Predicting Vulnerable Road Users Using Stacked LSTM Network

Md Shiful Islam
Technische Hochschule Ingolstadt (THI)

***Abstract*— Recently, researchers have been looking into predicting what pedestrians and other vulnerable road users might do. This is important for making driving systems better and safer. Most methods for this focus on how people move and plan their paths. These methods often need specific details about each situation, which makes it hard to use them in new situations. This paper talks about a new way to predict what pedestrians might do. It looks only at how pedestrians move over time. By watching a short sequence of their movements, researchers can guess where they'll be in the future, up to 4 seconds ahead. To do this, they use a type of computer program called a neural network, specifically Long-Short Term Memory networks (LSTM).The researchers tested this method using a big dataset that has different traffic situations, like busy streets. The results show that this method works well. It can predict where pedestrians will be pretty accurately, even up to 4 seconds ahead. This could help make driving safer, especially in urban areas.**

I. INTRODUCTION

In recent years, there has been significant momentum in the advancement of highly and fully automated vehicles. This surge is primarily driven by their potential to reduce road accidents and fatalities caused by human driver errors. Nevertheless, these vehicles encounter challenges in comprehending the behaviors and intentions of other road users, particularly vulnerable ones like pedestrians [1]-[3].

For that reason, autonomous vehicles (AVs) are becoming more popular, with big car companies like Toyota, Mercedes Benz, GM, Ford, Audi, and tech giants like Google and Uber all getting involved. Even though AVs are expected to reduce accidents caused by human drivers, they still face many challenges [4]. Especially in busy city areas, AVs struggle, especially with things like interacting with people walking on the road [5]. Right now, when people drive, they use unspoken signals to understand each other. One of these signals is how people first move, which can show what they're planning to do [6]. Researchers have been studying how to predict what pedestrians will do in cities by looking at how they move for the past 5 years. But most of this work focuses on one type of model, which isn't great at predicting different kinds of movements. For example, if someone suddenly stops while crossing the road, this model might not predict it well [6]-[8]. Some other approaches, like ones from robotics, also try to predict what pedestrians will do based on how they move. These are better at handling longer predictions, but they still need to know where people are going [9]. This can be hard to figure out from the perspective of a moving car. Both of these methods have limitations because they need a lot of manual work to set up, which makes them less useful in new situations [10].

In this study, a data-driven method is introduced for predicting pedestrians' intentions in urban traffic environments based on their motion trajectories. The task is approached as a time series prediction problem, and a variant of Recurrent Neural Networks (RNN) known as Long-Short Term Memory networks (LSTM) is employed. This model predicts a long-term sequence (up to 4 seconds) of pedestrians' motion trajectories from the perspective of a moving observer, such as a vehicle. By adopting this approach, we aim to overcome the limitations associated with dynamical motion models and planning-based models discussed in prior research [11].

The rest of this paper is organized as follow. Section II, provides an overview of the related work. In Section III, details of our proposed RNN-LSTM model. In Section IV, quantitative and qualitative experimental results will be presented. Finally, we summarize our paper in Section V.

II. RELATED WORK

In recent years, there has been growing momentum in the intelligent transportation and robotic communities regarding intent prediction of pedestrians in urban traffic environments. This section will provide a brief overview of the related work in these areas.

A. Dynamical Motion Models

The predominant approach in intelligent transportation systems for predicting pedestrians' intentions from motion trajectories has been the dynamical motion model approach. Schneider et al. [6] utilized Bayesian filters, such as the Extended Kalman filter (EKF) [12], to predict pedestrian motion trajectories from a vehicle's perspective within a short prediction horizon (less than 2 seconds). They applied this approach across four different motion dynamics: bending-in, crossing, starting, and stopping. Additionally, they introduced another Bayesian filter based on the Interacting Multiple Model (IMM) KF to accommodate various dynamical motion models of pedestrians, including constant velocity (CV), constant acceleration (CA), and constant turn (CT).

Similarly, Kooij et al. [7] proposed another dynamic motion model using a Dynamic Bayesian Network (DBN) for predicting a pedestrian's trajectory when intending to cross the street while walking on the curb. They hypothesized that a pedestrian's decision to stop and cross the street or continue walking on the curb depends on three factors: the presence of an approaching vehicle at a potential collision point, the pedestrian's awareness of this scenario, and the layout of the physical environment surrounding the pedestrian. By treating these factors as unobservable variables within a Switching

Linear Dynamical System (SLDS) as part of the DBN, they could predict to some extent the changes in pedestrians' motion dynamics.

B. Planning-based Models

In contrast, planning-based models for predicting pedestrians' intentions do not explicitly consider the dynamical motion of pedestrians' trajectories. Instead, they frame the problem as a motion or path planning task, assuming that pedestrians are rational agents with hidden intentions to reach known specific destinations. These models predict that pedestrians will choose an optimal path, typically the shortest path, to reach their goal.

For instance, Rehder et al. [13] proposed a model for long-term prediction of pedestrian intention to reach a specific destination. They achieved this by estimating the probability distribution over the pedestrian's future positions using path planning techniques. By considering the pedestrian's position, orientation, and an online-recorded grid occupancy map of the environment, they could estimate the pedestrian's goal destination as a latent variable. To do this, they discretized the grid occupancy map into independent cells, each containing a vector of location-weighted features. These weights were calculated using a supervised learning model trained with ground truth trajectories of pedestrians and their corresponding grid maps. The goal destination of the pedestrian was modeled as a Gaussian Mixture Model, which was iteratively improved using a Particle filter.

C. Data-Driven Approaches

Recently, data-driven approaches, particularly those utilizing deep hierarchical representation layers like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated significant success across various spatial and sequential tasks [14], [15]–[17]. For instance, Alahi et al. [18] employed an RNN-based data-driven approach to develop a social behavior model for estimating pedestrians' motion trajectories in crowded environments using surveillance cameras. This model implicitly captured the interactions between pedestrians and their neighbors.

In a more recent study, Volz et al. [19] proposed a data-driven vehicle-based approach for recognizing pedestrians' intentions at urban traffic intersections, distinct from the surveillance camera-based approach of [18]. They framed the intention recognition problem as a time series binary classification task, where given a processed sequence of features, they classified whether a pedestrian would cross the road. Their approach utilized temporal features (position/velocity) and geometrical features (distance to curb) extracted from 3D LiDAR data collected at the intersection. Additionally, they introduced two data-driven approaches for pedestrian intention recognition based on CNNs and RNNs, respectively.

D. Another planning-based approach

As detailed in [20], focused on an on-road pedestrian avoidance system designed for an autonomous mobile robot. This system integrated considerations of pedestrians' intentions and associated uncertainty into the robot's motion planning framework. The problem was formulated as a Mixed Observable Markov Decision Process (MOMDP), where the pedestrian's motion model variables were fully observable, but their intention remained unknown. It was assumed that pedestrians followed the shortest path trajectory toward their goal. To address this, a sampling-based approximate algorithm called Successive Approximations of the Reachable Space under Optimal Policies (SARSOP) was utilized. This approach enabled the solving of the MOMDP model and the inference of a probability distribution over the potential directions of pedestrians.

III. METHODOLOGY

In this section, the underlying operation of recurrent neural networks (RNNs) will be discussed, particularly in their application to time series prediction tasks. Subsequently, the formulation of our problem of interest, namely, intent prediction of pedestrians from their motion trajectories, as a time series prediction task using RNNs will be described. Secondly, our proposed approach to address the formulated problem will be introduced. Finally, the details of the proposed model for achieving long-term intent prediction of pedestrians in urban traffic environments will be presented.

A. Time Series Prediction using Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNN) are often used for tasks like imitating handwriting [21], analyzing how people walk [22], or studying human interactions [23]. Unlike other neural networks, RNNs have loops that help them remember previous information, called "hidden units." However, traditional RNNs struggle with remembering long sequences. To solve this, Long Short-Term Memory (LSTM) networks were introduced [24]. LSTM networks have a special unit called a memory block, similar to the hidden units in regular RNNs. Each block has memory cells and three gates: forget, input, and output. These gates decide what information to remember, update, or output based on the input.

In LSTM networks, the hidden layer's operations are calculated using formulas like these:

$$f_t = \text{sigm}\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \quad (1)$$

$$i_t = \text{sigm}\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \quad (3)$$

$$o_t = \text{sigm}\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right) \quad (4)$$

$$h_t = o_t * \tanh\left(c_t\right) \quad (5)$$

- f_t decides what to forget from previous memory (c_{t-1}).
- i_t decides what new information to add to the memory.
- c_t updates the memory by combining old and new information.
- o_t decides what to output based on the updated memory.
- h_t is the final output based on the updated memory.

While W_{*f}, $W_{*i}, W_{*o}, W_{*c}, b_f, b_i, b_o, b_c$ are their respective weight matrices and variable biases. x_t, h_t are the memory cell input and final output at time t. From previous equations, we can notice that, each gate from the three gates of the LSTM is just a composition of a sigmoid neural network layer and an elementwise multiplication operation. The sigmoid layer clips its input into a value between one and zero, whereas zero means "no input will be passed through" and one means "let the input to be passed through".

B. Intent Prediction of VRUS using Stacked LSTM Network

The motion paths of pedestrians, essentially a timeline of their sideways movement on the ground, are like signals recorded at regular intervals, as described in different models [6], [10]. Understanding the intentions of pedestrians in city traffic can be seen as predicting their future movements based on this recorded path. At any given moment, we can guess a pedestrian's intentions by looking at where they are on their path. By tracking their position from the start to a certain point, we can make an educated guess about where they'll go next.

Our solution involves using a deep network made up of stacked LSTM blocks to predict pedestrian intentions from their paths. This network has three layers of LSTM blocks, as seen in Fig. 1.

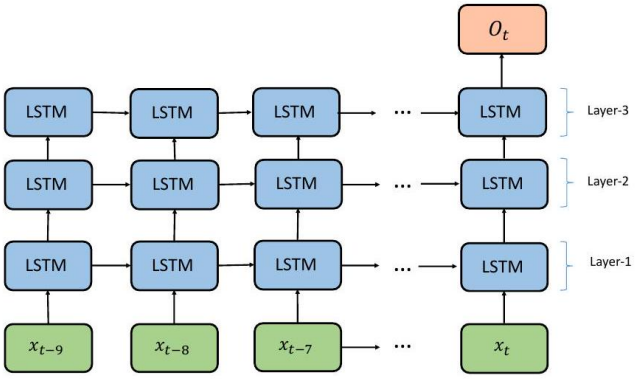

Fig. 1. Proposed LSTM based model for intent prediction of pedestrians's of their motion trajectories[11].

The first layer takes in a sequence of lateral positions (usually ten points) and passes it on to the next layer, which has 100 hidden units. This process repeats twice more, each time with another layer of 100 hidden units. Finally, the last layer has just one neuron, predicting the pedestrian's next lateral position.

We use a linear activation function in the last layer since we're predicting real numbers. During training, the model learns to predict just the next position in the sequence. But during testing, it can predict sequences of any length, giving us flexibility in how we use it. This way, we can adapt it to different scenarios without needing to change its structure or the data we test it on.

C. Stacked LSTM Network Training

Training a learning-based model for time-series prediction involves minimizing a loss function, typically measured by mean squared error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2 \tag{9}$$

where N is the number of training samples, $\hat{Y}_i$ and Y_i are the predicted and target values for each sample.

This function calculates the average squared difference between predicted and actual values for each sample. To optimize this loss function, we used the Adam optimizer for training our LSTM model [25]. Adam is a stochastic gradient descent algorithm that efficiently estimates the gradient mean and squared gradient using exponential moving averages.

a. Crossing.

b. Starting.

c. Bending In.

d. Stopping.

Fig. 2. The four scenarios of Daimler pedestrian path prediction benchmark dataset [9].

One of the main advantages of using Adam is its low number of hyperparameters, with only the learning rate needing tuning compared to other optimizers. For our stacked LSTM architecture, we set the learning rate to 0.001 based on several training experiments using 3-fold cross-validation. Additionally, we applied dropout of 20% after each LSTM layer to prevent overfitting during training [26]. We trained the stacked LSTM network for 10,000 iterations with a batch size of 512.

IV. EXPERIMENTS

In this section, the details of the dataset used for training and testing the stacked LSTM model will be provided, along with an overview of the preprocessing steps undertaken to prepare the dataset for the model. Subsequently, the experimental results of the stacked LSTM model's performance on the testing data will be presented, with comparisons to two other major approaches for pedestrian intent prediction.

A. Data Description

For training and testing the performance of the proposed stacked LSTM model for pedestrian intent prediction from motion trajectories, the widely used Daimler pedestrian path prediction benchmark dataset [6] will be utilized (as depicted in Fig. 2). This dataset comprises 68 stereo image sequences captured from vehicle-based cameras positioned behind the vehicle's windshield. These sequences cover four primary scenarios, as defined in [6], encompassing various dynamics of pedestrian motion in urban traffic environments.

The first scenario involves a pedestrian walking laterally towards the street with the intent to cross, labeled as "Crossing." This subset of the dataset contains 18 crossing sequences, with 9 designated for training and 9 for testing. The second scenario, termed "Stopping," entails a pedestrian initially walking laterally towards the street and then stopping. This subset comprises 18 stopping sequences, with 10 for training and 8 for testing. In the third scenario, a pedestrian starts from a stationary position at the road curb and begins walking laterally facing the street, termed as "Starting." This subset contains 9 starting sequences, with 5 for training and 4 for testing. Finally, the fourth scenario involves a pedestrian walking alongside the curb road and attempting to bend towards crossing the street, termed "Bending in." This subset comprises 23 bending in sequences, with 12 for training and 11 for testing.

Each sequence within the dataset is meticulously labeled frame by frame, including bounding boxes of pedestrians, median disparity of the upper body area of the pedestrian, pedestrian position in the vehicle coordinate system, and event tags denoting "Time to Event" (TTE) values. These TTE values indicate the moment in each sequence of the four scenarios when the pedestrian is expected to cross, stop, start to cross, or bend in to cross.

B. Data Preparation

The dataset undergoes preprocessing to prepare it for feeding into the stacked LSTM model during the training phase. Initially, the dataset, comprising 36 training sequences and 32 testing sequences, is already divided, covering the four scenarios.

In the preprocessing stage, the ground truth labels for lateral positions (in meters) of pedestrians in each sequence from both the training and testing datasets are parsed. Subsequently, a sliding window approach is applied to each sequence in the training dataset, with a window size of w+1 and an overlap of 1. Here, w represents the size of the input layer feeding into the first LSTM layer of the stacked LSTM model, which is illustrated in Fig. 1 with an input window size of 10. This results in 4492 training samples of window size w+1 from all training sequences.

TABLE I
MEAN LATERAL POSITION ERROR (IN METERS) OVER ALL THE TESTING SEQUENCES OF EACH SCENARIO WITH TWO PREDICTION WINDOWS OF 70 AND STEPS AHEAD [11]

		Bending in	Crossing	Starting	Stopping
EKF-CV [6]	Mean	1.09	0.72	1.31	**0.22**
	± Std	0.27	0.39	0.50	0.34
IMM-CV/CA [6]	Mean	1.08	0.68	1.32	0.24
	± Std	0.27	0.40	0.52	0.35
LSTM-2L [19]	Mean	0.79	1.21	0.76	1.01
	± Std	0.19	0.30	0.06	0.50
Stacked LSTM	Mean	**0.39**	**0.48**	**0.46**	0.51
	± Std	0.24	0.32	0.07	0.37

Further, the 4492 window samples of window size w+1 are split into separate training samples of window size w and target values of size 1. Finally, approximately 5% of the training window samples, roughly 255 window samples, are randomly selected for use as a cross-validation set during the training of the stacked LSTM model [11].

V. OUTCOME

We tested our stacked LSTM model on the testing sequences outlined in Section IV-A. Our evaluation process followed the methodology used in [3]. We evaluated the lateral position of each sequence within the TTE range [10, -50], which spans a window of 60 steps in total, from 0.60 seconds before the event to 3.0 seconds after the event. In

contrast to [3], where a 32-step prediction (1.9 seconds) was made, we predicted 70 steps ahead (over 4 seconds) to focus on long-term pedestrian intent prediction. During testing, we input a sequence of size 10 before the start of the TTE range and predicted the entire range. To compare our model's performance, we implemented similar dynamical models to EKF (CV) and IMM (CV, CA) models used in [3] as baselines. Additionally, we compared against an LSTM-based model proposed in [17], denoted as LSTM-2L, which showed limited improvement over an SVM-based model in a similar task. Table I summarizes the average mean error in lateral position (in meters) for a 70-step prediction window across each scenario of the testing sequences. Our stacked LSTM model outperforms EKF-CV, IMM-CV/CA, and LSTM-2L models in all scenarios except "Stopping".

Further analysis of the "Stopping" scenario revealed discrepancies due to unique testing sequences not present in the training dataset. This affected the mean lateral position error. The lateral position error for our LSTM model compared to EKF-CV and IMM-CV/CA dynamical models across the four testing scenarios is illustrated. Our stacked LSTM model shows improvement, with errors reduced by up to 0.85 m, 0.7 m, and 0.24 m in the "Starting", "Bending In", and "Crossing" scenarios, respectively. We also compared our stacked LSTM model's performance with a smaller prediction horizon of 15 steps. Despite this, our model generally outperforms both the dynamical motion models and the LSTM-2L model.

TABLE II
MEAN LATERAL POSITION ERROR (IN METERS) OVER ALL THE TESTING SEQUENCES OF EACH SCENARIO WITH PREDICTION WINDOWS OF 15 STEPS AHEAD [11]

		Bending in	Crossing	Starting	Stopping
EKF-CV [6]	Mean	0.44	0.58	0.44	**0.03**
	± Std	0.12	0.07	0.03	0.01
IMM-CV/CA [6]	Mean	0.48	0.66	0.49	0.05
	± Std	0.13	0.08	0.05	0.01
LSTM-2L [19]	Mean	0.32	0.54	0.34	0.76
	± Std	0.09	0.08	0.04	0.34
Stacked LSTM	Mean	**0.04**	**0.07**	**0.05**	0.09
	± Std	0.02	0.03	0.01	0.05

Notably, the stacked LSTM layers contribute to improved prediction results compared to a shallow LSTM network (LSTM-2L).

VI. CONCLUSIONS

In this paper, a novel data-driven approach for long-term intent prediction of pedestrians from motion trajectories is presented. The problem is framed as a time-series prediction task, and a stacked LSTM (Long Short-Term Memory) architecture is employed to model the sequential nature of pedestrian movement. This LSTM-based model is trained using a widely recognized public dataset, which captures pedestrian motion trajectories in various urban traffic scenarios. The proposed stacked LSTM architecture effectively captures the temporal dependencies in pedestrian motion data, enabling it to make accurate predictions over both short and long-term horizons. By leveraging the LSTM's ability to retain and process information across extended sequences, the model demonstrates a significant improvement in predictive performance compared to traditional dynamical motion models. These models, which include the Extended Kalman Filter (EKF) and the Interacting Multiple Model (IMM), have been commonly used for pedestrian intent prediction but typically struggle with long-term forecasting due to their reliance on simpler assumptions about pedestrian movement dynamics. Evaluation results indicate that the stacked LSTM architecture consistently outperforms these dynamical motion models in terms of mean lateral position error, a key metric for assessing prediction accuracy. This superior performance is observed across most of the four distinct scenarios tested within the dataset, which include pedestrians crossing the street, stopping, starting to walk, and bending in towards the street. Specifically, the LSTM model shows marked improvements in prediction accuracy over longer horizons, which is crucial for enhancing the safety and efficiency of intelligent transportation systems.

In summary, the data-driven LSTM approach not only provides a robust framework for capturing the intricate patterns in pedestrian trajectories but also offers significant advancements in predictive accuracy for long-term pedestrian intent prediction. This makes it a valuable tool for urban traffic management and autonomous vehicle navigation, where understanding and anticipating pedestrian behavior is critical for preventing accidents and ensuring smooth traffic flow.

ACKNOWLEDGMENT

We would like to acknowledge the authors Khaled Saleh, Mohammed Hossny, and Saeid Nahavandi for their work on "Intent Prediction of Vulnerable Road Users from Motion Trajectories Using Stacked LSTM Network", which served as a foundation for the methodology employed in this study.

REFERENCES

[1] M. Mara and C. Lindinger, "Talking to the robocar—new research ap- proaches to the interaction between human beings and mobility machines in the city of the future," pp. 86–91, 2015.

[2] M. Wagner and P. Koopman, "A philosophy for developing trust in self- driving cars," in Proc. Road Vehicle Autom., Springer, 2015, pp. 163–171.

[3] D. Rothenbu cher, J. Li, D. Sirkin, B. Mok, and W. Ju, "Ghost driver: A field study investigating the interaction between pedestrians and driver - less vehicles," in Proc. 25th IEEE Int. Symp. Robot Human Interactive Commun., 2016, pp. 795–802.

[4] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," Transportation Research Part A: Policy and Practice, vol. 77, pp. 167–181, 2015.

[5] J. Wang, Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style Submitted for publication), IEEE J. Quantum Electron., submitted for publication.

[6] N. Schneider and D. M. Gavrila, "Pedestrian path prediction with recur- sive Bayesian filters: A comparative study," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8142 LNCS, pp. 174–183, 2013.

[7] C. G. Keller and D. M. Gavrila, "Will the pedestrian cross? A study on pedestrian path prediction," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 2, pp. 494–506, 2014.

[8] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-based pedestrian path prediction," in European Conference on Computer Vision. Springer, 2014, pp. 618–633.

[9] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, "Activity forecasting," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7575 LNCS, no. PART 4, pp. 201–214, 2012.

[10] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, "Planning-based prediction for pedestrians," 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, pp. 3931–3936, 2009.

[11] Saleh, K., Hossny, M., and Nahavandi, S, "Intent Prediction of Vulnerable Road Users from Motion Trajectories Using Stacked LSTM Network,"IEEE 20th International Conference on Intelligent Transportation Systems Workshops, pp. 327-332, 2017.

[12] S. M. Mohamed and S. Nahavandi, "Robust finite-horizon kalman filtering for uncertain discrete-time systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1548–1552, 2012.

[13] E.RehderandH.Kloeden,"Goal-DirectedPedestrianPrediction,"IEEE International Conference on Computer Vision Workshops, 2015.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 91–99.

[16] R.Girshick,J.Donahue,T.Darrell,andJ.Malik,"Richfeaturehierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 580–587.

[17] K.Saleh,M.Hossny,andS.Nahavandi,"Earlyintentpredictionofvulnerable road users from visual attributes using multi-task learning network," in Proc. IEEE Int. Conf. Syst., Man, Cybern., 2017, pp. 3367–3372.

[18] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 961–971.

[19] B. Vo lz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "A data-driven approach for pedestrian intention estimation," in Proc. IEEE 19th Int. Conf. Intell. Transp. Syst., 2016, pp. 2607–2612.

[20] T. Bandyopadhyay, C. Z. Jie, D. Hsu, M. H. Ang, D. Rus, and E. Frazzoli, "Intention-aware pedestrian avoidance," in Experimental Robotics. Heidelberg, Germany: Springer, 2013, pp. 963–977.

[21] A. Graves, "Generating sequences with recurrent neural networks," arXiv preprint arXiv:1308.0850, 2013.

[22] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent network models for human dynamics," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4346-4354.

[23] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-fei, and S. Savarese, "Social LSTM : Human Trajectory Prediction in Crowded Spaces," CVPR, 2016.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[25] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[26] S. Nitish, "Improving neural networks with dropout," Diss., University of Toronto, 2013.