2019 IEEE Intelligent Transportation Systems Conference (ITSC)
Auckland, NZ, October 27-30, 2019

A Survey on Path Prediction Techniques for Vulnerable Road Users: From Traditional to Deep-Learning Approaches

Ariyan Bighashdel and Gijs Dubbelman

***Abstract*—Behavior analysis of Vulnerable Road Users (VRU)s has become a crucial topic in the computer vision research area. In recent decades, numerous papers have extensively addressed the problem of VRU path prediction, which has a wide range of applications such as video surveillance and autonomous driving. The behavioral complexity of VRUs has forced researchers to employ various techniques in order to develop more comprehensive models that potentially respond better to VRUs movement patterns. This indeed has led to development of a large variety of models and approaches in the literature. The aim of this paper is to a) provide a comprehensive review of developed path prediction methods, b) individuate and classify the proposed methods from multiple viewpoints, and c) present a framework for better understanding of various aspects in VRUs path prediction problems.**

I. Introduction

During recent decades, Advanced Driving Assistance Systems (ADAS)s have proposed numerous benefits in reducing the number of traffic accidents, which are mainly due to driver errors [1]. Although encouraging, ADAS still have significant challenges when it comes to driving in crowded areas, where the interaction with Vulnerable Road Users (VRUs) becomes of paramount importance. In these situations, forecasting the future paths and movements of VRUs is a vital task in producing collision-free paths. Even a small misestimation of VRUs position can cause a serious hazard, as this might place the pedestrian just inside or outside the driving corridor [2]. However, path prediction of VRUs is extremely challenging, which is due to the high variability of their movement patterns. This is even worse when it comes to pedestrians, as they often are not using specific traffic lanes, can abruptly change their motion state (e. g., starting, stopping, bending in) and do not actively communicate their intention by turn indicators or brake lights [3]. The subject of pedestrian's path prediction has been extensively addressed in the literature with a wide range of applications. Nonetheless, the proposed methods have never been well categorized in an academically structured fashion. Recently, Ridel et al. [4] provided a comprehensive literature review on the prediction of pedestrian behavior in urban scenarios. However, their study is mostly limited to pedestrian intention estimation. In this paper, the evolution of all path prediction models from traditional methods to state-of-the-art, deep-learning ones are reviewed and individuated from multiple perspectives.

The key to successful path prediction highly relies on an effective understanding of the target's dynamics model. Dynamics models describe the evolution of the target states with respect to time. One can say without exaggeration that a good dynamics model is worth a thousand pieces of data [5]. However, sometimes it is quite infeasible to reach a closed-form dynamics model, which is fully responsive to all target behaviors. This is especially the case with pedestrians. The complexity of human behavior originate from various factors and most of the time a simple dynamics model cannot cover all of them. For the past twenty years, numerous papers have been published, aiming at discovering the most influencing factors and proposing more thorough models. Analyzing the published papers shows that throughout time, the authors have taken three different approaches. In the first approach, which is referred to as *interaction-based approach*, the authors mainly address the interactions between the pedestrians themselves as well as the environment, such as a set of static objects in the scene. In this approach, these interactions are considered as the main influencing factor and the goal is to develop a model, which better covers the interactions. In the second approach. i.e. *path-planning approach*, it is assumed that the pedestrian movement is highly affected by the destinations they are headed to. The problem of path prediction is normally turned into a path-planning one and the authors try to either estimate the destinations or recover the policy and reward function. Finally, in the third approach, which is referred to as the *intention-based approach*, the focus is on discovering the next movement patterns of VRUs in the form of intention estimation.

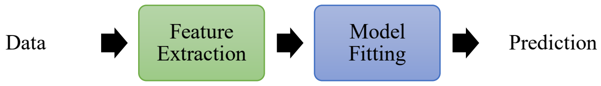

Figure 1. Prediction procedure in traditional methods

Generally, the prediction process in these approaches consists of two fundamental steps: 1) *feature extraction* and 2) *model fitting*. Features can be defined as a set of influencing factors. They could be some motion cues of different body parts or could be the proximity of different pedestrians to a target, or all other effective factors that can be derived from the data. As depicted in Fig. 1, given suitable data, first these features are extracted and then a model is fitted based on these extracted features.

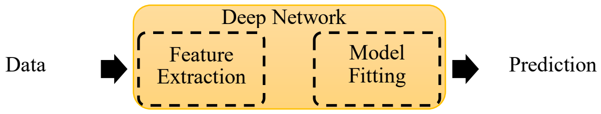

Figure 2. Prediction procedure in deep-learning methods

In recent years, development of deep-learning methods has allowed the removal of manual feature extraction and model

Ariyan Bighashdel (a.bighashdel@tue.nl) and Gijs Dubbelman (g.dubbelman@tue.nl) are with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. This research has received funding from ECSEL JU in collaboration with the European Union’s 2020 Framework Programme and National Authorities, under grant agreement no. 783190.

978-1-5386-7024-8/19/$31.00 ©2019 IEEE

fitting (see Fig. 2). In deep-learning methods, all the processes are done automatically in a single, deep neural network, which makes the prediction model more powerful by capturing more hidden features and fitting stronger models.

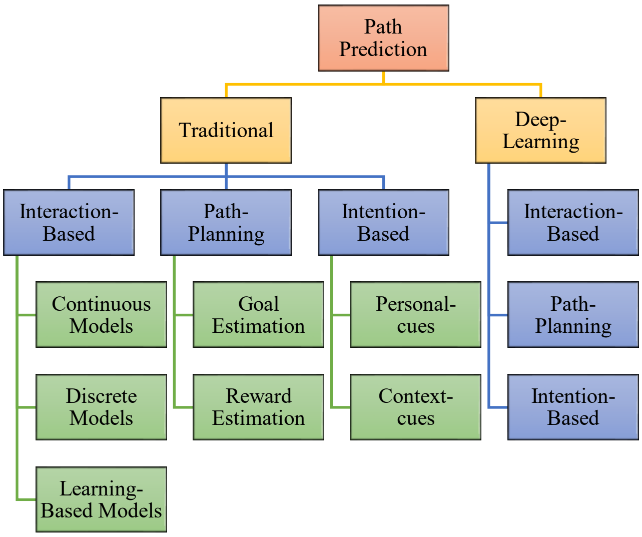

Figure 3. Overview of path-prediction approaches

Fig. 3 shows a general overview of recently proposed methods for VRUs path prediction applications. Apart from the three previously mentioned approaches, the methods can be divided from a broader point of view into traditional and deep-learning methods. The traditional methods can be further divided into some subdivisions based on the type of handcrafted features that are extracted, or the manually-designed models that are fitted. The remainder of the paper is structured as follows. In Section II, the different approaches in traditional methods are thoroughly reviewed. The deep-learning models are addressed in Section III. Finally, Section VI is devoted to discussion.

II. Traditional Methods

A. Interaction-Based Approach

One of the most influencing factors on the human movement pattern is the interaction, which can be due to proximity of other humans or some objects in the scene. In Interaction-based approach, the main emphasis is on accurately modeling of these interactions. The proposed interaction models are reviewed in this section.

Continuous models: The pioneering work of Helbing and Molnar [6], is one of the first approaches in modelling the interaction patterns among pedestrians. The model, called Social Force Model (SFM), describes the pedestrian motions as if they subject to some social forces. These social forces are the internal motivations of the individuals to perform certain actions [6]. Later, Yamaguchi et al. [7] improved the original SFM by exploiting more behavioral factors such as damping and collision, and tried to incorporate them into the prediction model. Based on SFM, Stefano et al. [8] developed a dynamic social behavior based on trajectory avoidance. The choice of parameters in the SFM, has also been shown to be of primary importance for oscillation avoidance in the pedestrians movements [9]. The success of the SFM, convinced the authors to extend the model by fitting the parameters of the force functions to some observed crowd behaviors such as evacuation scenarios and abnormal behaviors [10], [11] and even for other traffic participants like the dynamics of interacting vehicles [12]. Similarly to SFM, Treuille et al. [13] used continuum dynamics to model the moving crowds with the help of a dynamic potential field. Another line of work includes the use of some well-engineered features and attributes to improve prediction. Alahi et al. [14] proposed a feature descriptor, termed as social affinity map, to capture the behavioral factors of neighboring pedestrians. On the other hand, Yi et al. [15] proposed the use of personal attributes for better improvement of forecasting in dense crowds, with the presence of some stationary crowd groups.

Discrete models: Unlike the SFM, which models the human movement in a continuous domain, some models are proposed which describe the human movement in a discrete environment [16]. Using the theory of *Cellular Automaton* (CA), Schadschneider [17] modeled the dynamics of pedestrian in a discretized environment. They described the local movements of the pedestrians with a matrix of preferences which contains the probabilities for a move, related to the preferred walking direction and speed, toward the adjacent directions [18]. Other variants of CA have also been developed by some authors during the years [18]–[21]. In contrary to CA which uses static discretization of the environment, Bierlaire et al. [16] proposed an agent-based method, based on dynamic discretization. In the model, called Dicrete Choice Model (DCM), the behavior of the agents are modeled as a sequence of specific choices in a discrete choice framework. DCM has been employed in various number of works [22]–[28].

Learning-based models: One of the regular ways of path prediction for pedestrians is through learning their motion patterns. It is based on the idea that in an area, the pedestrians tend to follow typical motion patterns that depend on their nature and the structure of the area [29]–[35]. Given a set of trajectories, the proposed learning-based approaches try to learn the motion patterns by various traditional techniques like hierarchical clustering (Large et al. [29]), K-means based methods (Bennewitz et al. [30], Hu et al. [31], and Fan et al. [35]), Gaussian mixture model (Tay and Laugier [36]), mean-shift (Ellis et al. [37]) and kriging-based models (Kawamoto et al. [38], [39]). Some other learning methods have also been proposed in the literature, which are reviewed briefly. Chen et al. [32], and Chen and Yung [34], incorporated velocity profiles and heading angles, along with the positions and clustered the observed trajectories using constraints gravitational method. Goldhammer et al. [40] presented a learning approach based on the artificial neural network. Given a number of trajectory points and time steps prior to the current time, they trained a Multi-Layer Perceptron (MLP) to predict future trajectory points. Later, they improved the model by means of polynomial approximation and considering the velocity profiles [41]. Rather than pure offline training, Vasquez et al. [33] proposed a continuously learning framework based on growing neural gas algorithm. Chen et al. [42] combined Markovian-based and clustering-based approaches in developing a dictionary learning algorithm, called augmented semi- non-negative sparse coding. Their model later improved by Japuria et al. [43] by incorporating semantic features from the environment for more confident prediction, including relative distances to curbside. In order to explicitly model the interactions among the pedestrians, Trautman and Krause [44] developed an Interacting Gaussian process model. The proposed statistical model is based on dependent output Gaussian processes, which is also employed in [45]. Li et al. [46] extracted the motion features related to the target pedestrian and the nearby ones, and applied an auto-encoder in order to strongly consider the interactions. The learned features are then fed into a Gaussian mixture model for prediction.

B. Path-Planning Approach

The other approach can be considered as a path-planning method, which is inspired by the approaches heavily used in the robotics field. Given an environment, these approaches, try to either estimate the destinations for the agents (goal-estimation) or recover a reward function that induces the demonstrated behaviors (reward-estimation).

Goal-estimation: In an early work, Bruce and Gordon [47] tried to estimate the destinations using clustering techniques and modeled the pedestrian motion model as a Markov Decision Process (MDP) to predict how a person moves from his/her location to estimated goals. Madrigal and Hayet [48] also tried to determine the goals by clustering the trajectories key points. Ikeda et al. [49] introduced the concept of sub-goals and they assumed that pedestrians tend to segment their path to final destinations by passing thorough some common sub-goals. Rehder and Kloeden [50], also used a path-planning approach to predict the future trajectories of pedestrians. They modeled the goal destinations of the pedestrians as a Gaussian Mixture model and tried to improve estimated destinations by employing particle filters.

Reward-estimation: Another idea, related to imitation learning, is to structure the space of learned policies to be the solutions of path planning [51]. The idea is based on the assumption that people plan paths in an optimal way or in other words, the agents act to optimize an unknown reward function [51], [52]. Therefore, the problem can be turned into finding the reward function that makes the demonstrated behavior of the agents, optimal. In general, the people are considered as path planners, however, contrary to traditional planners where the problem is finding the optimal trajectory, here the problem is finding the cost function that best explains the previously observed trajectories [53]. Ziebart et al. [51], [53], employed the principals of maximum entropy and developed a model called maximum entropy inverse optimal control or inverse reinforcement learning (IRL). In their model, it was assumed that the agents have full access to all factors that influence the decisions. Henry et al. [52] employed the same method and at the same time tried to estimate some factors like density and flow of people. Later, Kitani et al. [54] improved the model by incorporating vison-based, scene features and noisy tracker observations to forecast activities and destinations. Other variants of the maximum entropy IRL can be seen in [55]–[57].

C. Intention-Based Approach

In the development of collision avoidance systems, one of the key issues is the ability of intention estimation for the pedestrians approaching the curbside. In this regard, various methods have been proposed in the literature, aiming to answer this crucial question: “Will the pedestrian cross?” [58], [59]. In general, the intention prediction methods are based on two essential assumptions for pedestrians; first, pedestrian behavior is intention-driven and second, intention can be derived from specific features [60]. A wide range of features and information can be extracted which can be originated from the pedestrian itself (personal-cues) or the context (context-cues).

Personal-cues: Schneider and Gavrila [61] proposed a model in which the pedestrian is treated as a point mass and limited the features to positioning information. They employed multiple Bayesian filters and tried to predict different motion modes by use of Interacting Multiple Model (IMM). Apart from the use of positional cues, Keller et al. [58] and Keller and Gavrila [59] extracted motion features from dense optical flow and used them in a probabilistic trajectory matching. They also employed two Gaussian Process Dynamics Models to make predictions for walking and stopping motions. Other examples of motion features can be seen in the works of Koehler et al. [62]–[64] where the motion features are extracted using contour based HOG-like descriptors and intention classification is done by the help of a linear Support Vector Machine (SVM). Moreover, the direction in which the pedestrian is facing or in general, the head pose has been used in various works as an indicator of interesting areas as well as situational awareness of the pedestrian [65]–[68]. Gandhi et al. [66] used HOG features and SVM classifier to infer the future orientations and modeled the orientation transitions over time by a hidden Markov model. Schulz and Stiefelhagen [67], [68] proposed a model, based on latent-dynamic conditional random field for intention recognition. The model integrates the pedestrian dynamics and situational awareness using observations from a stereo-vison system for pedestrian detection and head pose estimation. Apart from head pose, body-pose estimation has also gained a great deal of attention in literature [69]–[73], where the authors use Gaussian process dynamical models to infer the three-dimensional pedestrian body language.

Context-cues: It has been shown that taking into account the context information, like the criticality of the situation and the environment layout, may yield to better estimation of the pedestrian intention [2], [74]. Kooij et al. [2] incorporated the pedestrian situational awareness, situation criticality and spatial layout of the environment as the latent states on top of a switching linear dynamical system to predict the changes in motion modes of the pedestrians. Bonnin et al. [74] fused various models to predict crossing intentions via contextual information such as lateral distances and time to reach goals. In general, context cues have been incorporated in various works [65], [75], [76] and sometimes are combined with motion features [77].

III. Deep-Learning Methods

A. Interaction-Based Approach

Before the dawn of deep learning, the traditional interaction methods were used extensively in human trajectory prediction. After the growth of data-driven approaches, the question was raised that whether the handcrafted features and models are able to capture all the complexities regarding human movements or not. Instead of learning, the traditional methods, either manually design the behavioral functions with hand-tuned settings, or use shallow learning-based models to describe the dynamics of pedestrians. Consequently, these methods only cover the dominant interactions between agents and the surrounding environment, which makes them unreliable in movement prediction in more crowded scenarios. However, it should be borne in mind that mere use of deep-learning methods does not guarantee correct interaction modeling. The deep interaction models use various mechanisms for better capturing the interactions. These mechanisms are summarized in Table 1 and are discussed in this section.

In an early work, Yi et al. [78] proposed a 3-stage deep Convolutional Neural Network (CNN), called "Behavior-CNN", to model pedestrians behaviors in crowded scenes. In their model, they used the concept of trajectory volumes and a walking behavior encoding scheme, in order to consider the human-human interactions. Alahi et al. [79] turned the problem of trajectory prediction to a sequence generation task, using Recurrent Neural Network (RNN). In order to capture interactions between different sequences, they introduce a Long Short-Term Memory (LSTM) network with a social pooling layer (referred to as "Social-LSTM"), which allows different sequences to share their features in their hidden states. The pooling layers also preserve the spatial information through grid-based pooling which makes the network immune in case of prohibitively high number of neighboring sequences. Cheng et al. [80] adopted the social pooling strategy of Alahi et al. [79] and applied it to a two-dimensional Grid LSTM [81] and called the model "Social-Grid-LSTM". The Social-LSTM was later improved by Bartoli et al. [82] with adding a context pooling layer. The model, called "Context-Aware Social-LSTM", also takes into account the positions of static objects in the environment.

Su et al. [83] proposed an Encoder-Decoder framework using LSTM for path prediction. In order to capture the coherent motion phenomena, they introduced a coherent LSTM (C-LSTM) in which the memory unit is updated by incorporation its own states as well as its neighboring agents with a coherent regularization. Later, they proposed a social-aware recurrent Gaussian process model (SRGP) to better account for motion uncertainties [84]. Using velocity fields, they employed a LSTM with a social-aware gate to capture the dynamics of the pedestrian and the neighboring crowd. The hidden features derived from the social-aware LSTM are fed into a deep Gaussian process, which is designed for explicit modeling of the prediction uncertainties. The problems of prediction uncertainties and multimodality of human behaviors have also been addressed in the work of Gupta et al. [85]. To solve the problem, they proposed a model based on Generative Adversarial Networks (GANs). The model, called "Social GAN", uses a global pooling mechanism, which encodes the interactions between all people that are involved in a scene.

Fernando at al. [86] proposed a LSTM Encoder-Decoder framework which utilizes both soft- and hard-wired attention for path prediction of pedestrians in crowded environments . With the help of a soft attention context, the network is able to focus different levels of attention towards parts of the input trajectory sequence and using hard-wired attention, makes the model applicable in combining the hidden states of neighboring sequences. The required attention for the neighbors are approximated through the hard-wired attention weights, which are designed to incorporate the notation of the distance of the neighbor pedestrians to the pedestrian of the interest. By introducing the structured LSTM (St-LSTM) cells, later they improved their model for considering long-term scene context [87]. They proposed a structured memory network (SMN) which hierarchically stores and reads the short-term contexts for long-term usage.

Vemula et al. [88] proposed a social attention RNN model using structural-LSTM [89], which captures the relative importance of neighboring pedestrians, irrespective of their proximities. They modeled the human-human interactions using a soft attention model over all humans in the crowd. Recently, Haddid et al. [90] improved the work of Vemula et al. [90] for better modeling of multiple trajectories correlations over space-time dimensions by incorporating the interactions between Human-Human-Obstacles (H-H-O) using a spatio-temporal attention module.

Xu et al. [91] introduced a Crowd-Interaction Deep Neural Network (CIDNN) for displacement prediction in crowded environments. Using a kernel trick, they proposed a MLP framework for automatic learning of the spatial affinity measurement. Rather than distance-based affinity measurement, they could weigh the influence level of different pedestrians to the target pedestrian.

Xue et al. [92] proposed a two-stage framework to predict multiple trajectories with different probabilities toward different regions. They employed a Bidirectional LSTM (B-LSTM) with two convolutional and pooling layers for dealing with classification of the trajectory coordinates, which is then topped with a softmax layer to output the probability distribution on different sub-regions. Finally, several LSTMs with encoder-decoder architectures were employed to predict the future trajectories based on different route classes. In another work, they proposed a hierarchical LSTM-based network with three LSTM encoders, namely person-scale, social-scale, and scene-scale encoders, to take into account the different types of interactions [93]. In the model, called "SS-LSTM", a person-scale encoder uses the information regarding the pedestrian's past trajectory and a social-scale encoder captures the influence from other pedestrians in the neighborhood by building an occupancy map matrix for each target pedestrian. Finally, a scene-scale encoder, combined with a CNN, extracts the features regarding the scene layout. More recently, they suggested that considering the relationship between location and velocity information, may lead to better results in path prediction [94]. Accordingly, they proposed a joint location-velocity, attention LSTM (LVA-LSTM) which uses a tweak module to build a relationship between location and velocity.

Similar to the work of Xue et al. [93], Pfeiffer et al. [95] proposed a network for better incorporation of static objects and surrounding pedestrians for trajectory forecasting. The proposed model, here referred to as Static-LSTM, employs some occupancy grids and a pre-trained CNN to capture the information regarding neighboring pedestrians and static obstacles.

TABLE I. SUMMARY OF DEEP-LEARNING INTERACTION MODELS

Name	Deep Network	Type of Interaction	Interaction Mechanism
Behavior-CNN [78]	CNN	Human-Human	Displacement Volume encoder
Social-LSTM [79]	LSTM	Human-Human	Social pooling layer
Social-Grid-LSTM [80]	Grid-LSTM	Human-Human	Social pooling layer
Context-Aware Social-LSTM [82]	LSTM	Human-Human & Human-Scene	Social-context pooling layer
C-LSTM [83]	LSTM	Human-Human	Coherent memory update
SRGP [84]	LSTM	Human-Human	Social-aware gate
Social GAN [85]	GAN-LSTM	Human-Human	Global pooling mechanism
Soft+ hardwired attention [86]	LSTM-Attention	Human-Human	Soft and hard-wired attention
St-LSTM [87]	LSTM-Attention	Human-Human	Soft and hard-wired attention & SMN
Social-attention [88]	Structural-LSTM	Human-Human	Soft attention
H-H-O [90]	Structural-LSTM	Human-Human & Human-Scene	Spatio-temporal attention
CIDNN [91]	LSTM	Human-Human	MLP
B-LSTM [92]	Bidirectional-LSTM	Human-Scene	Multi-trajectory prediction
SS-LSTM [93]	LSTM, CNN	Human-Human & Human-Scene	occupancy map matrix & CNN feature extraction
LVA-LSTM [94]	LSTM-Attention	Human-Human	location-velocity attention mechanism
Static-LSTM [95]	LSTM, CNN	Human-Human & Human-Scene	angular pedestrian grid, CNN
SR-LSTM [96]	LSTM-Attention	Human-Human	Refinement module and attention mechanism

Zhang et al. [96] proposed a LSTM-based network which accounts for the current intention of the neighbors, rather than only relying on their previous states. The network, called “state-refinemnt LSTM (SR-LSTM)”, employs a refinement module which activates the utilization of the neigbors current intention and iteratively refines the states of all paricipants. To adaptively focus on the most useful information of the neighbors, they introduced a social-aware information selection mechanism which consists of an element-wise motion gate and a pedestrian-wise attention mechanism. More recently, Shi et al. [97] proposed an LSTM-based prediction model for extremely crowded scenarios using the concept of the relativity of motion trajectory. In their model, both motion trajectories and human interaction are represented with relative motions and the interactions are dynamically modeled by incorporating an attention-weighted pooling.

B. Path-Planning Approach

The traditional path-planning methods, build the visual representations based on handcrafted features and try to model the context in a relatively simple and shallow way [98]. In more complex and crowded scenarios, abundant semantic information and deep feature extraction are needed for better visual representation. This can be accomplished with the help of deep networks that improve the task of path prediction by deep, scene analyses and powerful automatic feature extractions. In this section, various deep path-planning methods are briefly explained.

Huang et al. [98] proposed a two-stage, deep-learning framework for visual path prediction task. In the first stage, they perform deep feature learning for visual representation in conjunction with spatio-temporal context modeling, then in the second stage, they use a unified path-planning scheme to make path prediction based on the analytical results of the context model. For the context modeling, they employed two CNNs, one for generating the spatial matching cost, which shows the structure relationship between object and the scene, and the other one for building the orientation cost, showing the moving direction of the object. Having the total cost, they turned the problem to an optimization one, aiming to find the optimal path with the lowest cost. The proposed CNN mechanism was also employed by Varshneya and Srinivasaraghavan [99], with some modifications.

Rehder at al. [60] targeted both destination estimation and trajectory prediction using a deep neural network. The whole network contains three parts. In the first part, a feature vector output from a CNN, concatenated with position vectors, have been used as the input to a Recurrent Mixture Density Network (RMDN) with a LSTM cell. With a constant additive layer, the output of the RMDN is a discretized probability distribution map of the position and orientation. In the next step, the output of the RMDN are fed to a fully convolutional network, which is trained by IRL policy prediction. Given the policy, in the final step, two planning techniques namely MDP as the optimal path generator, and Forward-Backward algorithm as a sub-optimal one, are employed to do the planning part.

Saleh et al. [100] proposed a data-driven, two-stage framework based on IRL and Bidirectional-LSTM. In the first stage, they used both contextual information and demonstrated trajectories of pedestrians and recovered the reward map using maximum entropy IRL [51], [53]. Then, in the second stage, the reward map along with the trajectories are fed to a bidirectional LSTM [100] which is topped with a mixture density network to account for the uncertainty of pedestrian trajectories in urban traffic environments. The output of the network, which is called B-LSTM-MDN, is the probability distribution of predicted trajectories.

Lee et al. [101] introduced a deep stochastic encoder-decoder model based on IRL framework for path prediction. They formulated the problem as an optimization process, through a two-stage framework. In the first stage, a diverse set of hypothetical future prediction samples are obtained employing a conditional variational auto-encoder [102] with gated recurrent units structures. In the second stage, the samples are ranked and refined in a RNN scoring-regression module. Given the samples, the scoring module is processed considering the 1) motion context, based on the samples motions, 2) semantic scene context, using CNN features, and 3) the interaction between agents, based on social pooling layer [79]. A regression-based refinement module is also employed which using an iterative feedback mechanism to further boost the prediction accuracy.

Zou et al. [103] proposed a three-stage framework to mimic the decision-making process of pedestrians in order to achieve a better path planning procedure. In the first stage, which is called the generator (policy) stage, the future actions (trajectories) are generated through an encoder-decoder RNN. During the action generation, three social-aware components are added to the network: 1) Intention with latent code, which ensures that pedestrians with similar observed trajectories do not necessarily take similar future paths, 2) Collision avoidance, inspired by optimal reciprocal collision avoidance [104] in a stop/go framework, and 3) social interaction as proposed by Alahi et al. [79]. In the second stage, the outputs of the generator that are combined with the observed trajectories, are scored in the discriminator with a RNN network. Inspired by the work of [105], the whole network is optimized in IRL framework. Finally, in the third stage, a variational posterior estimator [106] is employed to establish the relationship between the latent code and the policy.

C. Intention-Based Approach

One thing that can be noticed from traditional, intention-based methods is that in all of them, the model is forced to capture some specific features, such as the head orientation or some relative distances, and intention prediction is done based on these predefined features. In this way, some hidden features may be easily ignored. In deep-learning methods on the other hand, the model has the freedom to capture as much features as needed to better infer the intention. In this regard, various deep-learning frameworks have been proposed, aiming to predict the intention of the pedestrians in a data-driven manner.

Völz at al. [107] used both data-driven features as well as some handcrafted ones at the same time and proposed a deep network for intention analysis and prediction of the pedestrians approaching crosswalks. First, they employed a CNN to produce some feature vectors from images in several time steps. Then, in a concatenation stage, some handcrafted features, like distances to curb and crosswalk and some relative velocities, are added and finally, all the features are fed to a deep fully connected neural network in which the output is the movement action of the pedestrian (crossing, not crossing). They also employed a LSTM to do the path prediction for the specific action.

Fang et al. [108] presented a date-driven model to detect pedestrian intention using skeleton features. After detection of the bounding boxes, they employed a two-branch, multi-stage CNN to perform the skeleton fitting. Based on the fitted skeleton, the features are extracted for all the past frames and combined with each other. Finally, two binary classifiers namely Random Forest and SVM are used for intention prediction of the pedestrians. In another work, Saleh et al. [109] proposed an end-to-end approach for intent-prediction of VRUs. They formulated the problem as time-series prediction and tried to predict the future position of the objects solely based on their motion trajectories. For their network, they utilized a RNN architecture and formed a Stacked LSTM network.

IV. Discussion

In this paper, a comprehensive and contemporary overview of path prediction models for VRUs is presented. Considering the evolution of models, one can easily observe that the models are gradually getting more and more complex. If once a relatively simple, physics-based model was promising the best results, now there are some hesitations in calling deep learning-based methods as the state-of-the-art. This is due to this important fact that humans are a complex being and throughout time, more aspects related to its motion and path are discovered. The main difference that can be seen in the proposed methods is that in traditional approaches, researchers first tried to find a reason for specific human behaviors, and then propose a model that captures those kinds of behaviors. This is especially the case for the intention-based approaches, in which it was seen how researchers tried to discover various cues as the reasons for sudden behaviors. On the other hand, the introduction of deep leaning to human path prediction, created the opportunity for researchers to be able to identify models that well respond to human behavior regardless of the underlying behavioral reasons. However, deep learning itself does not guarantee to be responsive to all the complexities of human behavior. A combination of handcrafted models and deep-networks should be considered in order to cover the complete range of human behaviors. The diversity of proposed models and mechanisms suggests that the current understanding of the pedestrian's path is still far from the fully understood and a perfectly detailed prediction model requires further studies.

References

[1] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transp. Res. Part A Policy Pract.*, vol. 77, pp. 167–181, 2015.

[2] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-Based Pedestrian Path Prediction," 2014.

[3] M. Goldhammer, S. Köhler, S. Zernetsch, K. Doll, B. Sick, and K. Dietmayer, "Intentions of Vulnerable Road Users-Detection and Forecasting by Means of Machine Learning," *arXiv Prepr. arXiv1803.03577*, 2018.

[4] D. Ridel, E. Rehder, M. Lauer, C. Stiller, and D. Wolf, "A Literature Review on the Prediction of Pedestrian Behavior in Urban Scenarios," in *ITSC*, 2018, pp. 3105–3112.

[5] X. R. LI and V. P. JILKOV, "Survey of Maneuvering Target Tracking. Part I: Dynamic Models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, p. 1333, 2003.

[6] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Phys. Rev. E*, vol. 51, no. 5, p. 4282, 1995.

[7] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg, "Who are you with and where are you going?," in *CVPR 2011*, 2011, pp. 1345–1352.

[8] P. Stefano, E. Andreas, S. Konrad, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *ICCV*, 2009.

[9] T. Kretz, "On oscillations in the social force model," *Phys. A Stat. Mech. its Appl.*, vol. 438, pp. 272–285, 2015.

[10] A. Johansson, D. Helbing, and P. K. Shukla, "Specification of the social force pedestrian model by evolutionary adjustment to video tracking data," *Adv. Complex Syst.*, vol. 10, pp. 271–288, 2007.

[11] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009*, pp. 935–942.

[12] D. Helbing and B. Tilch, "Generalized force model of traffic dynamics," *Phys. Rev. E*, vol. 58, no. 1, p. 133, 1998.

[13] A. Treuille, S. Cooper, and Z. Popović, "Continuum crowds," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1160–1168, 2006.

[14] A. Alahi, V. Ramanathan, and L. Fei-Fei, "Socially-aware large-scale crowd forecasting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2203–2210.

[15] S. Yi, H. Li, and X. Wang, "Understanding pedestrian behaviors from stationary crowd groups," in *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition, 2015, pp. 3488–3496.

[16] M. Bierlaire, G. Antonini, and M. Weber, “Behavioral dynamics for pedestrians,” 2003.

[17] A. Schadschneider, “Cellular automaton approach to pedestrian dynamics-theory,” *arXiv Prepr. cond-mat/0112117*, 2001.

[18] P. Shao, “A more realistic simulation of pedestrian based on cellular automata,” in *Open-source Software for Scientific Computation (OSSC), 2009 IEEE International Workshop on*, 2009, pp. 24–29.

[19] J. Wąs, B. Gudowski, and P. J. Matuszyk, “New cellular automata model of pedestrian representation,” in *International Conference on Cellular Automata*, 2006, pp. 724–727.

[20] K. Yamamoto, S. Kokubo, and K. Nishinari, “Simulation for pedestrian dynamics by real-coded cellular automata (RCA),” *Phys. A Stat. Mech. its Appl.*, vol. 379, no. 2, pp. 654–660, 2007.

[21] X. Li, X. Yan, X. Li, and J. Wang, “Using cellular automata to investigate pedestrian conflicts with vehicles in crosswalk at signalized intersection,” *Discret. Dyn. Nat. Soc.*, vol. 2012, 2012.

[22] G. Antonini, S. Venegas, J.-P. Thiran, and M. Bierlaire, “A discrete choice pedestrian behavior model for pedestrian detection in visual tracking systems,” 2004.

[23] S. Venegas-Martinez, G. Antonini, J. P. Thiran, and M. Bierlaire, “Automatic pedestrian tracking using discrete choice models and image correlation techniques,” in *International Workshop on Machine Learning for Multimodal Interaction*, 2004, pp. 341–348.

[24] S. Venegas, G. Antonini, J.-P. Thiran, and M. Bierlaire, “Bayesian integration of a discrete choice pedestrian behavioral model and image correlation techniques for automatic multiobject tracking,” in *Image Processing, 2004. ICIP’04*, vol. 2, pp. 1037–1040.

[25] G. Antonini, S. V. Martinez, and J.-P. Thiran, “A model-based framework for automatic tracking and counting of pedestrians in video sequences,” in *International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2005.

[26] G. Antonini, “A discrete choice modeling framework for pedestrian walking behavior with application to human tracking in video sequences,” 2005.

[27] G. Antonini, S. V. Martinez, M. Bierlaire, and J. P. Thiran, “Behavioral priors for detection and tracking of pedestrians in video sequences,” *Int. J. Comput. Vis.*, vol. 69, no. 2, pp. 159–180, 2006.

[28] G. Antonini, M. Bierlaire, and M. Weber, “Discrete choice models of pedestrian walking behavior,” *Transp. Res. Part B Methodol.*, vol. 40, no. 8, pp. 667–687, 2006.

[29] F. Large, D. Vasquez, T. Fraichard, and C. Laugier, “Avoiding cars and pedestrians using velocity obstacles and motion prediction,” in *Intelligent Vehicles Symposium, 2004 IEEE*, 2004, pp. 375–379.

[30] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, “Learning Motion Patterns of People for Compliant Robot Motion,” 2005.

[31] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, “A system for learning statistical motion patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, 2006.

[32] Z. Chen, D. C. K. Ngai, and N. H. C. Yung, “Pedestrian behavior prediction based on motion patterns for vehicle-to-pedestrian collision avoidance,” in *ITSC*, 2008, pp. 316–321.

[33] D. Vasquez, T. Fraichard, O. Aycard, and C. Laugier, “Intentional motion on-line learning and prediction,” *Mach. Vis. Appl.*, vol. 19, no. 5–6, pp. 411–425, 2008.

[34] Z. Chen and N. H. C. Yung, “Improved multi-level pedestrian behavior prediction based on matching with classified motion patterns,” in *ITSC*, 2009, pp. 1–6.

[35] Z. Fan, Z. Wang, J. Cui, F. Davoine, H. Zhao, and H. Zha, “Monocular pedestrian tracking from a moving vehicle,” in *Asian Conference on Computer Vision*, 2012, pp. 335–346.

[36] M. K. C. Tay and C. Laugier, “Modelling smooth paths using gaussian processes,” in *Field and Service Robotics*, 2008, pp. 381–390.

[37] D. Ellis, E. Sommerlade, and I. Reid, “Modelling pedestrian trajectory patterns with gaussian processes,” in *2009 IEEE ICCV Workshops*, 2009, pp. 1229–1234.

[38] K. Kawamoto, Y. Tomura, and K. Okamoto, “Learning pedestrian dynamics with kriging,” in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–4.

[39] K. Kawamoto, Y. Tomura, and K. Okamoto, “Kriging-based prediction and interpolation for modeling pedestrian dynamics,” in *Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS),* 2017, pp. 1–4.

[40] M. Goldhammer, K. Doll, U. Brunsmann, A. Gensler, and B. Sick, “Pedestrian’s trajectory forecast in public traffic with artificial neural networks,” in *Pattern Recognition (ICPR)*, pp. 4110–4115.

[41] M. Goldhammer, S. Köhler, K. Doll, and B. Sick, “Camera based pedestrian path prediction by means of polynomial least-squares approximation and multilayer perceptron neural networks,” in *SAI Intelligent Systems Conference (IntelliSys), 2015*, 2015, pp. 390–399.

[42] Y. F. Chen, M. Liu, and J. P. How, “Augmented dictionary learning for motion prediction,” in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, 2016, pp. 2527–2534.

[43] N. Japuria, G. Habibi, and J. P. How, “CASNSC: A context-based approach for accurate pedestrian motion prediction at intersections,” 2017.

[44] P. Trautman and A. Krause, “Unfreezing the robot: Navigation in dense, interacting crowds,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 797–803.

[45] P. Trautman, J. Ma, R. M. Murray, and A. Krause, “Robot navigation in dense human crowds: Statistical models and experimental studies of human--robot cooperation,” *Int. J. Rob. Res.*, vol. 34, no. 3, pp. 335–356, 2015.

[46] Y. Li, M. L. Mekhalfi, M. M. Al Rahhal, E. Othman, and H. Dhahri, “Encoding Motion Cues for Pedestrian Path Prediction in Dense Crowd Scenarios,” *IEEE Access*, vol. 5, pp. 24368–24375, 2017.

[47] A. Bruce and G. Gordon, “Better motion prediction for people-tracking,” in *Proc. of the Int. Conf. on Robotics & Automation (ICRA), Barcelona, Spain*, 2004.

[48] F. Madrigal and J.-B. Hayet, “Goal-oriented visual tracking of pedestrians with motion priors in semi-crowded scenes,” in *Robotics and Automation (ICRA),* 2015, pp. 720–725.

[49] T. Ikeda, Y. Chigodo, D. Rea, F. Zanlungo, M. Shiomi, and T. Kanda, “Modeling and prediction of pedestrian behavior based on the sub-goal concept,” *Robotics*, vol. 10, 2013.

[50] E. Rehder and H. Kloeden, “Goal-directed pedestrian prediction,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 50–58.

[51] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, “Maximum entropy inverse reinforcement learning.,” in *Aaai*, 2008, vol. 8, pp. 1433–1438.

[52] P. Henry, C. Vollmer, B. Ferris, and D. Fox, “Learning to navigate through crowded environments,” in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 981–986.

[53] B. D. Ziebart *et al.*, “Planning-based prediction for pedestrians,” in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 3931–3936.

[54] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, “Activity forecasting,” in *European Conference on Computer Vision*, 2012, pp. 201–214.

[55] M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard, “Feature-Based Prediction of Trajectories for Socially Compliant Navigation.,” in *Robotics: science and systems*, 2012.

[56] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, “Socially compliant mobile robot navigation via inverse reinforcement learning,” *Int. J. Rob. Res.*, vol. 35, no. 11, pp. 1289–1307, 2016.

[57] H. Kretzschmar, M. Kuderer, and W. Burgard, “Learning to predict trajectories of cooperatively navigating agents,” in *2014 IEEE ICRA*, 2014, pp. 4015–4020.

[58] C. G. Keller, C. Hermes, and D. M. Gavrila, “Will the pedestrian cross? probabilistic path prediction based on learned motion features,” in *Joint Pattern Recognition Symposium*, 2011, pp. 386–395.

[59] C. G. Keller and D. M. Gavrila, “Will the pedestrian cross? A study on pedestrian path prediction,” *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 494–506, 2014.

[60] E. Rehder, F. Wirth, M. Lauer, and C. Stiller, “Pedestrian prediction by planning using deep neural networks,” *arXiv Prepr. arXiv1706.05904*, 2017.

[61] N. Schneider and D. M. Gavrila, “Pedestrian path prediction with recursive Bayesian filters: A comparative study,” in *German Conference on Pattern Recognition*, 2013, pp. 174–183.

[62] S. Koehler *et al.*, “Stationary detection of the pedestrian? s intention at intersections,” *IEEE Intell. Transp. Syst. Mag.*, vol. 5, no. 4, pp. 87–99, 2013.

[63] S. Köhler, M. Goldhammer, K. Zindler, K. Doll, and K. Dietmeyer, “Stereo-vision-based pedestrian’s intention detection in a moving vehicle,” in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 2317–2322.

[64] S. Köhler, M. Goldhammer, S. Bauer, K. Doll, U. Brunsmann, and K. Dietmayer, “Early detection of the pedestrian’s intention to cross the street,” in *ITSC*, 2012, pp. 1759–1764.

[65] F. Schneemann and P. Heinemann, “Context-based detection of pedestrian crossing intention for autonomous driving in urban environments,” in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, 2016, pp. 2243–2248.

[66] T. Gandhi and M. M. Trivedi, “Image based estimation of pedestrian orientation for improving path prediction,” in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 506–511.

[67] A. T. Schulz and R. Stiefelhagen, “A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction,” in *ITSC*, 2015, pp. 173–178.

[68] A. T. Schulz and R. Stiefelhagen, “Pedestrian intention recognition using latent-dynamic conditional random fields,” in *Intelligent Vehicles Symposium (IV), 2015 IEEE*, 2015, pp. 622–627.

[69] R. Q. Mínguez, I. P. Alonso, D. Fernández-Llorca, and M. Á. Sotelo, “Pedestrian Path, Pose, and Intention Prediction Through Gaussian Process Dynamical Models and Pedestrian Activity Recognition,” *IEEE Trans. Intell. Transp. Syst.*, 2018.

[70] R. Quintero, J. Almeida, D. F. Llorca, and M. A. Sotelo, “Pedestrian Path Prediction using Body Language Traits,” 2014.

[71] R. Quintero, I. Parra, D. F. Llorca, and M. A. Sotelo, “Pedestrian path prediction based on body language and action classification,” in *ITSC*, 2014, pp. 679–684.

[72] R. Quintero, I. Parra, D. F. Llorca, and M. A. Sotelo, “Pedestrian intention and pose prediction through dynamical models and behaviour classification,” in *ITSC*, 2015, pp. 83–88.

[73] R. Quintero, I. Parra, J. Lorenzo, D. Fernández-Llorca, and M. A. Sotelo, “Pedestrian intention recognition by means of a hidden markov model and body language,”in *ITSC*, 2017, pp. 1–7.

[74] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmüdderich, “Pedestrian crossing prediction using multiple context-based models,” in *ITSC*, 2014, pp. 378–385.

[75] B. Völz, H. Mielenz, G. Agamennoni, and R. Siegwart, “Feature relevance estimation for learning pedestrian behavior at crosswalks,” in *ITSC*, 2015, pp. 854–860.

[76] J.-Y. Kwak, E.-J. Lee, B. Ko, and M. Jeong, “Pedestrian’s intention prediction based on fuzzy finite automata and spatial-temporal features,” *Electron. Imaging*, vol. 2016, no. 3, pp. 1–6, 2016.

[77] S. Neogi, M. Hoy, W. Chaoqun, and J. Dauwels, “Context based pedestrian intention prediction using factored latent dynamic conditional random fields,” in *Computational Intelligence (SSCI), 2017 IEEE Symposium Series on*, 2017, pp. 1–8.

[78] S. Yi, H. Li, and X. Wang, “Pedestrian behavior understanding and prediction with deep neural networks,” in *European Conference on Computer Vision*, 2016, pp. 263–279.

[79] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social lstm: Human trajectory prediction in crowded spaces,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971.

[80] B. Cheng, X. Xu, Y. Zeng, J. Ren, and S. Jung, “Pedestrian trajectory prediction via the Social-Grid LSTM model,” *J. Eng.*, vol. 2018, no. 16, pp. 1468–1474, 2018.

[81] N. Kalchbrenner, I. Danihelka, and A. Graves, “Grid long short-term memory,” *arXiv Prepr. arXiv1507.01526*, 2015.

[82] F. Bartoli, G. Lisanti, L. Ballan, and A. Del Bimbo, “Context-aware trajectory prediction,” in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 1941–1946.

[83] H. Su, Y. Dong, J. Zhu, H. Ling, and B. Zhang, “Crowd Scene Understanding with Coherent Recurrent Neural Networks.,” in *IJCAI*, 2016, vol. 1, p. 2.

[84] H. Su, J. Zhu, Y. Dong, and B. Zhang, “Forecast the Plausible Paths in Crowd Scenes.,” in *IJCAI*, 2017, vol. 1, p. 2.

[85] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, “Social gan: Socially acceptable trajectories with generative adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264.

[86] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, “Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection,” *Neural networks*, vol. 108, pp. 466–478, 2018.

[87] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, “Pedestrian trajectory prediction with structured memory hierarchies,” *arXiv Prepr. arXiv1807.08381*, 2018.

[88] A. Vemula, K. Muelling, and J. Oh, “Social attention: Modeling attention in human crowds,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1–7.

[89] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, “Structural-RNN: Deep learning on spatio-temporal graphs,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5308–5317.

[90] S. Haddad, M. Wu, H. Wei, and S. K. Lam, “Situation-Aware Pedestrian Trajectory Prediction with Spatio-Temporal Attention Model,” *arXiv Prepr. arXiv1902.05437*, 2019.

[91] Y. Xu, Z. Piao, and S. Gao, “Encoding Crowd Interaction With Deep Neural Network for Pedestrian Trajectory Prediction,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5275–5284.

[92] H. Xue, D. Q. Huynh, and M. Reynolds, “Bi-Prediction: Pedestrian Trajectory Prediction Based on Bidirectional LSTM Classification,” in *Digital Image Computing: Techniques and Applications (DICTA), 2017 International Conference on*, 2017, pp. 1–8.

[93] H. Xue, D. Q. Huynh, and M. Reynolds, “SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 1186–1194.

[94] H. Xue, D. Huynh, and M. Reynolds, “Location-Velocity Attention for Pedestrian Trajectory Prediction,” in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 2038–2047.

[95] M. Pfeiffer, G. Paolo, H. Sommer, J. Nieto, R. Siegwart, and C. Cadena, “A data-driven model for interaction-aware pedestrian motion prediction in object cluttered environments,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1–8.

[96] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, “SR-LSTM: State Refinement for LSTM towards Pedestrian Trajectory Prediction,” *arXiv Prepr. arXiv1903.02793*, 2019.

[97] X. Shi, X. Shao, Z. Guo, G. Wu, H. Zhang, and R. Shibasaki, “Pedestrian trajectory prediction in extremely crowded scenarios,” *Sensors*, vol. 19, no. 5, p. 1223, 2019.

[98] S. Huang *et al.*, “Deep learning driven visual path prediction from a single image,” *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5892–5904, 2016.

[99] D. Varshneya and G. Srinivasaraghavan, “Human trajectory prediction using spatially aware deep attention models,” *arXiv Prepr. arXiv1705.09436*, 2017.

[100] K. Saleh, M. Hossny, and S. Nahavandi, “Long-Term Recurrent Predictive Model for Intent Prediction of Pedestrians via Inverse Reinforcement Learning,” 2017.

[101] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr, and M. Chandraker, “Desire: Distant future prediction in dynamic scenes with interacting agents,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 336–345.

[102] K. Sohn, H. Lee, and X. Yan, “Learning structured output representation using deep conditional generative models,” in *Advances in neural information processing systems*, 2015, pp. 3483–3491.

[103] H. Zou, H. Su, S. Song, and J. Zhu, “Understanding human behaviors in crowds by imitating the decision-making process,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[104] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha, “Reciprocal n-body collision avoidance,” in *Robotics research*, Springer, 2011, pp. 3–19.

[105] J. Ho and S. Ermon, “Generative adversarial imitation learning,” in *Advances in Neural Information Processing Systems*, 2016, pp. 4565–4573.

[106] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, “Infogan: Interpretable representation learning by information maximizing generative adversarial nets,” in *Advances in neural information processing systems*, 2016, pp. 2172–2180.

[107] B. Völz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, “A data-driven approach for pedestrian intention estimation,” in *ITSC*, 2016, pp. 2607–2612.

[108] Z. Fang, D. Vázquez, and A. López, “On-board detection of pedestrian intentions,” *Sensors*, vol. 17, no. 10, p. 2193, 2017.

[109] K. Saleh, M. Hossny, and S. Nahavandi, “Intent prediction of vulnerable road users from motion trajectories using stacked LSTM network,” in ITSC, 2017, pp. 327–332.