Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks

Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, Alexandre Alahi
Stanford University, EPFL
CVPR 2018 Poster
Problem

- Given the spatial coordinates of each pedestrian from frame $1$ to $t$, trajectory prediction aims at predicting the coordinates in the future time period from $t + 1$ to $t + T$. 
Previous Method

Inherent Properties of Human Motion

• **Interpersonal.** The motion of each person depends on the people around them.

• **Socially Acceptable.** Some trajectories are physically possible but socially unacceptable. (Right-of-way or respecting personal space)

• **Multimodal.** Given a partial history, there is no single correct future prediction.
Drawbacks of Existing Methods

• **No capacity to model interactions between all people.** Existing methods model a local neighborhood around each person when making the prediction.

• **“Average Behavior”**. Existing methods commonly use the loss function that minimize the euclidean distance between the ground truth and forecasted outputs.
Pipeline

Initialization: $h_{di}^t = [MLP(P_i, h_{ei}^t), z]$

• Input trajectories: $X = \{X_1, X_2, ..., X_n\}, X_i = (x_i^t, y_i^t)$.
• Predicted future trajectories: $\hat{Y} = \{\hat{Y}_1, \hat{Y}_2, ..., \hat{Y}_n\}, \hat{Y}_i = (x_i^t, y_i^t)$.
• Latent variable $z$, randomly sampled from $\mathcal{N}(0,1)$. 

Xu Gao, Peking University, gaoxu1024@pku.edu.cn

2018/12/4
Pooling Module
Reference: PointNet

Loss Function

- Adversarial loss: $\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p(z)}[\log(1 - D(G(z)))]$
- Variety loss: $\mathcal{L}_{\text{variety}} = \min_k \| Y_i - \hat{Y}_i^{(k)} \|_2$
- For each scene, generate $k$ possible output predictions by randomly sampling $z$, and choose the best prediction in $L2$. 

Xu Gao, Peking University, gaoxu1024@pku.edu.cn
Experiments

• **Datasets:** ETH & UCY.

• **Metrics:**
  • ADE: The average L2 distance between the ground truth and the prediction.
  • FDE: The distance between the predicted final destination and the true final destination at the end of the prediction.

• **Baselines:**
  • Linear: A linear regressor to estimate by minimizing the least square error.
  • LSTM: A simple LSTM with no pooling mechanism
Experiments

• **Name of this method:** SGAN-kVP-N
• k is the parameter in the variety loss.
• N is the number of time which this paper sample from the model.
• P means the usage of the pooling module.

**Strategy:**
• Observe 8 timesteps and predict 8/12 timesteps.
## Experiments

<table>
<thead>
<tr>
<th>Metric</th>
<th>Dataset</th>
<th>Linear</th>
<th>LSTM</th>
<th>S-LSTM [1]</th>
<th>SGAN (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1V-1</td>
<td>1V-20</td>
<td>20V-20</td>
<td>20VP-20</td>
</tr>
<tr>
<td>ADE</td>
<td>ETH</td>
<td>0.84 / 1.33</td>
<td>0.70 / 1.09</td>
<td>0.73 / 1.09</td>
<td>0.79 / 1.13</td>
</tr>
<tr>
<td></td>
<td>HOTEL</td>
<td>0.35 / 0.39</td>
<td>0.55 / 0.86</td>
<td>0.49 / 0.79</td>
<td>0.71 / 1.01</td>
</tr>
<tr>
<td></td>
<td>UNIV</td>
<td>0.56 / 0.82</td>
<td>0.36 / 0.61</td>
<td>0.41 / 0.67</td>
<td>0.37 / 0.60</td>
</tr>
<tr>
<td></td>
<td>ZARA1</td>
<td>0.41 / 0.62</td>
<td>0.25 / 0.41</td>
<td>0.27 / 0.47</td>
<td>0.25 / 0.42</td>
</tr>
<tr>
<td></td>
<td>ZARA2</td>
<td>0.53 / 0.77</td>
<td>0.31 / 0.52</td>
<td>0.33 / 0.56</td>
<td>0.32 / 0.52</td>
</tr>
<tr>
<td>AVG</td>
<td></td>
<td>0.54 / 0.79</td>
<td>0.43 / 0.70</td>
<td>0.45 / 0.72</td>
<td>0.49 / 0.74</td>
</tr>
<tr>
<td>FDE</td>
<td>ETH</td>
<td>1.60 / 2.94</td>
<td>1.45 / 2.41</td>
<td>1.48 / 2.35</td>
<td>1.61 / 2.21</td>
</tr>
<tr>
<td></td>
<td>HOTEL</td>
<td><strong>0.60 / 0.72</strong></td>
<td>1.17 / 1.91</td>
<td>1.01 / 1.76</td>
<td>1.44 / 2.18</td>
</tr>
<tr>
<td></td>
<td>UNIV</td>
<td>1.01 / 1.59</td>
<td>0.77 / 1.31</td>
<td>0.84 / 1.40</td>
<td>0.75 / 1.28</td>
</tr>
<tr>
<td></td>
<td>ZARA1</td>
<td>0.74 / 1.21</td>
<td>0.53 / 0.88</td>
<td>0.56 / 1.00</td>
<td>0.53 / 0.91</td>
</tr>
<tr>
<td></td>
<td>ZARA2</td>
<td>0.95 / 1.48</td>
<td>0.65 / 1.11</td>
<td>0.70 / 1.17</td>
<td>0.66 / 1.11</td>
</tr>
<tr>
<td>AVG</td>
<td></td>
<td>0.98 / 1.59</td>
<td>0.91 / 1.52</td>
<td>0.91 / 1.54</td>
<td>1.00 / 1.54</td>
</tr>
</tbody>
</table>

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).
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.
Experiments

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.
Experiments
Experiments

Ground Truth Observed

Our Model Observed
Conclusion

• + Merge many methods from other areas into trajectory prediction.
• + Generate multiple predictions with “good behavior”.
• + Good introduction.

• - Experiment comparison is tricky.
• - Bad Writing.