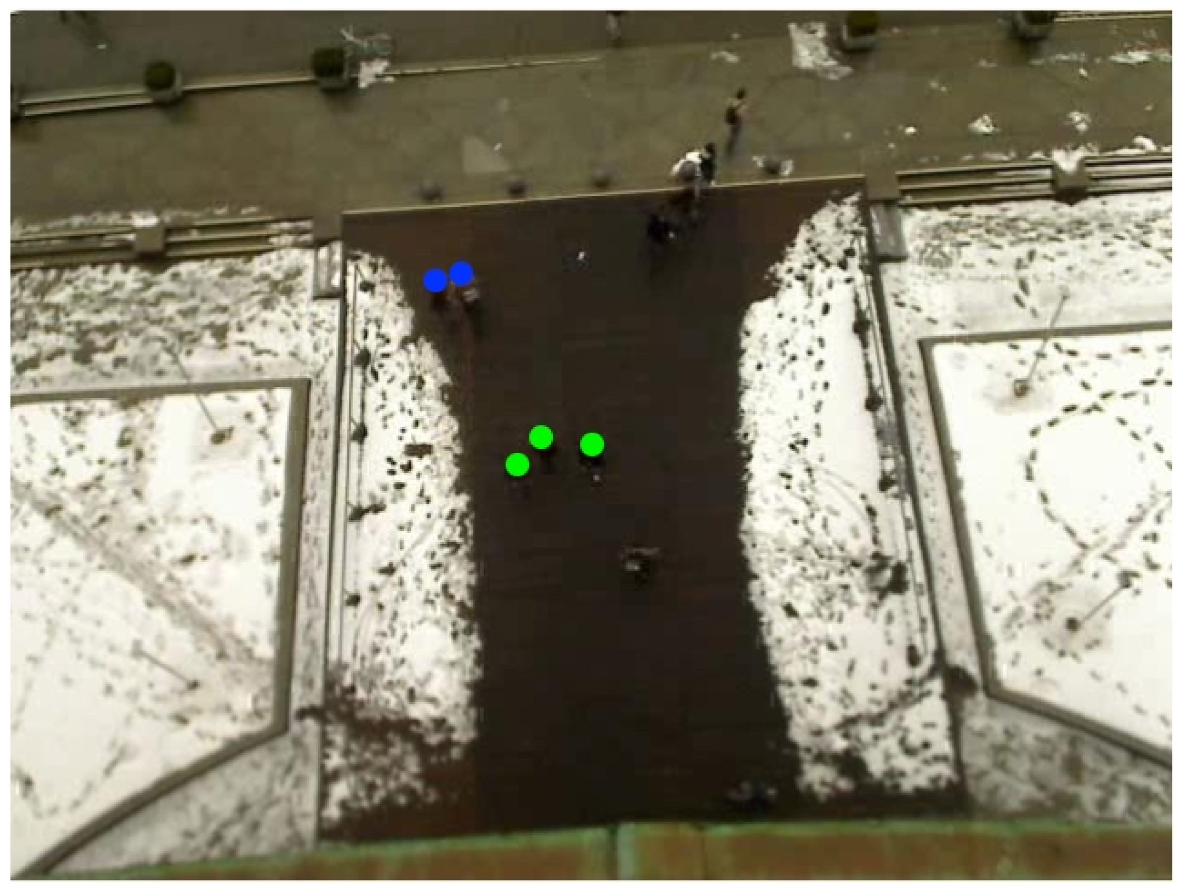

Figure 4:Pedestrians walking coherently are clustered into one group s_i. In this frame, two sets of pedestrians going in opposite directions are identified.

AG-GAN: An Attentive Group-Aware GAN for pedestrian trajectory prediction

Yue Song, Niccoló Bisagno, Syed Zohaib Hassan and Nicola Conci[1]

[1]University of Trento, Department of Information Engineering and Computer Science - DISI
Email: yue.song@studenti.unitn.it, niccolo.bisagno@unitn.it, syedzohaib.hassan@unitn.it, nicola.conci@unitn.it

Evaluation Metrics

- *Average Displacement Error (ADE)*
- *Final Displacement Error (FDE)*
- *Dynamic Time Warping (DTW)*
- *Collision rate.* We evaluate the effectiveness of interaction modeling by calculating the collision rate per frame.

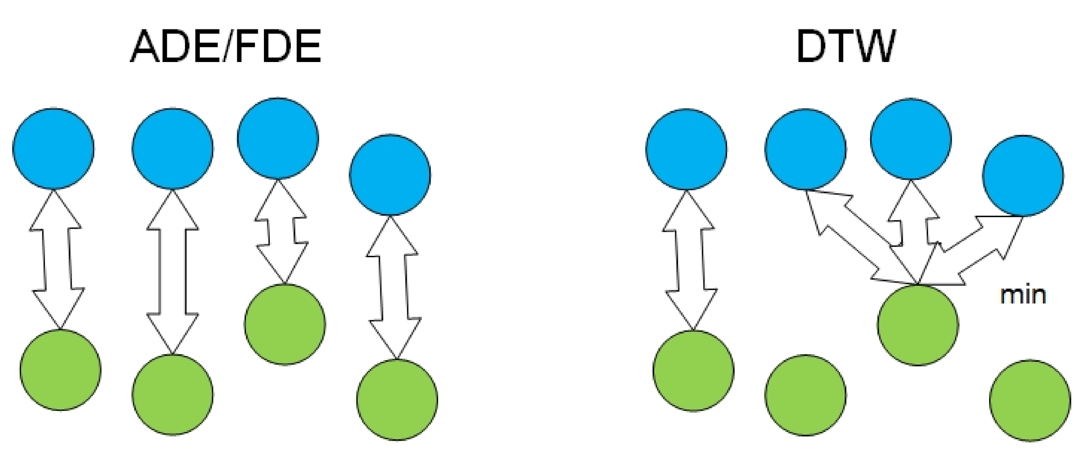

Figure 5:Using ADE/FDE vs DTW in comparing two trajectories. ADE/FDE takes two locations at the same timestamp while DTW compares two timestamps with minimum distance.

References

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese.
Social lstm: Human trajectory prediction in crowded spaces.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016.

[2] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi.
Social gan: Socially acceptable trajectories with generative adversarial networks.
In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[3] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, and Silvio Savarese.
Sophie: An attentive gan for predicting paths compliant to social and physical constraints.
arXiv preprint arXiv:1806.01482, 2018.

Quantitative Results

Table 1:Comparison of our model with other baselines. ADE/FDE are reported in meters.

	Other baselines				AG-GAN		
Datasets	**Lin**	**S-LSTM [1]**	**S-GAN [2]**	**Sophie [3]**	**G**	**A**	**G+A**
ETH	1.33/2.94	1.09/2.35	0.81/1.52	0.70/1.43	0.72/1.28	0.67/1.27	**0.67/1.21**
HOTEL	0.39/0.72	0.79/1.76	0.72/1.61	0.76/1.67	0.43/0.87	0.47/0.95	**0.33/0.67**
UCY	0.82/1.59	0.67/1.40	0.60/1.26	**0.54/1.24**	0.63/1.30	0.62/1.31	0.61/1.28
ZARA1	0.62/1.21	0.47/1.00	0.34/0.69	**0.30/0.64**	0.35/0.70	0.35/0.72	0.36/0.73
ZARA2	0.77/1.48	0.56/1.17	0.42/0.84	0.38/0.78	0.32/0.65	0.37/0.76	**0.31/0.64**
AVG	0.79/1.59	0.72/1.54	0.59/1.18	0.54/1.15	0.49/0.96	0.50/1.00	**0.46/0.90**

Table 2:The displacement error of S-GAN [2] and our AG-GAN in the setting of single generation.

ADE/FDE	S-GAN [2]	S-GAN-P [2]	AG-GAN
ETH	1.11/2.20	1.11/2.24	**1.01/2.14**
HOTEL	0.78/1.70	0.74/1.57	**0.62/1.40**
UNIV	0.77/1.69	1.07/2.18	**0.74/1.68**
ZARA1	**0.64/1.40**	0.68/1.28	0.66/1.44
ZARA2	0.55/1.20	0.57/1.21	**0.53/1.16**
AVG	0.77/1.64	0.83/1.70	**0.71/1.56**

Table 3:For each model, the mean DTW is calculated between the prediction S_{pred}^{N} and ground truth S_{gt}^{N}. The lower the value, the higher the similarity of the generated trajectories.

mean DTW	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
LIN	2.08	0.76	5.80	1.60	1.93	2.43
S-LSTM [1]	1.72	1.33	2.78	0.72	0.94	1.50
S-GAN [2]	1.53	1.08	**2.71**	**0.63**	0.84	1.36
AG-GAN	**1.43**	**0.73**	2.82	0.67	**0.78**	**1.29**

Table 4:Comparison of the average collision rate % of the prediction trajectories S_{pred}^{N} per frame for each model.

%	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
LIN	0	**0.06**	4.57	0.15	1.09	1.17
S-LSTM [1]	0	0.18	7.21	**0.11**	0.52	1.60
S-GAN [2]	0	0.22	7.40	**0.11**	0.46	1.64
AG-GAN	0	0.13	**4.45**	0.14	**0.40**	**1.02**

Qualitative Results

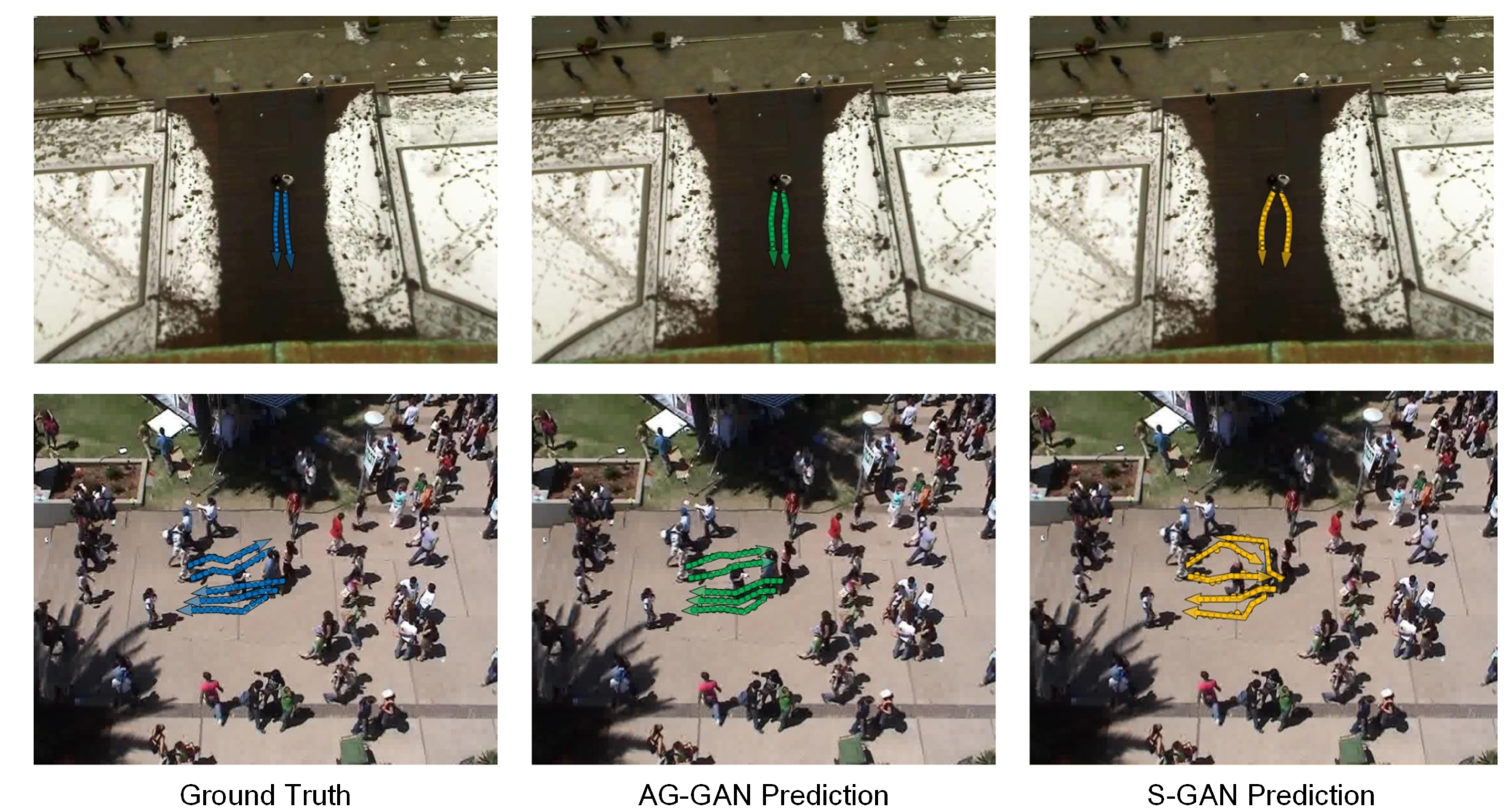

Figure 6:Example of two groups walking in opposite direction. Our model preserves the coherent group motion, therefore displaying a behavior closer to ground truth.

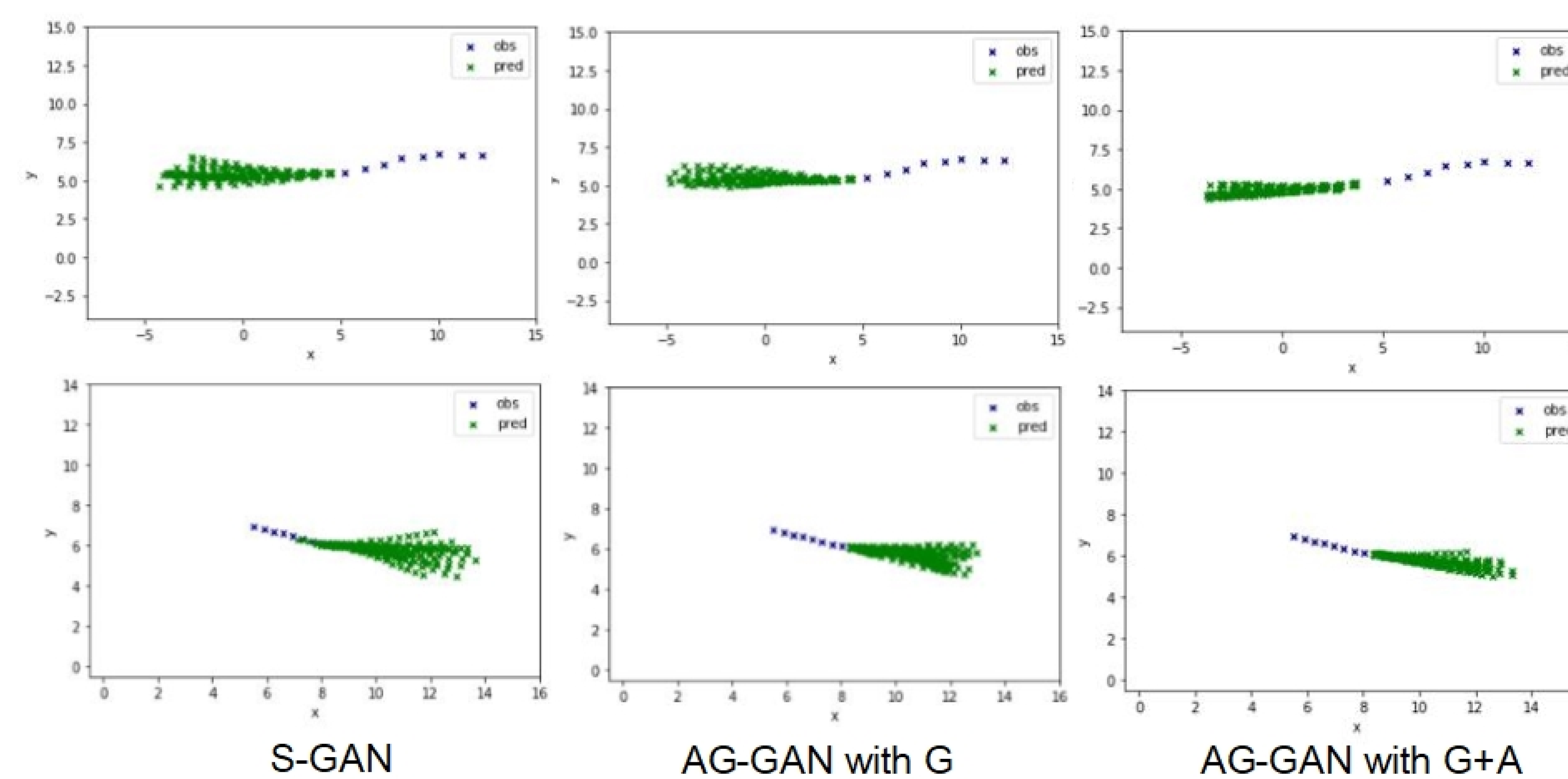

Figure 7:20 generated trajectories for a pedestrian in ETH dataset (top row) and ZARA2 dataset (bottom row). The blue crosses represent the ground truth and the green crosses represent the predicted paths. Thanks to the use of Group-Aware pooling and attention mechanism, generation convergence is improved, getting closer to the ground truth.