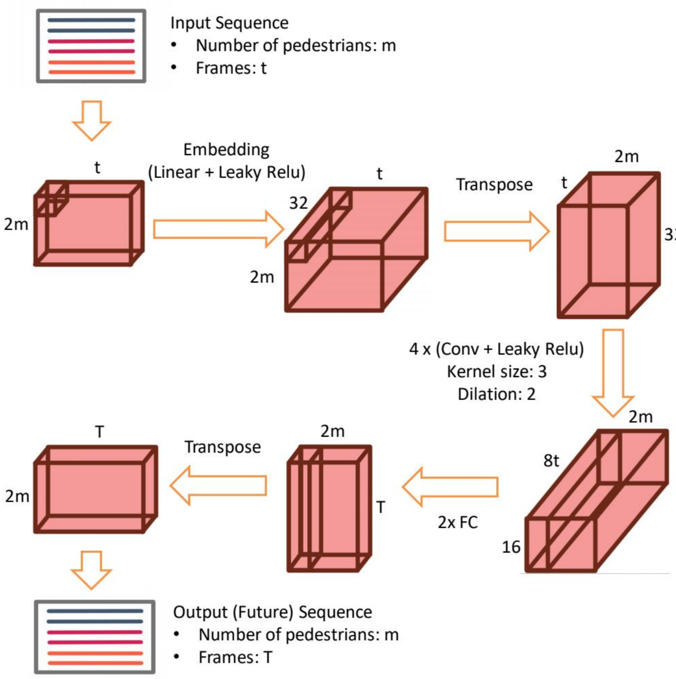
Human Trajectory Prediction in Socially Interacting Crowds

Bilan Jackie Yang and Haruki Nishimura
{bilany, hnishimura}@stanford.edu

Problem Description

This project aims to predict human trajectories in dense crowds for motion planning applications using Convolutional Neural Networks (CNN). Inspired by recent success of CNN for sequence-to-sequence tasks [1], we will compare the proposed CNN model against Social-LSTM [2], a more standard Recurrent Neural Network (RNN) recently proposed.

Proposed Network Architecture



Inspired by prior work [3], we designed the model above to map the input sequence to the output sequence.

- Dilated convolution over pedestrians is for capturing interrelations.
- Batch Norm performed after each Conv layer
- Dropout added between two FC layers
- L2 Loss function with Adam optimizer

$$\ell(x,y) = L = \{l_1, \ldots, l_N\}^\top, \quad l_n = (x_n - y_n)^2$$

Future Research Directions

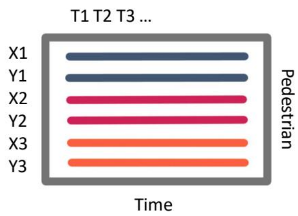
- Extend the current model to multimodal prediction (e.g. Gaussian mixture fitting), which is important for safe motion planning.
- Conduct extensive comparable study with other methods.
- Incorporate scene-specific information in addition to human-human interactions.

Data Acquisition and Preprocessing

We use a publicly available dataset from Stanford Trajectory Forecasting Benchmark [6]. The dataset contains the ground truth (x,y)-coordinates of each pedestrian in the scene at various locations, along with the unique pedestrian ID and the timestamp.

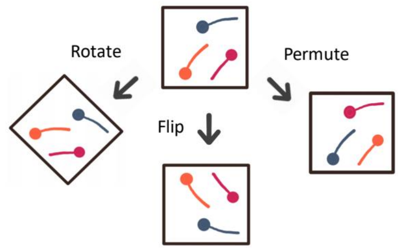
Data Preprocessing

A custom input pipeline was created where we represent the joint trajectories in a 2D matrix. Each row contains either x or y values of a pedestrian over time.



Data Augmentation

We performed data augmentation by flipping and rotating the coordinates, as well as permuting pedestrian orders in the matrices to learn from the small dataset.

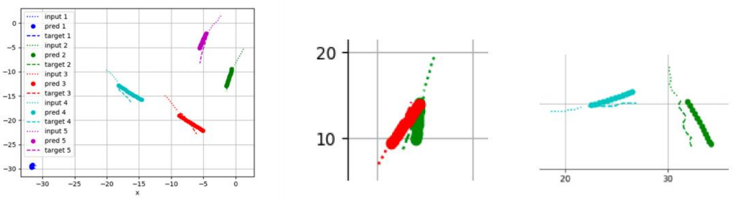


Results & Discussion

Models		Train time (min) / epoch	Test size / train size	Train error	Test error	Train error - test error	Mean displacement error	Final displacement error
(1) normal, mix_all_data	Input/output sequence: 5/5	0.62	1173 / 9384 = 0.13	0.72	0.53	0.192	0.843	1.085
(2) normal, specify_test_set		1.01	576 / 11010 = 0.05	0.74	0.49	0.246	0.852	1.124
(3) individual, mix_all_data		0.69	74818 / 598552 = 0.13	0.46	0.16	0.3	0.363	0.548
(4) individual, specify_test_set		0.75	34143 / 679902 = 0.05	0.45	0.32	0.125	0.614	0.841
(1) normal, mix_all_data	Input/output sequence: 8/12	0.15	141 / 1255 = 0.11	0.75	1.54	-0.798	1.408	1.984
(2) normal, specify_test_set		0.13	24 / 1380 = 0.02	0.7	0.93	-0.229	1.209	1.414
(3) individual, mix_all_data		0.08	6772 / 60279 = 0.11	0.91	0.83	0.083	0.81	1.464
(4) individual, specify_test_set		0.07	3093 / 61542 = 0.05	0.9	0.79	0.11	0.856	1.614

Note: Nomal: all pedestrians in one frame form an example; individual: one pedestrian forms an example

- Input/output sequence of 5/5 has better results than 8/12
- The individual dataset performas better than the normal; it contains more data but was trained faster



(a) (b) (c)

Test set trajectory plots (in meters): dashes are the ground truth; dots are predictions.

- (a) Our model succesfully predicts the general trend of the pedestrians
- (b) The pedestrian interactions should be better capture for predictions
- (c) Predictions suffer from the uncertainty of human behaviors.

[1] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. ArXiv preprint arXiv:1705.03122, 2017.
[2] 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
[3] Nishant Nikhil and Brendan Tran Morris. Convolutional neural network for trajectory prediction. ArXiv preprint arXiv:1809.00696, 2018.

Acknowledgements

The authors are grateful to Amir Sadeghian and the teaching staff, especially Abhijeet Shenoi for helpful guidance and discussions.