Social GAN : Socially Acceptable Trajectories with Generative Adversarial Networks

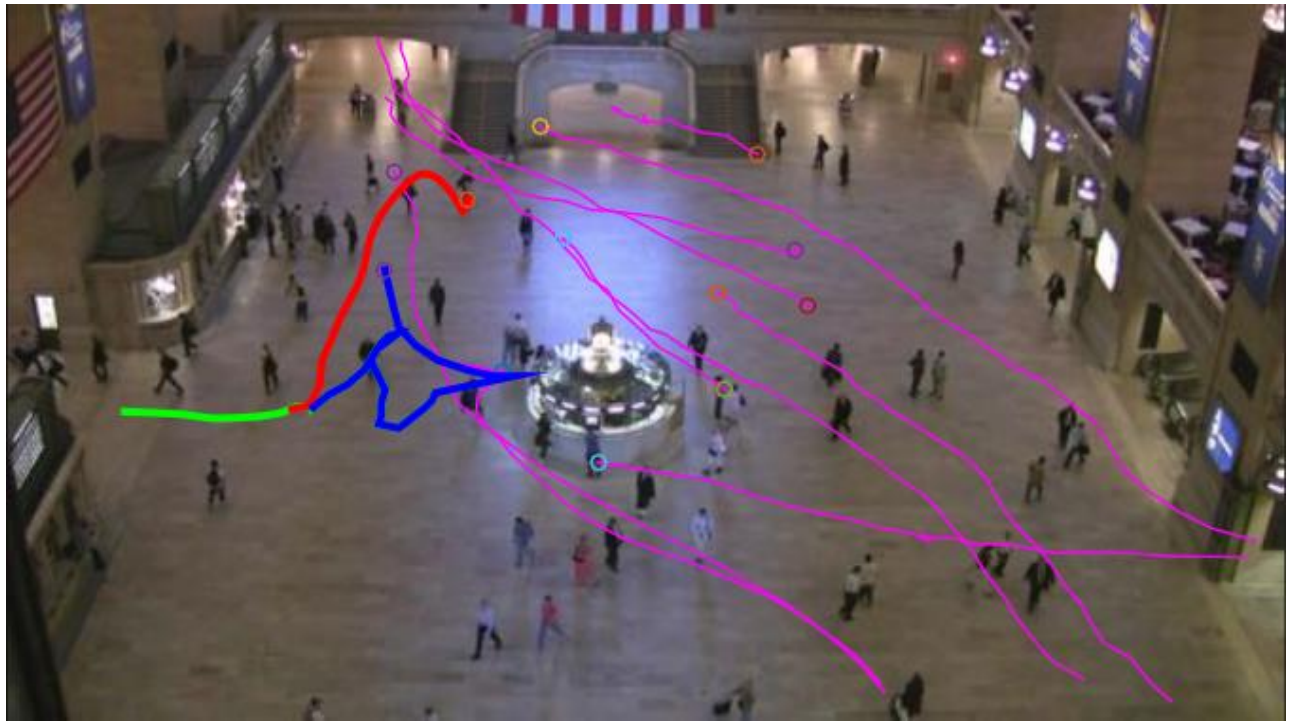
CVPR2018

2019.04.11

발표자 박성현

1 Introduction

Human Trajectory Prediction



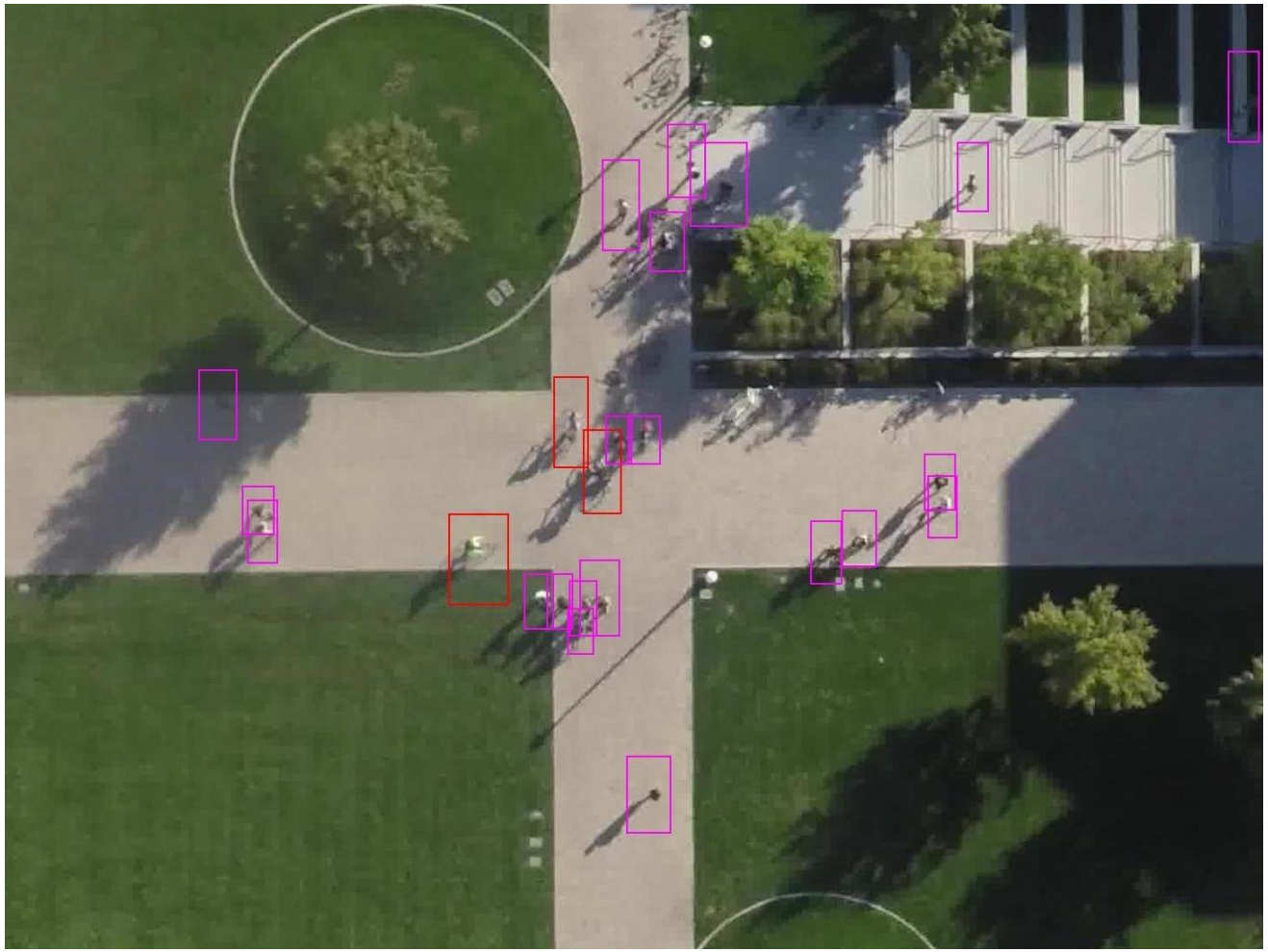
[사람의 경로를 예측하는 Task]

1. **Interpersonal** – 사람들끼리 서로의 경로에 영향
2. **Socially Acceptable**
3. **Multimodal** –여러 경로가 가능

1 Introduction

Human Trajectory Dataset

보통 Bird View의 이미지 + 좌표



Human Trajectory Dataset : http://trajnet.stanford.edu/

2 Model

Social GAN

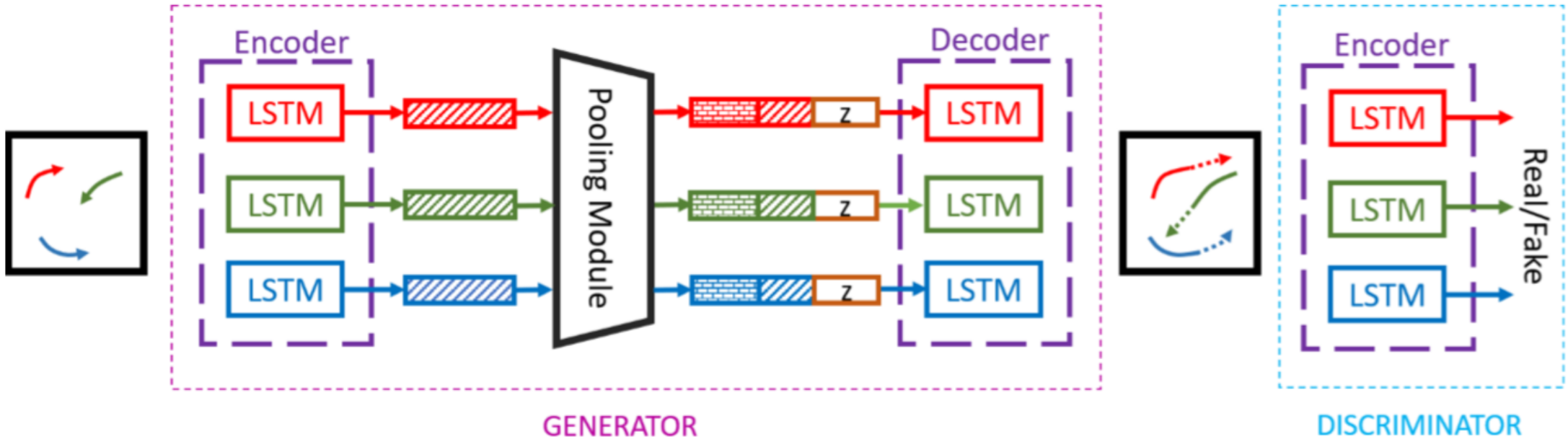


Figure 2: System overview. Our model consists of three key components: Generator (G), Pooling Module, and Discriminator (D). G takes as input past trajectories X_i and encodes the history of the person i as H_i^t. The pooling module takes as input all $H_i^{t_{obs}}$ and outputs a pooled vector P_i for each person. The decoder generates the future trajectory conditioned on $H_i^{t_{obs}}$ and P_i. D takes as input T_{real} or T_{fake} and classifies them as socially acceptable or not (see Figure 3 for PM).

2 Model

Social GAN

[Generator (Encoder)]

$$e_i^t = \phi(x_i^t, y_i^t; W_{ee})$$
$$h_{ei}^t = LSTM(h_{ei}^{t-1}, e_i^t; W_{encoder})$$
$$c_i^t = \gamma(P_i, h_{ei}^t; W_c)$$
$$h_{di}^t = [c_i^t, z]$$

[Generator (Decoder)]

$$e_i^t = \phi(x_i^{t-1}, y_i^{t-1}; W_{ed})$$
$$P_i = PM(h_{d1}^{t-1}, ..., h_{dn}^t)$$
$$h_{di}^t = LSTM(\gamma(P_i, h_{di}^{t-1}), e_i^t; W_{decoder})$$
$$(\hat{x}_i^t, \hat{y}_i^t) = \gamma(h_{di}^t)$$

[Variety Loss]

$$\mathcal{L}_{variety} = \min_k \|Y_i - \hat{Y}_i^{(k)}\|_2$$

2 Model

Social GAN

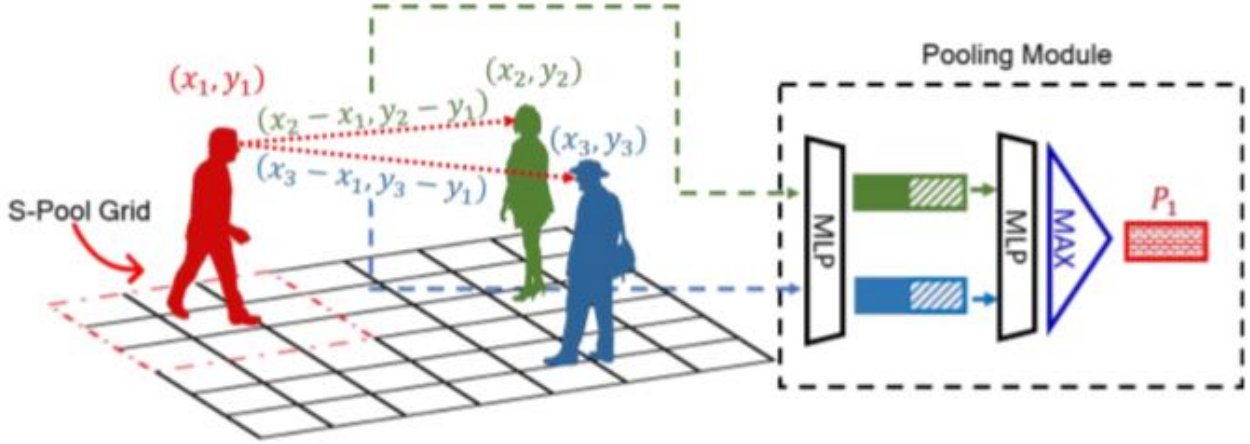


Figure 3: Comparison between our pooling mechanism (red dotted arrows) and Social Pooling [1] (red dashed grid) for the red person. Our method computes relative positions between the red and all other people; these positions are concatenated with each person's hidden state, processed independently by an MLP, then pooled elementwise to compute red person's pooling vector P_1. Social pooling only considers people inside the grid, and cannot model interactions between all pairs of people.

3 Experiments

Quantitative Evaluation

[Evaluation Metrics]

1. *Average Displacement Error (ADE)*: Average $L2$ distance between ground truth and our prediction over all predicted time steps.
2. *Final Displacement Error (FDE)*: The distance between the predicted final destination and the true final destination at end of the prediction period T_{pred}.

[Baseline]

1. *Linear*: A linear regressor that estimates linear parameters by minimizing the least square error.
2. *LSTM*: A simple LSTM with no pooling mechanism.
3. *S-LSTM*: The method proposed by Alahi *et al.* [1]. Each person is modeled via an LSTM with the hidden states being pooled at each time step using the social pooling layer.

3 Experiments

Quantitative Evaluation

Metric	Dataset	Linear	LSTM	S-LSTM [1]	SGAN (Ours) 1V-1	SGAN (Ours) 1V-20	SGAN (Ours) 20V-20	SGAN (Ours) 20VP-20
ADE	**ETH**	0.84 / 1.33	0.70 / 1.09	0.73 / 1.09	0.79 / 1.13	0.75 / 1.03	0.61 / **0.81**	**0.60** / 0.87
	HOTEL	**0.35 / 0.39**	0.55 / 0.86	0.49 / 0.79	0.71 / 1.01	0.63 / 0.90	0.48 / 0.72	0.52 / 0.67
	UNIV	0.56 / 0.82	0.36 / 0.61	0.41 / 0.67	0.37 / 0.60	0.36 / 0.58	**0.36 / 0.60**	0.44 / 0.76
	ZARA1	0.41 / 0.62	0.25 / 0.41	0.27 / 0.47	0.25 / 0.42	0.23 / 0.38	**0.21 / 0.34**	0.22 / 0.35
	ZARA2	0.53 / 0.77	0.31 / 0.52	0.33 / 0.56	0.32 / 0.52	0.29 / 0.47	**0.27** / 0.42	0.29 / **0.42**
AVG		0.54 / 0.79	0.43 / 0.70	0.45 / 0.72	0.49 / 0.74	0.45 / 0.67	**0.39 / 0.58**	0.41 / 0.61
FDE	**ETH**	1.60 / 2.94	1.45 / 2.41	1.48 / 2.35	1.61 / 2.21	1.52 / 2.02	1.22 / **1.52**	**1.19** / 1.62
	HOTEL	**0.60 / 0.72**	1.17 / 1.91	1.01 / 1.76	1.44 / 2.18	1.32 / 1.97	0.95 / 1.61	1.02 / 1.37
	UNIV	1.01 / 1.59	0.77 / 1.31	0.84 / 1.40	0.75 / 1.28	**0.73 / 1.22**	0.75 / 1.26	0.84 / 1.52
	ZARA1	0.74 / 1.21	0.53 / 0.88	0.56 / 1.00	0.53 / 0.91	0.48 / 0.84	**0.42** / 0.69	0.43 / **0.68**
	ZARA2	0.95 / 1.48	0.65 / 1.11	0.70 / 1.17	0.66 / 1.11	0.61 / 1.01	**0.54 / 0.84**	0.58 / 0.84
AVG		0.98 / 1.59	0.91 / 1.52	0.91 / 1.54	1.00 / 1.54	0.93 / 1.41	**0.78 / 1.18**	0.81 / 1.21

Table 1: Quantitative results of all methods across datasets. We report two error metrics Average Displacement Error (ADE) and Final Displacement Error (FDE) for $t_{pred} = 8$ and $t_{pred} = 12$ (8 / 12) in meters. Our method consistently outperforms state-of-the-art S-LSTM method and is especially good for long term predictions (lower is better).

3 Experiments

Quantitative Evaluation

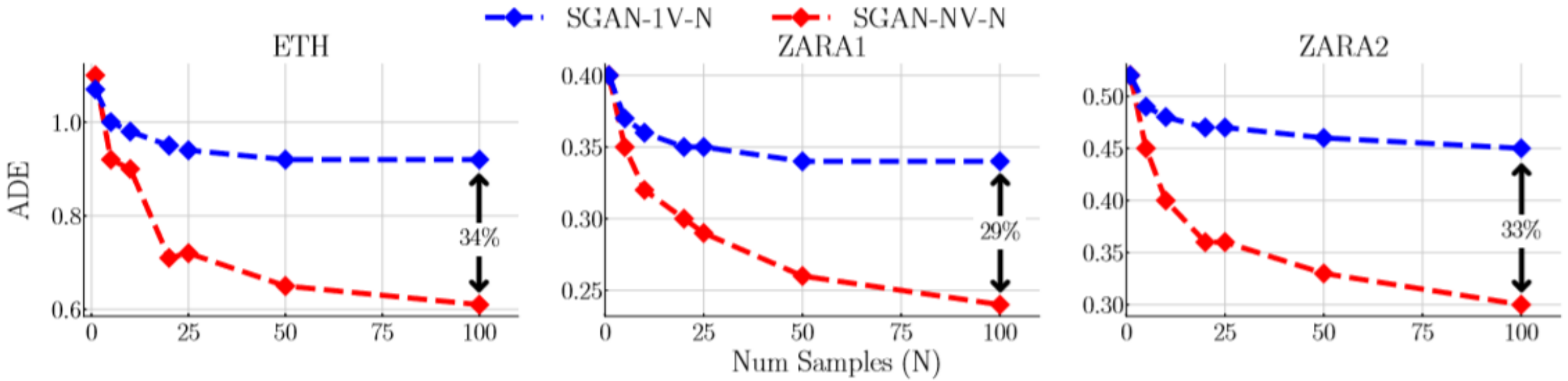


Figure 4: Effect of variety loss. For SGAN-1V-N we train a single model, drawing one sample for each sequence during training and N samples during testing. For SGAN-NV-N we train several models with our variety loss, using N samples during both training and testing. Training with the variety loss significantly improves accuracy.

3 Experiments

Quantitative Evaluation

	LSTM	S-LSTM	SGAN	SGAN-P
8	0.02	1.79	0.04	0.12
12	0.03	2.61	0.05	0.15
Speed-Up	82x	1x	**49x**	**16x**

Table 2: Speed (in seconds) comparison with S-LSTM. We get 16x speedup as compared to S-LSTM allowing us to draw 16 samples in the same time S-LSTM makes a single prediction. Unlike S-LSTM we don't perform pooling at each time step resulting in significant speed bump without suffering on accuracy. All methods are benchmarked on Tesla P100 GPU

3 Experiments

Quantitative Evaluation

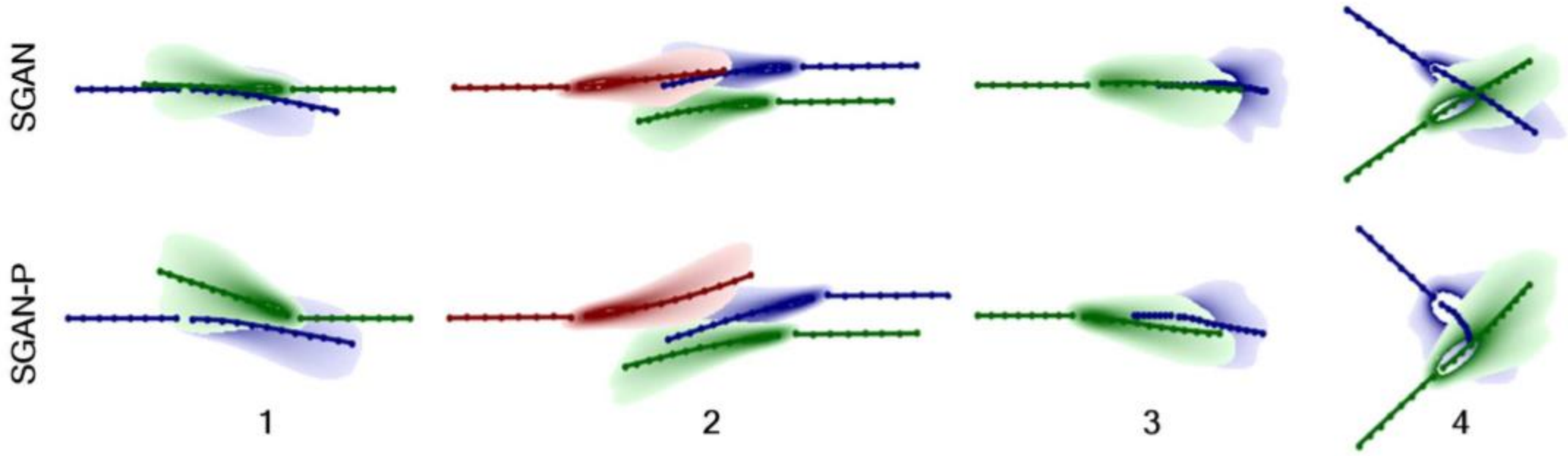


Figure 5: Comparison between our model without pooling (SGAN, top) and with pooling (SGAN-P, bottom) in four collision avoidance scenarios: two people meeting (1), one person meeting a group (2), one person behind another (3), and two people meeting at an angle (4). For each example we draw 300 samples from the model and visualize their density and mean. Due to pooling, SGAN-P predicts socially acceptable trajectories which avoid collisions.

3 Experiments

Quantitative Evaluation

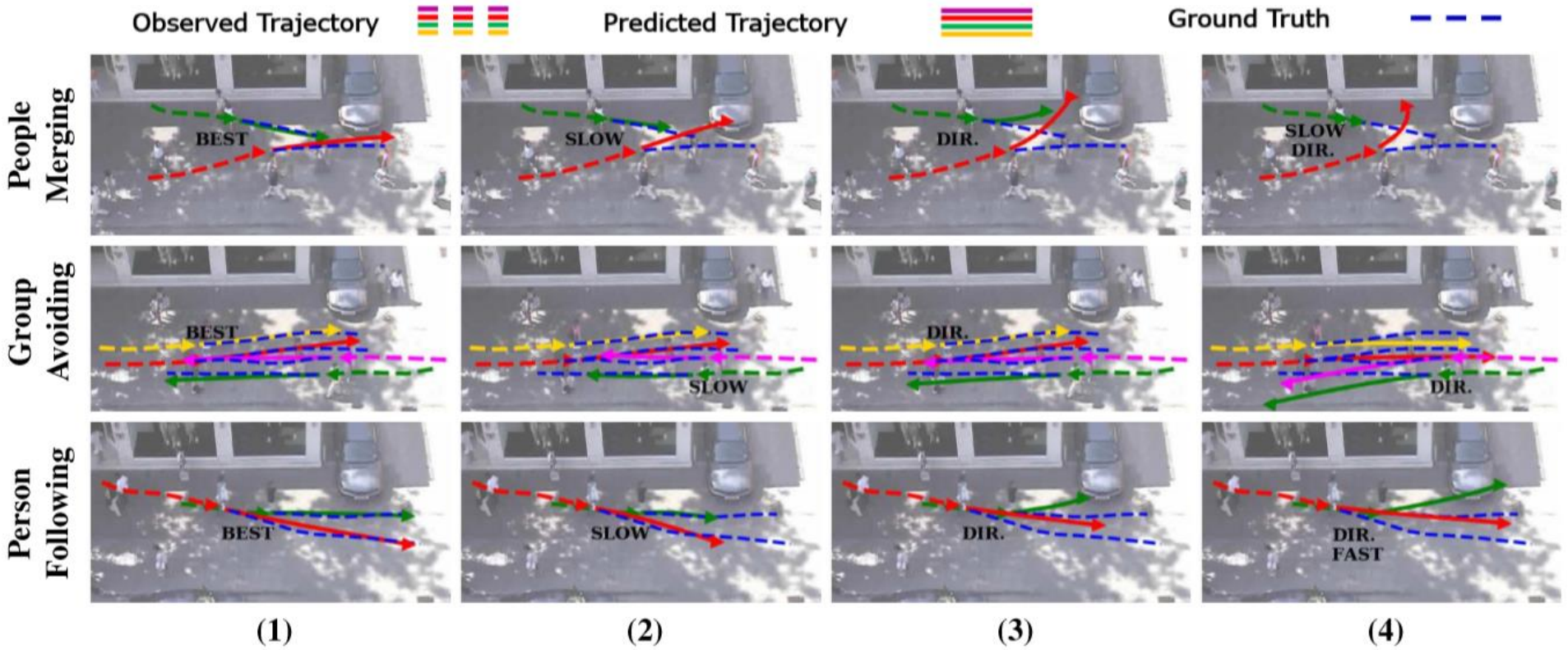


Figure 6: Examples of diverse predictions from our model. Each row shows a different set of observed trajectories; columns show four different samples from our model for each scenario which demonstrate different types of socially acceptable behavior. BEST is the sample closest to the ground-truth; in SLOW and FAST samples, people change speed to avoid collision; in DIR samples people change direction to avoid each other. Our model learns these different avoidance strategies in a data-driven manner, and jointly predicts globally consistent and socially acceptable trajectories for all people in the scene. We also show some failure cases in supplementary material.