Asymmetrical Bi-RNNs, 3rd place solution at the ICCV Trajnet++ Challenge

Raphaël Rozenberg\textsuperscript{1,2}, Joseph Gesnouin\textsuperscript{2,3} & Fabien Moutarde\textsuperscript{2}

\textsuperscript{1}École Normale Supérieure, Université PSL
\textsuperscript{2}MINES ParisTech, Université PSL, Centre de Robotique
\textsuperscript{3}Institut VEDECOM

October 17, 2021
Modern approaches specifically focused on the presence of social interactions.

We focus on the encoding part of the trajectories of individual people.

---

1[Kothari et al., 2021], ”Human trajectory forecasting in crowds: A deep learning perspective”

2[Alahi et al., 2016] ”Social lstm: Human trajectory prediction in crowded spaces”
A better encoder? From RNNs & Bi-RNNs to U-RNNs

The data has a **preferred direction** in time: the forward direction.

→ We accumulate information while knowing which part of the information will be useful in the future.

\(^1\) [Xue et al., 2017], ”Bi-prediction: pedestrian trajectory prediction based on bidirectional LSTM classification.”
## Results on Trajnet++ real world dataset

<table>
<thead>
<tr>
<th>Model (Encoder - Decoder)</th>
<th>Interaction</th>
<th>ADE (m) ± 0.01 m</th>
<th>FDE (m) ± 0.01 m</th>
<th>Col-I (%) ± 0.5%</th>
<th>Col-II (%) ± 1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant velocity</td>
<td>None</td>
<td>0.68</td>
<td>1.42</td>
<td>14.3</td>
<td>15.2</td>
</tr>
<tr>
<td>None - GRU</td>
<td>Dir.</td>
<td>0.63</td>
<td>1.33</td>
<td>6.9</td>
<td>12.1</td>
</tr>
<tr>
<td>LSTM - LSTM</td>
<td>Occ.</td>
<td>0.58</td>
<td>1.23</td>
<td>11.5</td>
<td>13.9</td>
</tr>
<tr>
<td>U-LSTM - LSTM</td>
<td>Occ.</td>
<td>0.57</td>
<td>1.22</td>
<td>10.2</td>
<td>14.9</td>
</tr>
<tr>
<td>GRU - GRU</td>
<td>Dir.</td>
<td>0.58</td>
<td>1.24</td>
<td>6.5</td>
<td>12.4</td>
</tr>
<tr>
<td>Bi-GRU - GRU</td>
<td>Dir.</td>
<td>0.59</td>
<td>1.26</td>
<td>6.7</td>
<td>11.7</td>
</tr>
<tr>
<td>U-GRU - GRU</td>
<td>Dir.</td>
<td>0.58</td>
<td>1.25</td>
<td>6.5</td>
<td>11.7</td>
</tr>
<tr>
<td>reversed U-GRU - GRU</td>
<td>Dir.</td>
<td>0.58</td>
<td>1.25</td>
<td>6.5</td>
<td><strong>11.0</strong></td>
</tr>
<tr>
<td>LSTM - LSTM</td>
<td>Dir.</td>
<td>0.58</td>
<td>1.25</td>
<td>6.4</td>
<td>11.4</td>
</tr>
<tr>
<td>Bi-LSTM - LSTM</td>
<td>Dir.</td>
<td>0.59</td>
<td>1.28</td>
<td>6.2</td>
<td>11.9</td>
</tr>
<tr>
<td>U-LSTM - LSTM</td>
<td>Dir.</td>
<td><strong>0.56</strong></td>
<td><strong>1.22</strong></td>
<td><strong>5.2</strong></td>
<td>11.9</td>
</tr>
<tr>
<td>reversed U-LSTM - LSTM</td>
<td>Dir.</td>
<td>0.58</td>
<td>1.26</td>
<td>6.6</td>
<td><strong>11.1</strong></td>
</tr>
<tr>
<td>LSTM - LSTM</td>
<td>Soc.</td>
<td>0.55</td>
<td>1.18</td>
<td>6.9</td>
<td><strong>12.7</strong></td>
</tr>
<tr>
<td>Bi-LSTM - LSTM</td>
<td>Soc.</td>
<td>0.53</td>
<td>1.15</td>
<td><strong>6.5</strong></td>
<td>11.5</td>
</tr>
<tr>
<td>U-LSTM - LSTM</td>
<td>Soc.</td>
<td>0.53</td>
<td>1.14</td>
<td>5.3</td>
<td><strong>11.3</strong></td>
</tr>
</tbody>
</table>

1[1]([Liu et al., 2021], "Social NCE: Contrastive Learning of Socially-aware Motion Representations")
Conclusion

- A new sequence encoder.
- Interactions are **NOT** the only aspect on which pedestrian trajectory prediction can progress.
- Could be used to significantly **improve current trajectory prediction algorithms**:
  - TrouSPI-Net [Gesnouin et al., 2021]
  - (GO)-Home [Gilles et al., 2021a, Gilles et al., 2021b]
  
  → [Rozenberg et al., 2021], "Asymmetrical Bi-RNN for pedestrian trajectory encoding", https://arxiv.org/abs/2106.04419
Social Lstm: Human trajectory prediction in crowded spaces.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

TrouSPI-Net: Spatio-temporal attention on parallel atrous convolutions and U-GRUs for skeletal pedestrian crossing prediction.

Gohome: Graph-oriented heatmap output for future motion estimation.

Home: Heatmap output for future motion estimation.

Human trajectory forecasting in crowds: A deep learning perspective.
*IEEE Transactions on Intelligent Transportation Systems*.

Social nce: Contrastive learning of socially-aware motion representations.

Asymmetrical bi-rnn for pedestrian trajectory encoding.

Bi-prediction: pedestrian trajectory prediction based on bidirectional lstm classification.