Spatio-Temporal Graph Transformer Networks for Pedestrian Trajectory Prediction (Appendix)

Cunjun Yu*[1], Xiao Ma*[1,2], Jiawei Ren[1], Haiyu Zhao[1], and Shuai Yi[1]

[1] SenseTime Research, yucunjun@sensetime.com
[2] National University of Singapore, xiao-ma@comp.nus.edu.sg

Additional Attention Visualization

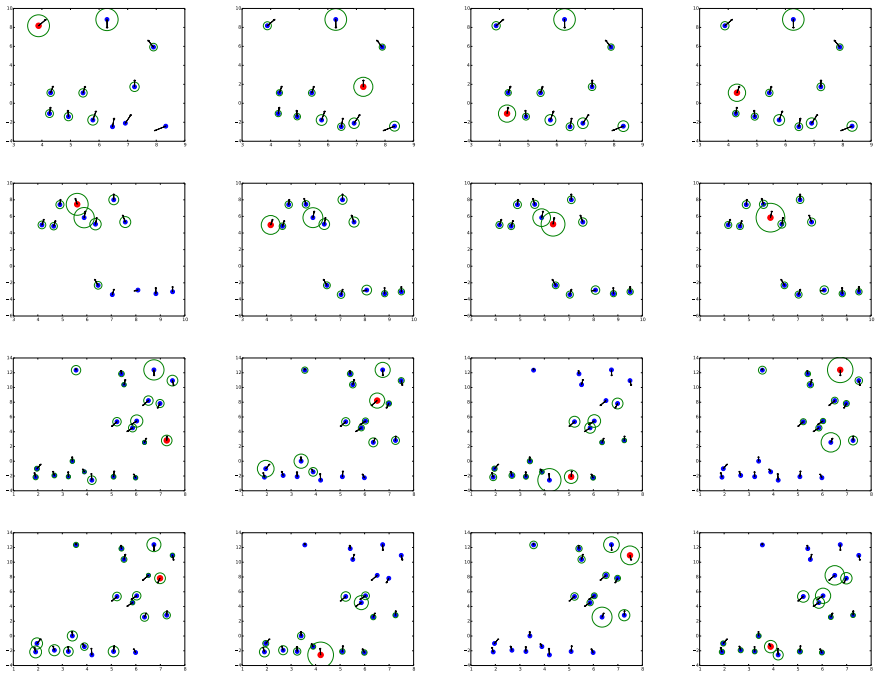
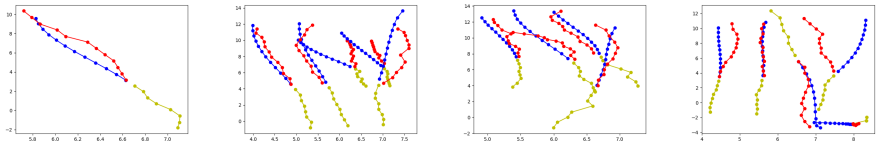


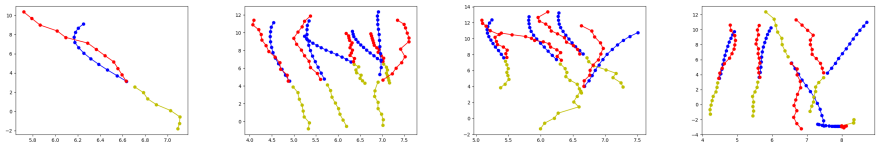
Fig. 1. Additional attention visualizations of the spatial Transformer, i.e., connected in the interaction graph, in encoder 2. We visualize the attention of neighbor pedestrians with respect to the red dotted pedestrian. The size of circles represents the attention value and bigger circles indicate higher attention. STAR learns reasonable spatial attention, the pedestrians have higher attentions over themselves and their neighbors.

* equal contribution, listed in alphabetical order

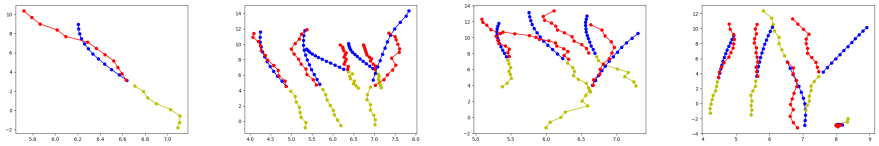
Ablation Trajectory Prediction Visualizations



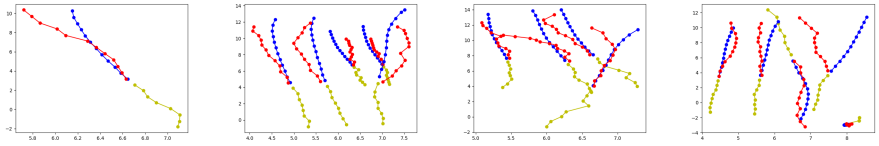
(a) GAT + STAR



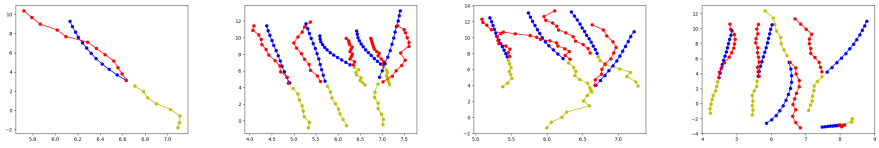
(b) MHA + STAR



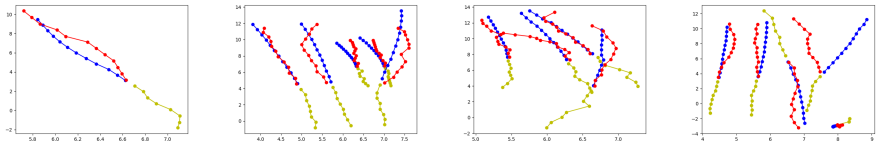
(c) STAR + LSTM



(d) STAR without Graph Memory



(e) Simplified STAR without Encoder 2



(f) STAR

Fig. 2. Trajectory visualization of all ablations. Yellow lines denote the history, red lines denote the ground-truth, and blue lines denote the prediction. Qualitatively, STAR produces best predictions both spatially and temporally.