SOCIAL CONTEXT BASED AGENT NETWORK TRAJECTORY PREDICTION

Manoj Bhat, Umut Soysal, Suyash Narain
Mechanical Engineering, Carnegie Mellon University

Abstract

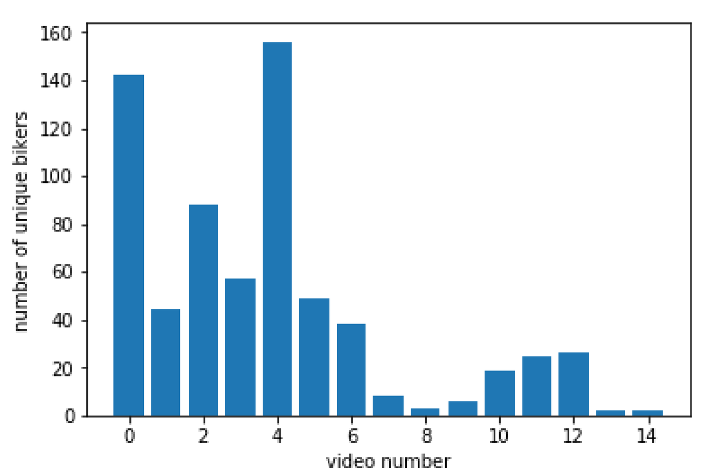
The aim of this project was to develop an agent based tracking and prediction model for subsequent use in autonomous vehicles. The moving agent tries to learn the motion behavior of external disturbances and predict the trajectory of moving disturbances and plan its path accordingly to avoid them. We analyze how the human behaves in external circumstances and make the agent learn to mimic these behaviors in given social context.

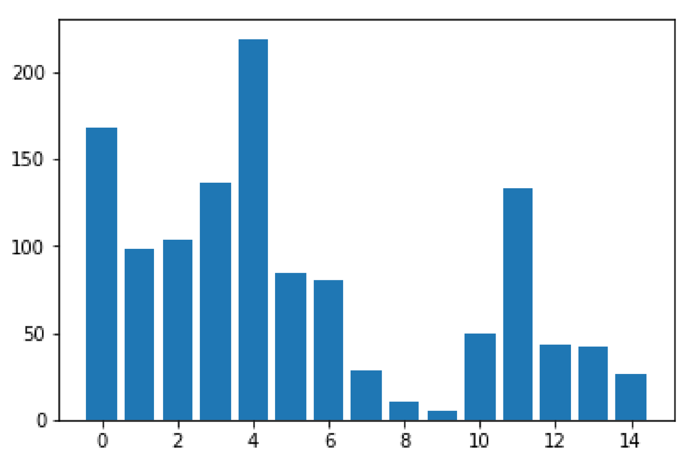
Problem Formulation

There exist many pedestrian path prediction models but none explain how the agent should behave in given social contexts. At the ground level of autonomous vehicles, if something goes wrong (e.g. Lidar fail), we can use continuous (GPS) real time data to avoid undesired motions. For this, the same agent should be able to take precise calculated steps and velocities. We will try to implement the same on RNNs and its derivatives like LSTM networks., and train using Stanford drone dataset.

The pedestrians are considered disturbances and powered vehicles like Bike, Carts and skaters are considered as Agents.

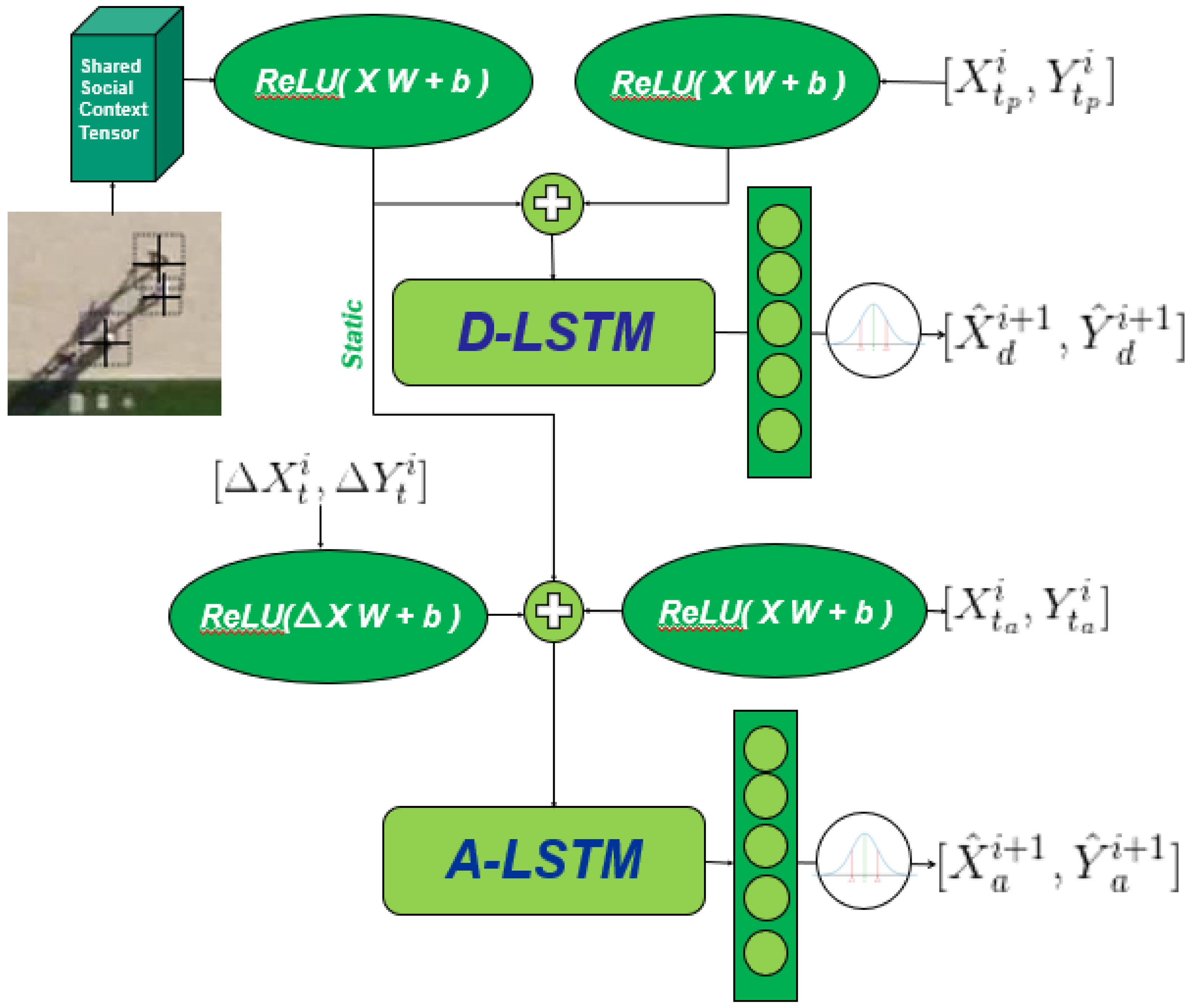
The dataset used for the training and testing is Stanford Drone Dataset which has top view video and pre-tracked well annotated data. The data structure chosen is [Time-frames, max people/agents, (PedId, x, y)]. Avg Frames in videos is 16,327 (seqs), there are maximum of 32 people/agent data in a frame.

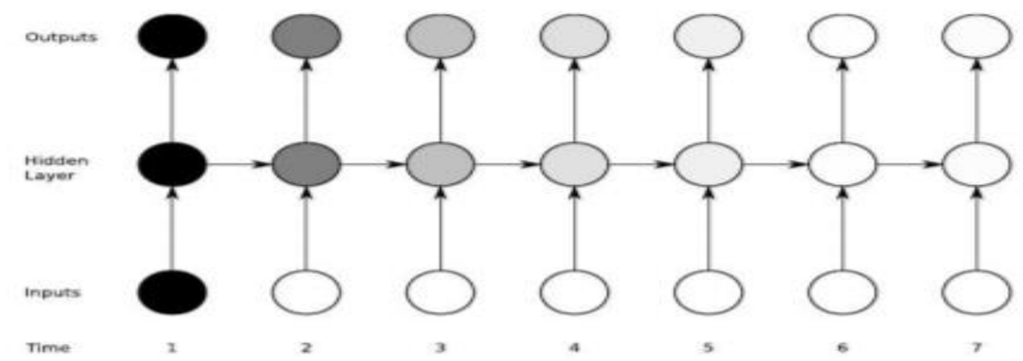




Proposed Methods and Results

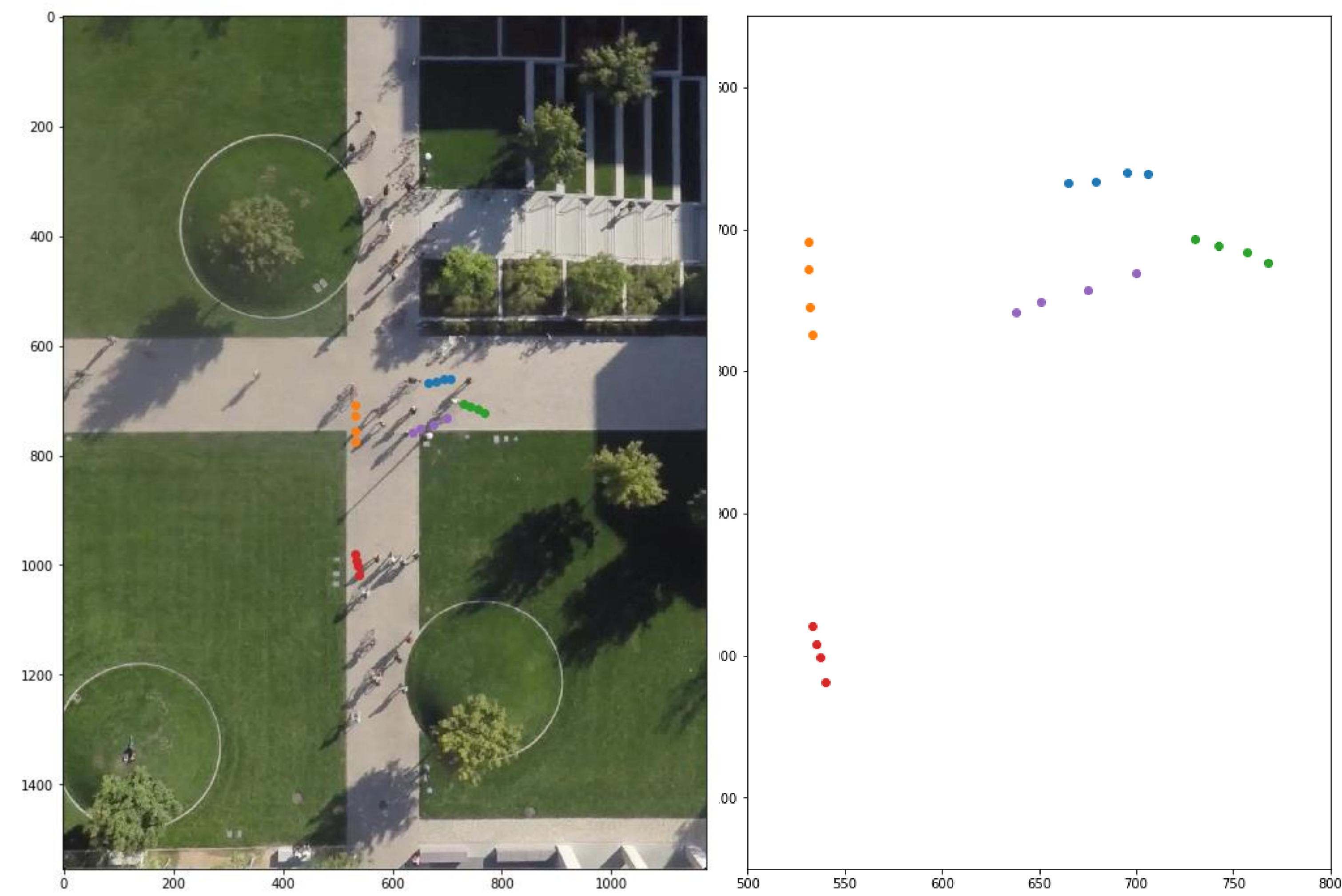
Prediction usually comes under the ambit of time series based sequential learning problems, and RNNs are best suited to tackle them. We try to work upon out model using vanilla RNN, and then proceed with LSTM and GRU for path prediction of agent vehicles. We then move on to create more robust model to tackle the problem.





The default learning rate is 0.005 for which the model was found to be stable-costed with a decay rate 0.95. Adagrad and RMSprop are used as optimizers. The loglikelihood is minimized for end parameters. And the LSTM outputs are mapped to 5 distribution parameters which describe the X,Y of pedestrians. The vanilla LSTM gave very poor results for the agent Biker model. But, the results without the velocity pool layer for proposed model as shown below is fascinating for the given timestep as the prediction of biker is to cross the person. More work will be done in improving metrics.

$$L^i(W_e, W_l, W_p) = -\sum_{t=T_{obs}+1}^{T_{pred}} \log\left(\mathbb{P}(x^i_t, y^i_t | \sigma^i_t, \mu^i_t, \rho^i_t)\right)$$

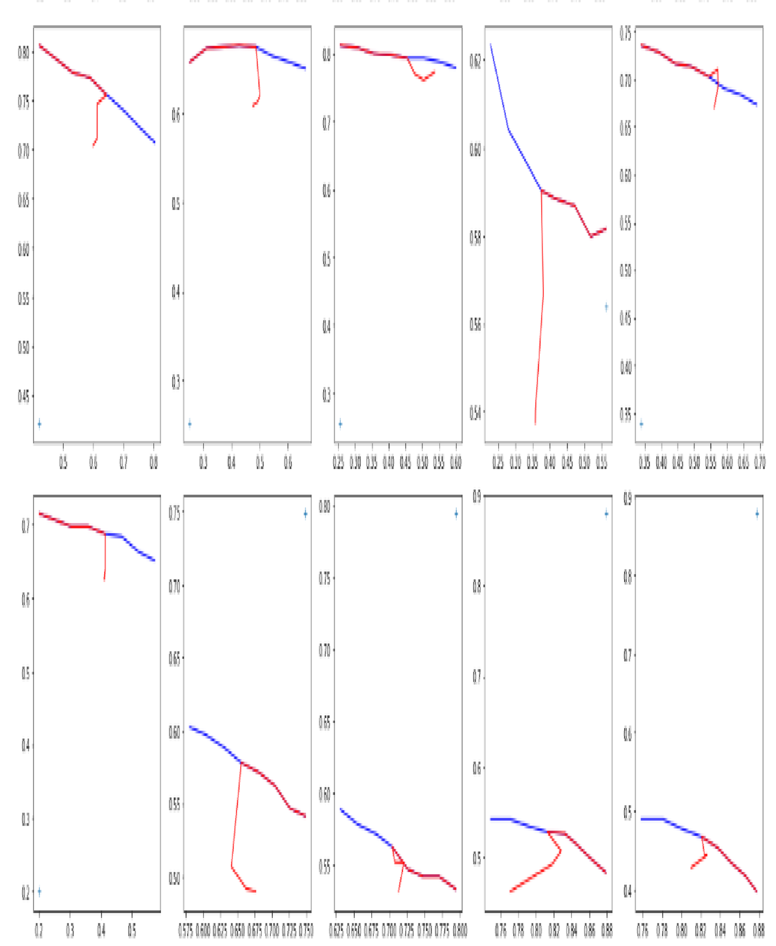


The proposed network architecture has a Coordinate Tensor which is based on the coordinates at the given time and a Social Context tensor which is shared with the agent network. This tensor is an embedding which follows the Social Attention LSTM model of social grids. We want our model to learn the essence of relative distances for which gradients are representative of velocity (pool layer). The concatenation helps the agent LSTM network to learn as much detailed inputs. The final prediction tensor consists sequence of position data of both the disturbances and the agents.

Conclusion and future work

The model has improved the predictions compared to vanilla LSTM model. The path observed are representative of human behaviors. We would like to do a complete clustered analysis of the predicted paths and actual trajectories to analyze the human behavioral tendency.

Which shows potential for the same. We intent use different loss and accuracy metrics to compare RNN models and we try to improve the accuracy with multilayered Bidirectional LSTM i.e Stacked LSTM and also HMM for probabilistic estimate of direction change and collision statistic.



References:

Alahi, Alexandre, et al. "Social lstm: Human trajectory prediction in crowded spaces." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

Manh, Huynh, and Gita Alaghband. "Scene-LSTM: A Model for Human Trajectory Prediction." arXiv preprint arXiv:1808.04018 (2018).

Su, Hang, et al. "Crowd Scene Understanding with Coherent Recurrent Neural Networks." IJCAI. Vol. 1. 2016.

Vemula, Anirudh, Katharina Muelling, and Jean Oh. "Social attention: Modeling attention in human crowds." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.