IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 12, 2021, accepted September 26, 2021, date of publication October 5, 2021, date of current version October 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3118224

A Survey on Motion Prediction of Pedestrians and Vehicles for Autonomous Driving

MAHIR GULZAR[1], YAR MUHAMMAD[2], (Senior Member, IEEE), AND NAVEED MUHAMMAD[1]

[1]Institute of Computer Science, University of Tartu, 51009 Tartu, Estonia
[2]Department of Computing & Games, School of Computing, Engineering & Digital Technologies, Teesside University, Middlesbrough TS1 3BX, U.K.

Corresponding author: Mahir Gulzar (mahir.gulzar@ut.ee)

This work was funded by the European Social Fund via IT Academy Programme.

ABSTRACT Autonomous vehicle (AV) industry has evolved rapidly during the past decade. Research and development in each sub-module (perception, state estimation, motion planning etc.) of AVs has seen a boost, both on the hardware (variety of new sensors) and the software sides (state-of-the-art algorithms). With recent advancements in achieving real-time performance using onboard computational hardware on an ego vehicle, one of the major challenges that AV industry faces today is modelling behaviour and predicting future intentions of road users. To make a self-driving car reason and execute the safest motion plan, it should be able to understand its interactions with other road users. Modelling such behaviour is not trivial and involves various factors e.g. demographics, number of traffic participants, environmental conditions, traffic rules, contextual cues etc. This comprehensive review summarizes the related literature. Specifically, we identify and classify motion prediction literature for two road user classes i.e. pedestrians and vehicles. The taxonomy proposed in this review gives a unified generic overview of the pedestrian and vehicle motion prediction literature and is built on three dimensions i.e. motion modelling approach, model output type, and situational awareness from the perspective of an AV.

INDEX TERMS Autonomous driving, road vehicles, roads, trajectory prediction, vehicle safety, human intention and behavior analysis.

I. INTRODUCTION

Safety Safety is a crucial aspect for an autonomous vehicle (AV). Other road users that an AV needs to interact with, come in many forms. It can be pedestrians, cyclists, skate-boarders or other vehicles etc. The challenge of predicting human motion comes from the complexity of modelling it using many underlying factors. For vehicles, it can depend on the behaviour of other vehicles, traffic rules, type of driving attitude and environmental context etc. In the case of pedestrians, human motion can be driven by personal goals, social relations, the behaviour of other agents and context of the environment etc. This means that for an AV to coexist with other road users, not only it should follow the traffic rules and regulations but also be socially aware i.e. it should understand the interactions of road users to ensure the flow of traffic [1]. Understanding these interactions helps an AV to forecast trajectories of road users giving an AV a complete overview of how the scene will unfold in next time-steps and what motion plan it should execute in order to ensure maximum safety of all the traffic participants.

This survey gives a comprehensive overview of motion prediction from the perspective of an AV. The scope of this survey is trajectory prediction of road users. We are interested in categorizing state-of-the-art literature into a novel taxonomy that incorporates motion prediction methods of the two road user classes i.e. pedestrians and vehicles. For these road user types, this work classifies the literature on the basis of modelling approach (physics, learning-based), output type (trajectories, intentions etc.) and situational awareness (interactions with scene objects). For this, we survey a collection of motion prediction methods, discuss their pros and cons and assign each method a certain generic category of our proposed taxonomy.

The paper is structured as follows. Section II discusses existing reviews on motion prediction. This includes

The associate editor coordinating the review of this manuscript and approving it for publication was Xiangxue Li.

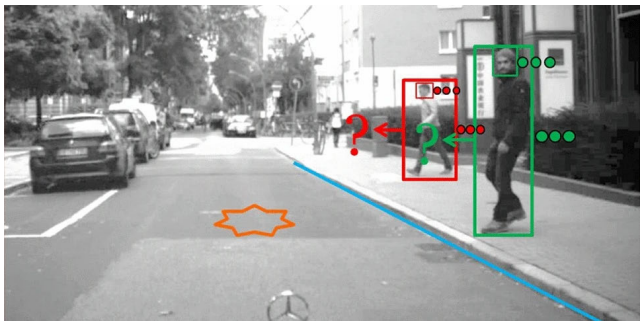

FIGURE 1. Will this pedestrian cross? [4].

surveys for both pedestrian and vehicle motion prediction. In section III, we discuss the novel taxonomy of motion prediction, the type of modelling methods and factors involved, also we present a table summarising the literature reflecting the proposed taxonomy. Section IV concludes this survey by highlighting useful takeaways, potential future work and challenges.

A. CHALLENGES

Motion prediction for road users comes with many inherent challenges which differ w.r.t road user classes. While AV's perception stack is responsible for detecting different types of road agents as well as predicting their future motion, motion prediction plays a pivotal role in understanding scene dynamics and drives efficient decision making of an AV [2]. In this section, we will discuss different challenges associated with both vehicle and pedestrian motion prediction.

1) PEDESTRIANS

In complex urban environments, AV's interact with different types of road users, pedestrians being one of them which belong to the most vulnerable road user class [3]. The interaction with pedestrians requires an AV to understand their intentions. For example, Figure-1 shows a scenario in which the AV has to reason about whether the pedestrian will cross or not.

In order to negotiate similar situations, we humans, in addition to traffic rules, employ informal social rules or often engage in non-verbal e.g. gesture-based communication to resolve a certain interaction. An example of such gestures is shown in Figure-2.

Understanding such informal social norms increases the safety of pedestrians. An AV's decision making could be designed in such a way that it acts cautiously to any interaction made with a pedestrian and always allow the pedestrian to lead the negotiation depending upon the scenario. However, this is prone to errors too. Several cases of robot bullying are reported in which pedestrians were seen attacking and stopping robots [5].

2) VEHICLES

Motion prediction of vehicles is governed by traffic rules and road geometry. Predicting possible future manoeuvre

FIGURE 2. An example of gestures used by pedestrians to negotiate an interaction [1].

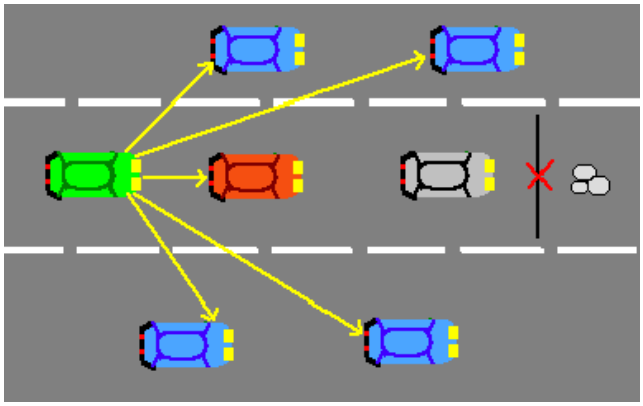

FIGURE 3. An illustration of how an occluded surrounding vehicle can affect the behaviour of a VOI. Here the green vehicle is our ego vehicle, the red vehicle is VOI whose future behaviour we intend to predict, the grey vehicle ahead of the red is an occluded surrounding vehicle that is about to apply emergency brakes due to a road obstruction ahead of it.

and trajectory of a vehicle has its own complexities e.g. a vehicle's motion is not only dependent on traffic rules but is also affected by other surrounding vehicles and other road users, which are sometimes occluded too. So an AV should take into account all surrounding vehicles to forecast the possible future trajectory of a vehicle of interest (VOI). For example, a leading vehicle might apply emergency brakes due to another vehicle ahead of it which is occluded to AV's onboard sensors, leading into a trickle-down effect of emergency braking. Figure-3 illustrates such a scenario. Similar to this, a vehicle's future behaviour is also dependent on road and traffic signs. Figure-4 shows a scenario in which the possible future trajectory of a VOI is affected by the road lane change markings.

II. RELATED WORK

A decent amount of literature is available on motion prediction for road users, but the surveys and reviews are somewhat limited and are specific to classes of road users. In [6] authors discuss pedestrian motion prediction methods for vehicle safety systems. Authors categorize pedestrian motion models into four types which include dynamic methods (methods that rely on target agent's motion), physiological methods

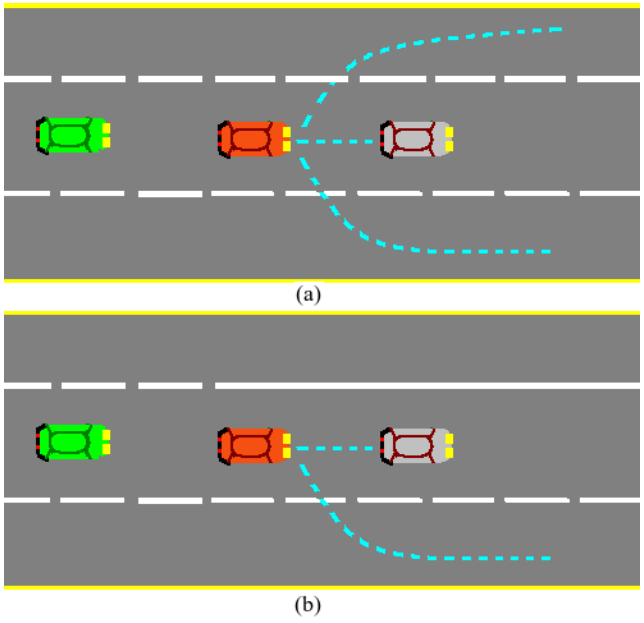

FIGURE 4. An example of where a vehicle's behaviour prediction is affected by road lane markings. a) shows that the possible future trajectory of the red vehicle includes left lane change, right lane change and follow-the-leading-vehicle manoeuvre. b) shows that the possible future trajectory of red vehicle includes right lane change and follow-the-leading-vehicle manoeuvre. The left lane change manoeuvre is excluded due to no lane change marking on the left side.

(methods that use information about relative pedestrian positioning w.r.t. to car, pedestrians velocity and direction), methods that use head orientation of the pedestrians, and lastly methods that use static environment context such as information about the position of sidewalks etc. Research in [7] surveys pedestrian motion and their interaction with AV. The literature presented here can be categorized as studies related to kinematic pedestrian models, models that include pedestrian's psychological constraints such as intention to accelerate and decelerate etc, models to estimate pedestrian's head orientation and lastly, the models that include environmental cues such as distance to curbs and obstacles.

In [8] authors use similar pedestrian motion prediction classification i.e. body pose (methods that rely on pedestrian body pose to estimate whether the pedestrian will cross the crossing or not), social-based (methods that consider social norms between people and use these norms for decision making), dynamics-based (methods that use tracking filters to estimate the future position of pedestrians), dynamics and awareness based (methods that incorporate pedestrian heading with dynamics to get a better estimation of pedestrian positioning).

The studies in [9] discuss motion prediction methods for vehicles. These methods are categorized into physics-based, manoeuvre-based, and interaction-aware models. Here physics-based models include models which use laws of physics to predict the future motion of vehicles. Manoeuvre based models predict vehicle's behaviour based on possible manoeuvres and lastly, interaction aware models use contextual awareness i.e. interaction with other vehicles. In addition to this, brief literature on risk evaluation for an AV is also presented.

[10] reviews deep learning-based vehicle behaviour prediction methods. Authors propose a classification of these methods based on input representation (track history, bird's eye view, raw sensor data), output type (manoeuvre intention, unimodal trajectory, multi-modal trajectory, occupancy maps) and prediction method (recurrent neural network, convolutional neural networks and others). The paper also discusses different evaluation metrics used for vehicle behaviour prediction.

A novel taxonomy that classifies human motion prediction is proposed in [11]. This taxonomy is not specific to any road user class and targets different applications of human motion prediction such as mobile robots, surveillance and autonomous driving. The taxonomy proposed here classifies motion prediction literature into two general categories i.e. on the basis of modelling approach (physics-based, pattern-based, planning-based) and using contextual cues (agent cues, dynamic and static environment cues).

Our survey extends the categorization used by [9], [10] and [11] and gives an overview of motion prediction for both pedestrians and vehicles. We generalize all deep learning-based [10] and pattern-based [11] methods into learning-based methods. Additionally, we also incorporate planning-based methods from [11] into learning-based methods. This is due to the fact that most of these methods essentially try to learn the future goal of agents based on some reward function or learn the reward function itself. This work reduces the sub-hierarchy of contextual cues [11] into unaware, interaction-aware, scene aware and map-aware methods. Here it is argued that some sub-categories proposed in [11] such as articulated pose and semantic attributes of agent do not reflect well in terms of vehicle motion prediction. Additionally, we leverage the classification of [10] and add an additional dimension of output type to our motion prediction taxonomy. The manoeuvre intention sub-category is redefined as intent prediction which makes it generic enough to incorporate vehicles as well as pedestrians.

III. MOTION PREDICTION TAXONOMY

This section explains our motion prediction taxonomy. The proposed taxonomy is built on three dimensions i.e. modelling approach, output type and situational awareness as shown in Figure-5. Here we discuss how these dimensions classify the motion prediction literature. The papers we discuss in each category may also be a part of other categories. Each category discussed here is independent of others, giving the researchers ease of exploring a specific category of papers. For instance, while exploring literature classified based on output type, a researcher might not want to limit their search to a specific modelling approach. At the end of this section, the taxonomy and literature are summarized into a table giving an overview of the classification of each paper.

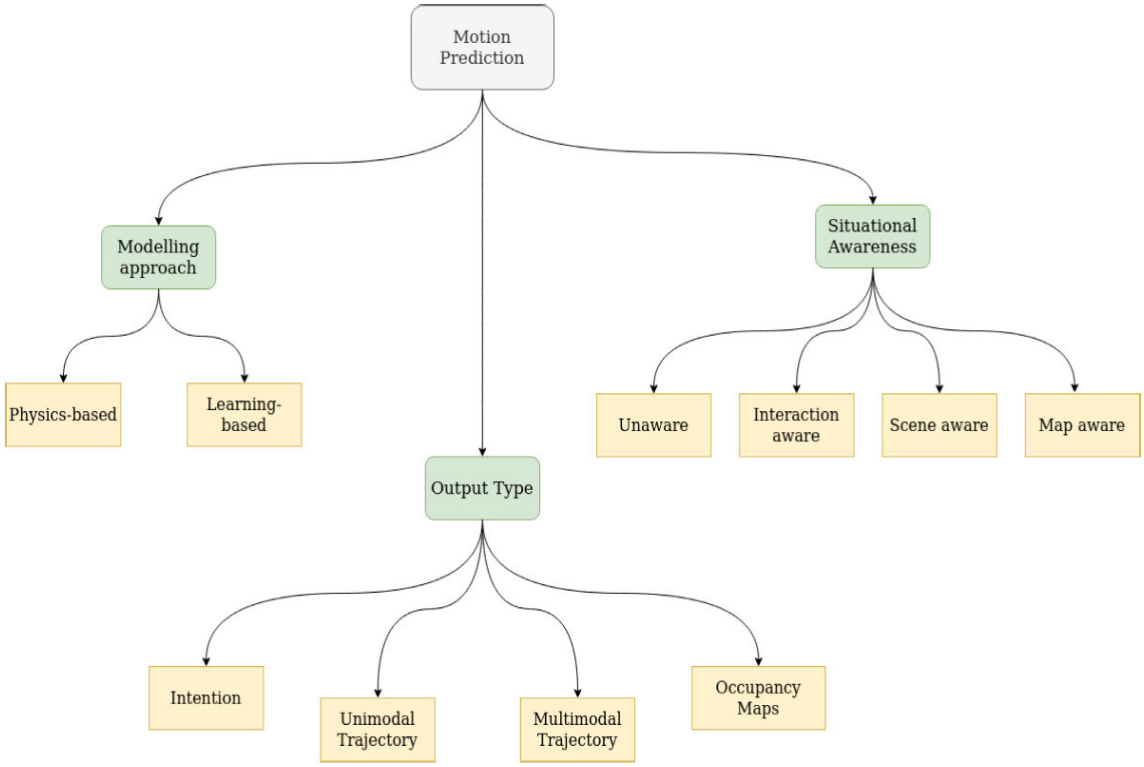

FIGURE 5. Proposed taxonomy of motion prediction for pedestrians and vehicles.

A. MODELLING APPROACH

Modelling approaches fall under the umbrella of motion prediction. Here modelling approaches are sub-categorized into physics-based and learning-based approaches. Physics-based approaches use equations of motion and physics to simulate the forward motion of agents. Learning-based approaches learn from data and statistics and recognize motion patterns from learned models. Here learning based models can further be categorized based on the type of method used, for example, clustering, Bayesian networks, convolutional neural networks etc. We will discuss these modelling approaches in detail in the following sections.

1) PHYSICS-BASED

Physics-based methods are governed by the laws of physics. These include dynamic and kinematic models based on Newton's laws. Dynamic models consider all forces that govern motion. Dynamic models get very complex due to the factors involved. For example, in case of vehicles, dynamic models consider forces acting on tires, drivers actions and their effect on the vehicle's engine and transmission. For trajectory prediction, it is rather irrelevant to model such complex behaviour using dynamic models unless we intend to do control-oriented applications [9]. Kinematic models on the other hand describe the motion in terms of mathematical relationship between movement parameters. Kinematic models are very common for trajectory prediction due to their simplicity of use. A simple example of a kinematic model is the constant velocity (CV) model used by [12]. A CV model assumes that the recent relative motion of an object drives its future trajectory. Figure-6 and Figure-7 show an example of such a model in action. If we denote the position of a pedestrian or a vehicle by

$$p = (x^t, y^t) \tag{1}$$

at time-step t, where x and y are top-down coordinates of the scene then, p denotes the position of the pedestrian or vehicle.

$$\Delta p = p^t - p^{t-1} \tag{2}$$

Δp is the most recent information to predict the future trajectory.

Another popular example of kinematic models is the constant acceleration (CA) model which assumes recent relative change in acceleration to be the factor that drives the future trajectory of the dynamic object. [13] is an example, where CA is used for collision warning and auto-braking system for a vehicle to help avoid collision with pedestrians.

A quite reasonable amount of literature is available on pedestrian and vehicle tracking by physics-based models. [14] uses Kalman filter (KF) and CA as a process model to filter dynamic obstacles. [15] uses a particle filter for hazard inference from linear motion predictions of pedestrians. [16] predict pedestrian motion along a road semantic graph. Using a unicycle model, the prediction algorithm assumes rational behaviour pedestrians i.e. pedestrians using crosswalks and accounts for road semantics in mathematical equations.

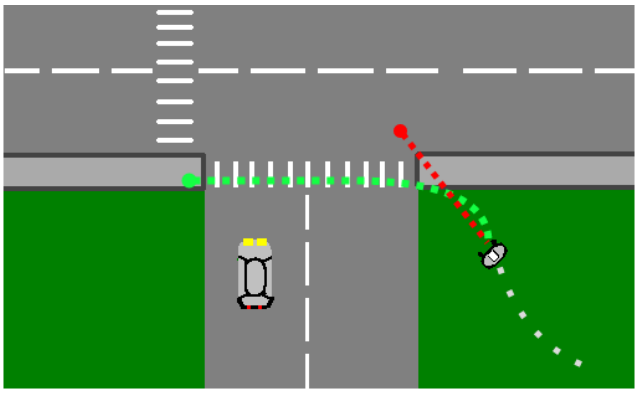

FIGURE 6. An illustration of pedestrian trajectory prediction from constant velocity (CV) model. Here the green trajectory denotes the ground truth trajectory of a pedestrian and the red trajectory shows the predicted trajectory at the current instant of time.

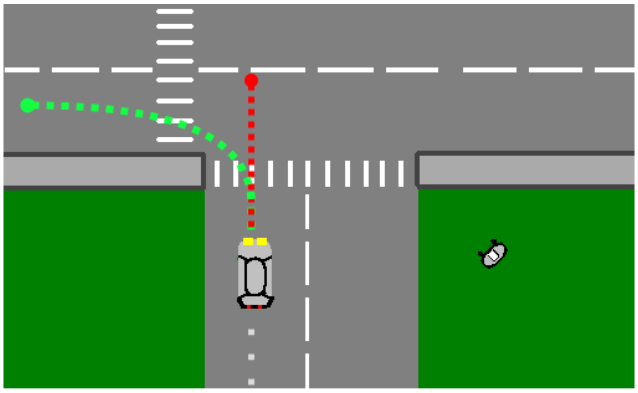

FIGURE 7. An illustration vehicular trajectory prediction from constant velocity (CV) model. Here the green trajectory denotes the ground truth trajectory of the vehicle which is about to do a turn left manoeuvre and the red trajectory shows the predicted trajectory at the current instant of time.

In addition to CV and CA models, some authors have presented Kalman filters using Constant Turn Rate and Velocity (CTRV) [17] and Constant Turn Rate and Acceleration (CTRA) [18] to capture the non-linearity of the trajectory. A recent extension of the physics-based approach uses IMMTP (Interactive multiple trajectory prediction) with Unscented Kalman filter (UKF) and Dynamic Bayesian Network (DBN) [19]. The integration of both the predictors using IMM gives non-linear trajectory prediction along with possible manoeuvre estimates (which in the above work is the lane changing manoeuvre). Such methods can be used for vehicular trajectory prediction if the uncertainties are handled well. Kalman filters represent these uncertainties in the form of Gaussian noise, where in the prediction step the filter outputs the position estimates based on kinematic or dynamic models and later the prediction estimates are updated based on sensor measurements. The problem of using filters like this is the unimodal representation of uncertainties which cannot capture the complex vehicle trajectory behaviours. A better representation of uncertainties is a mixture of Gaussians.

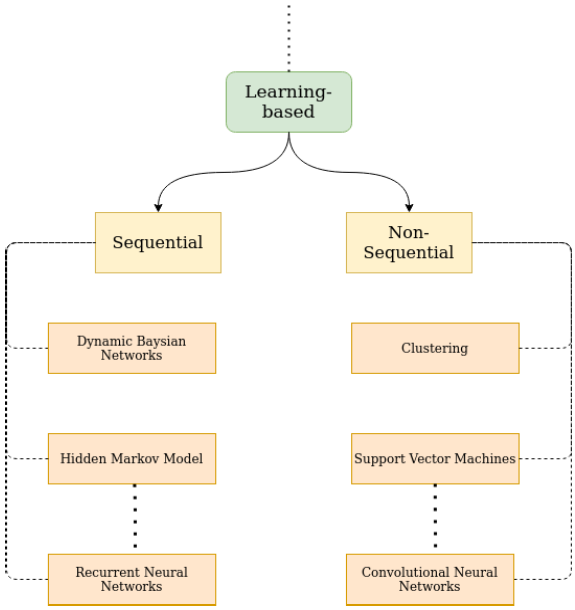

FIGURE 8. Categories of learning-based models. The dotted lines represent that the methods in these categories are not limited to the three methods mentioned in each category.

2) LEARNING-BASED

Learning-based methods have an element of learning in them i.e. they learn from data and history. The history and data can include past track history of vehicles, the birds-eye view of the environment, raw sensor data etc [10]. In comparison to physics-based models which are limited to low-level properties of motion and cannot estimate well the long term dependencies in motion, the learning-based models on the other hand tend to capture and incorporate long term dependencies and changes caused by external factors. Learning-based approaches have seen quite a lot of research in the past decade. The learning part here can be a function approximator, clustering algorithm or a hidden Markov model etc. Recent boom in deep learning has pushed these methods to an even higher level. We further sub-classify learning-based methods into two categories i.e. sequential and non-sequential, as shown in Figure-8.

a: SEQUENTIAL

Sequential models infer motion estimates of agents using the history of their states. Sequential models are quite similar to physics-based models in terms of Markovian assumption i.e. future motion of the agent is dependent on the current state of the agent. Sequential methods are often one step predictors similar to physics-based methods; the difference lies in learning functions from statistical observations instead of using motion models. One of the common types of sequential models is Dynamic Bayesian Network (DBN). A DBN essentially is a Bayesian network with temporal updates. These probabilistic models tend to be very useful for domains where observations are unrolled in time. A good example of DBN is the work done in [20] where authors employ a

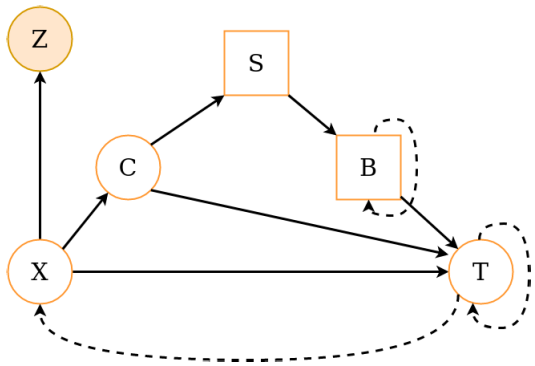

FIGURE 9. A DBN of vehicle motion prediction. Adapted from [21], where dashed lines represent temporal updates and solid lines represent updates within the current time frame.

DBN using agent dynamics and scene semantics to forecast pedestrian local patterns as local transition maps. Similar to pedestrians, the authors in [21] present a DBN to represent driver behaviour and vehicle trajectories.

The DBNs have Markov property where in order to satisfy Markov assumption we can enrich states with more information. In [21], this is done by adding all the relevant information of the process in the form of vectors to the DBN. This is illustrated in Figure-9.

Let, R be a set of state space random variables.

$$R = \{ X, C, S, B, T, Z \}$$

Assuming there are n number of vehicles in the scene at timestep T.

$$r = (r_1 \quad \ldots, \quad r_n)^T \quad \forall\, r_i \in R \tag{3}$$

Here, (3) is adapted from [21]. The state space random variables mentioned above are vectors, where X represents a vector containing vehicle states, the vector C represents local situational context (distance from other nearby relevant vehicles), S is vector recognized situations e.g. in context of distance to a vehicle, the recognized situation classes can be far-from-vehicle or close-to-vehicle etc, B is a vector of behaviours, the vector T represents vehicle trajectories and lastly, the vector Z represents sensor measurements about vehicle pose etc.

Recent works in capturing complex long-term dependencies of agents have a more generalizable approach in terms of location. These methods use sequence models, usually neural networks, to predict time series which is adapted from sequence modelling applications such as natural language processing. In particular, Recurrent Neural Networks (RNN) and Long Short-term Memory (LSTM) networks (a flavour of RNNs) have become popular to predict motion trajectories. In this regard, [22] was the first one to use joint-trajectories of pedestrians to predict multi-model paths, taking into account the social behaviour and common-sense rules that humans utilize while navigating. Here individual LSTMs were used to capture motion history of individual agents and later on a social pooling layer was applied to capture interactions among multiple agents. Here, social pooling is a network pooling layer that shares the information of interaction of individual spatially proximal LSTMs with each other. This work was later extended by [23] that, in addition to social pooling, applied a contextual pooling layer to encode environmental context into motion prediction in crowded spaces.

Some recent works in human trajectory predictions also include Spatio-temporal feature encoding and decoding architectures where temporal features are extracted from LSTMs and spatial features of the environment from convolutional neural networks (CNN). An example of such works is [24] that uses three LSTM encoders for three scales: (i) a person's scale observing an individual's trajectory, (ii) a social scale incorporating occupancy map and (iii) a scene scale that uses CNN to extract scene features. Later an LSTM decoder predicts human trajectory. Similar to pedestrians, a modified version of LSTM i.e ST-LSTM (Spatio-temporal LSTM) is used in [25] where the interaction of multiple vehicles and its effect on trajectory of VOI is estimated. The spatial information about other vehicles (calculated using safe distance function) is used to update the weights of the LSTM layers where more weight to an individual LSTM layer means that the particular trajectory will influence the trajectory of VOI more. Another example where both CNN and LSTM models are used is [26] where trajectories of VOI and other vehicles along with grid-based spatial positions are encoded into a social tensor using multiple LSTMs. Later a manoeuvre based decoder decodes manoeuvre based trajectories of VOI. This work is later extended into a multi-agent multimodal tensor fusion in [27] where, in addition to social convolution, the scene context is also encoded into the network.

Sequential methods have also leveraged Generative Adversarial Networks (GANs) along with recurrent neural networks. [28] extended the idea of social LSTM by using a social GAN to predict multi-modal human trajectories. Here a generator network G takes the input trajectory X and outputs predicted trajectories $\hat{Y}$, afterwards a discriminator network D takes the whole sequence comprising both input trajectory X and generator output $\hat{Y}$ and classifies them as real/fake. [29] extended the idea of social GAN and applied it to predict trajectories of vehicles.

b: NON-SEQUENTIAL

Non-sequential models learn over data and its distribution without constraints like Markovian assumption. These models tend to predict complete trajectories without relying on the feedback of past frames. One of the most common examples of such models is the clustering trajectories model. The notion behind the clustering approach is to understand the global motion pattern in the form of a cluster and impose it over local movement patterns of individual agents. A good example of such an approach is [30] where authors have used Ensemble Kalman Filter (EnKF) to track pedestrian state; later the individual tracks are clustered to get the local pattern and global flow of the crowd. To account for location

invariance, some methods use prototype trajectories which means learning the complete trajectories based on partially observed trajectories. [31] as an example used probabilistic trajectory matching to classify if a pedestrian is going to stop or cross the road.

For vehicles, a road network can be represented as a finite set of clusters of trajectories, where each cluster shows a motion pattern. Similar to transition maps, these patterns are learned from data. During inference, a partially observed trajectory can be assigned to a cluster, thus the most likely motion pattern can be obtained. Another way to represent prototype trajectories is to utilize digital road maps. A digital map can help us identify all possible manoeuvres at a certain location. In this case, the clustering process can be exempted which means a trajectory can directly be assigned to a cluster in the training set. Clustering-based methods for vehicles are mostly applied to data obtained from automated traffic vision systems. Examples of clustering vehicle trajectories include [32]–[34].

Another popular approach for non-sequential methods is CNNs. A CNN is a deep learning-based method that has convolution layers and learnable weights. CNNs are commonly used for extracting features from images; later these features can be passed to a fully connected network to obtain some useful output which, in the case of behaviour prediction, can be a trajectory, or an occupancy or transition map. An example of such a network employed for the case of pedestrians is [35] where the authors claim that CNNs can do better in capturing long-term temporal dependencies compared to LSTMs. Authors in [36] used human and machine annotated images to forecast pedestrian movement using a model built on resNet [37]. Similar to pedestrians, CNNs have also been used to estimate behaviour of road vehicles. In [38] intermediate representation of scene information is passed to a CNN and afterwards, a semantic occupancy map with vehicle trajectory is obtained. [39] uses CNN to predict the intention of surrounding vehicles. Vehicle intention and trajectory is predicted using backbone CNNs over lidar data and rasterized maps in [40]. To capture temporal features from the data, 3D convolutions along with 2D were applied in [41]. Here object detection, tracking and motion forecasting are performed in an end-to-end fashion.

B. OUTPUT TYPE

1) INTENTION

Predicting the intention of a pedestrian or a vehicle can give valuable insight into how the scene dynamics will change in future time steps. For vehicles, intentions are categorized based on manoeuvre which the vehicle is going to do. For example, follow a leading vehicle, steer left, steer right, turn left, turn right, stop etc. The possible manoeuvres can vary depending upon the scene situation. For example in a four-way intersection, a vehicle can either do the following, go forward, turn left, turn right and stop. In case of a highway, these possible manoeuvres would change to follow the leading vehicle, keep the lane, change to the left lane and change to the right lane. An example of such work is [42] where the model predicts lane change intention on a highway. The model here outputs probability distribution for the three classes, no lane change, left lane change, right lane change whereas in [43] the authors just predicted lane change and lane-keep intention of the VOI. The manoeuvre intention is good for a high-level understanding of vehicle behaviour but predicting vehicle motion is usually more complex and is not limited to high-level understanding. For example, at some point, these intentions can be further divided into more concrete intentions e.g. sharp left lane change or sharp right lane change.

For pedestrians, the intent prediction doesn't involve manoeuvres but the activity or state in which a pedestrian is going to be in future. An example of such a work is [44] in which a Gaussian process is used to predict a pedestrian's future state along with pose and path in 1 s time horizon. Here the predicted intentions were walking, starting, standing and stopping. Similar to this, in [45] walking, standing, walking-crossing and running-crossing were used as classification states of pedestrian's future motion. Some authors used binary classification to predict whether the pedestrian will cross or not [46], [47].

2) UNIMODAL TRAJECTORY

Addressing the problem with the previous output type, a better representation of pedestrian and vehicular behaviour is having a trajectory output from the model which is more precise than just intention. A unimodal trajectory thus outputs one trajectory with discrete trajectory points (which can later be made continuous using splines or Bézier curves).

In case of vehicles, the behaviour on road can be defined as a set of complex manoeuvres, a unimodal trajectory output can thus further be defined as trajectory independent / dependent on the intended manoeuvre. A trajectory independent of the intended manoeuvre is a unimodal trajectory without consideration of possible manoeuvres on it. Here the position of VOI is estimated over time. An example is [48] where the output trajectory of heterogeneous traffic participants is predicted as a unimodal trajectory. Here they predict mean, standard deviation and correlation coefficient of bivariate Gaussian w.r.t the x and y positions of each trajectory point. Despite being a better representation of vehicle motion compared to intention, manoeuvre independent trajectories tend to average out between two manoeuvres when there is an equal chance of making two manoeuvres at the same time. This can lead to dangerous encounters as illustrated in Figure-10.

In comparison to this, a trajectory dependent on intended manoeuvre gives safer and meaningful future estimates of the vehicle. This will make sure that whenever we get a trajectory output, it is valid in terms of manoeuvre. Figure-11 shows an illustration of manoeuvre constrained trajectory. An example of this is [49] where the authors demonstrated a policy anticipation network model that outputs trajectories constrained on manoeuvres using CARLA

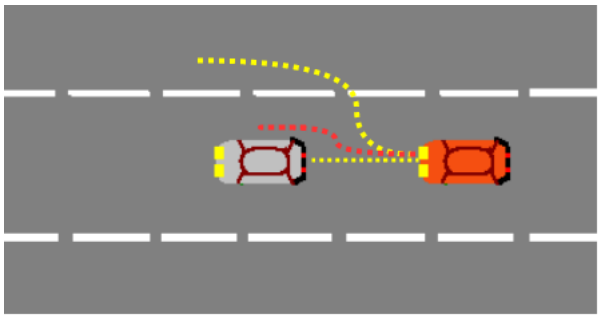

FIGURE 10. An Illustration of invalid manoeuvre adapted from [10]. The red vehicle is trailing and the grey vehicle is leading on a highway. There are two possible manoeuvres for the following vehicle i.e. take over from right or reduce speed and follow the lead vehicle shown with yellow dotted trajectories. A unimodal trajectory not constrained over manoeuvres might average out both options giving us the trajectory in red which results in a collision.

simulator [50]. The problem with uni-modal trajectory constrained on manoeuvres is that the possibility of exploring new trajectories gets limited to one trajectory only. This does not apply to pedestrians as pedestrians, in terms of unimodal trajectory, do not have manoeuvre categorizations i.e human walking behaviour is mostly influenced by social constraints. Some examples of work done on unimodal pedestrian trajectory include [12]–[16] and [24].

3) MULTIMODAL TRAJECTORY

An extension to unimodal trajectory is multimodal trajectory output. For vehicles, the limitation of having one mode as output gets addressed by having more than one trajectory output. At any particular instance of time, the VOI can choose from many correct manoeuvres or distribution of manoeuvres. Having the knowledge of this distribution makes the prediction algorithm more robust and less prone to unidentified trajectory outputs. In multimodal trajectory, we get a unimodal trajectory for each manoeuvre or mode. Like unimodal outputs, here the models can be dependent/independent of manoeuvres where the former means a probability distribution over finite sets of manoeuvres while later can have a fixed number of unimodal trajectory outputs independent of manoeuvres. An example of manoeuvre dependent multimodal output is convolution social pooling for vehicle trajectories [26] whereas a fixed-sized multimodal trajectory output independent of manoeuvres is given by [51]. Figure–12 shows the difference between unconstrained and constrained multimodal trajectory outputs.

The problem with unconstrained multimodal trajectory outputs is that they usually converge to one mode. This is called the mode collapse problem. This problem is usually addressed by carefully devising the loss function so that the model explores more varieties of outputs. The authors in [51] used a novel Multiple Trajectory Prediction (MTP) loss that gives some weightage of loss to other modes in addition to the one which is closest to the ground truth trajectory of the vehicle.

For pedestrians, having multimodal trajectories gives a better understanding of one's possible future motion. This is

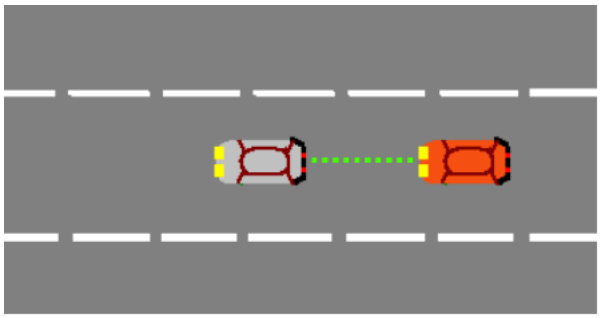

FIGURE 11. An Illustration of unimodal vehicle trajectory constrained on manoeuvre. Here the prediction shows that the trailing red vehicle will follow the leading grey vehicle.

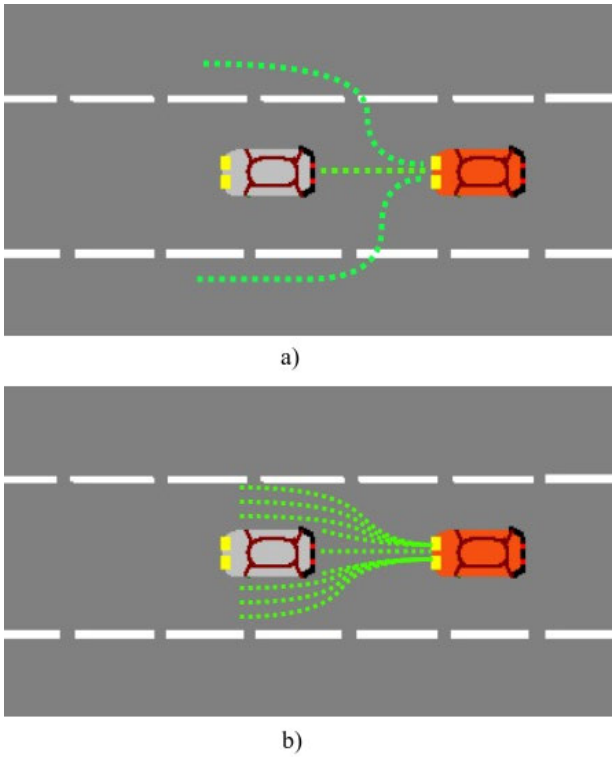

FIGURE 12. An illustration of constrained vs unconstrained multimodal trajectory output. Here a) is showing 3 manoeuvres (steer left, steer right and follow leading vehicle) while b) shows an unconstrained multimodal trajectory with fixed mode size = 9.

due to the fact that there are a variety of ways in which a pedestrian can interact with other pedestrians or avoid obstacles. In comparison to just predicting the intent or a single trajectory of a pedestrian, the multimodal output gives a much safer option for future motion planning of an AV. Figure-13 shows an illustration of unimodal vs multimodal pedestrian trajectory outputs in action. Some examples of work done on multimodal pedestrian trajectory prediction are [52]–[54].

4) OCCUPANCY MAPS

In occupancy maps and image rasters, a bird-eye view (BEV) of the scene occupancy is predicted in future timesteps, as illustrated in Figure-14. Occupancy maps are usually known to show occupancy of static obstacles in a scene,

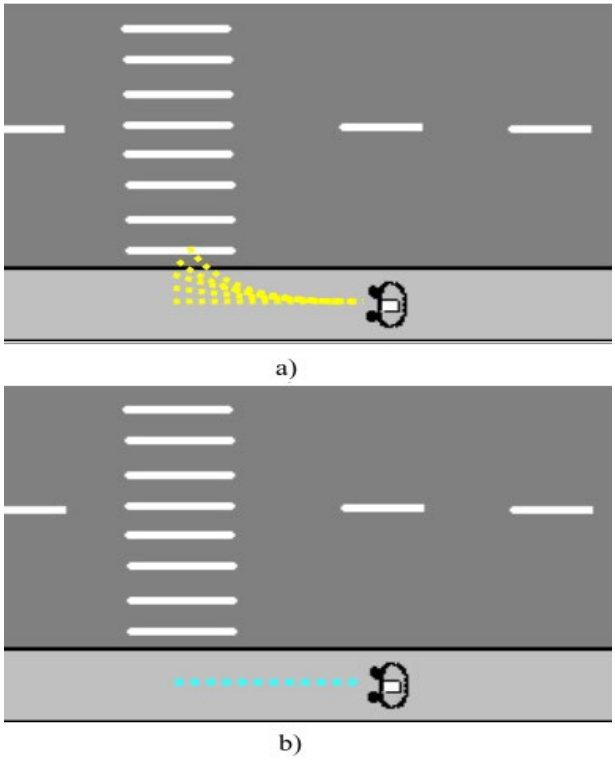

FIGURE 13. An illustration of unimodal vs multimodal pedestrian trajectory output. Here a) shows a distribution of possible trajectories where a possibility of pedestrian using the crosswalk is predicted while b) shows that the pedestrian will continue walking in the straight direction along the sidewalk.

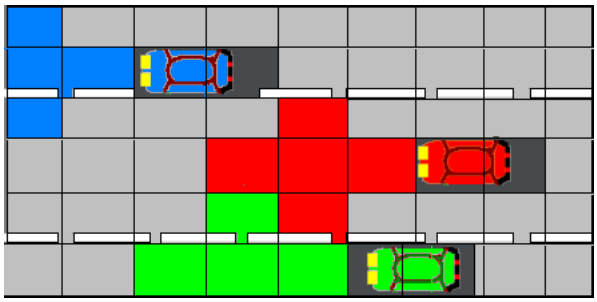

FIGURE 14. An example of a top-down dynamic occupancy grid. Here the grid cells in which vehicles currently exist are marked by dark grey colour showing the current occupancy while the colour coded grid cells represent possible future grid occupancies of the respective vehicles.

however dynamic occupancy grid maps [55] have also been used. In addition to occupancy grids, having coloured rasters in top-down grids give a more elaborative picture of the scene. An example of such a model is [38], where the motion prediction model outputs a grid map that illustrates vehicle trajectory as well as semantic segmentation of grid pixels giving rich information of what obstacles the model is trying to avoid. Similar to [38], the model proposed in [56] outputs multimodal grid occupancies for multiple vehicles. Here, for N number of vehicles in the scene, K best hypotheses are predicted on the occupancy grid map, where the grid size is (36 $\times$ 21) and each grid cell is 5.0 by 0.875 meters in longitudinal and lateral directions. Some more examples of occupancy predictions for road users are [57]–[59].

C. SITUATIONAL AWARENESS

Situational awareness corresponds to the information about the environment and other dynamic obstacles which can affect the possible future trajectory of the road agent. This information can be fed to the behaviour model in the form of numerical data e.g. coordinates of other road agents, top-down rasters showing the locations of static and dynamic obstacles, and road semantics etc. We divide the literature based on situational awareness into four broader categories.

1) UNAWARE

Following motion prediction methods predict the behaviour of traffic participants using the information obtained from the respective participant only i.e. they are unaware of the static and dynamic environment around them and cannot incorporate the behavioural changes which can be influenced by the surrounding elements. Most naive physics-based methods fall into the category of unaware models. CV and CA models discussed previously are good examples of such models. Some other examples include methods using Kalman filters with CTRA and CTRV process models [16], [17], [19].

2) INTERACTION AWARE

In comparison to unaware models, the interaction aware motion prediction models use information of surrounding agents as a guideline before estimating future motion of the road agents. This information can include coordinates of other road agents in the scene. The intuition here is that feeding this information to the model will make the model reason and consider possible future interactions with nearby agents before predicting the motion estimate of an agent. For pedestrians, an example of such a model is [22] where a social pooling layer captures interactions of a pedestrian of interest with others. In comparison to this, the authors in [28] used relative positioning of other pedestrians instead of social pooling grid. For vehicles, similar techniques are used to enrich the motion model with features of surrounding vehicles. For example, [21] adds a situational context by feeding longitudinal distance, lateral distance and relative velocities of surrounding vehicles. Similarly, the work done in [25] embeds spatial interactions of surrounding vehicles into LSTM layers.

3) SCENE AWARE

Scene awareness corresponds to the context of the environment. For example, a car driving on a highway road has a different scene context compared to a car at a four-way intersection or in a roundabout. Similarly, a pedestrian crossing the road using a crosswalk has a different scene context compared to a pedestrian using a sidewalk. The features of the corresponding scene are usually fed to the motion prediction model in the form of raw sensor data such as images. For pedestrians, an example of such a model is [23] where contextual aware

TABLE 1. Summary table of learning-based models showing existing works w.r.t proposed taxonomy categorization.

Road Agent	Output Type	Situational Awareness	Method	Code Availability	Works
Pedestrian	Intention	Unaware	SVM+Random-forest	✓	[47]
		Scene-aware	Gaussian Process	✗	[44]
	Unimodal	Unaware	Clustering	✗	[30]
			CNN	✗	[35]
			CNN	✓	[36]
		Scene-aware	LSTM+CNN	✓	[46]
		Interaction-aware	LSTM	✓	[22]
		Scene+Interaction-aware	LSTM	✗	[23]
			LSTM+CNN	✓	[24]
		Map+Interaction-aware	CNN	✗	[62]
			CNN	✗	[60]
	Multimodal	Interaction-aware	LSTM+GAN	✓	[28]
			LSTM+GAN	✗	[52]
		Scene+Interaction-aware	RNN	✗	[53]
			GAN+LSTM	✗	[54]
			CNN+LSTM+GAN	✓	[27]
	Occupancy Map	Map-aware	DBN	✗	[20]
Vehicle	Intention	Interaction-aware	CNN	✗	[39]
			CNN	✗	[40]
			CNN	✗	[41]
			GRU	✗	[43]
		Map-aware	SVM	✗	[42]
	Unimodal	Unaware	Clustering	✓	[32]
			Clustering	✗	[33]
			Clustering	✗	[34]
		Interaction-aware	DBN	✗	[21]
			LSTM	✗	[25]
			LSTM+GAN	✗	[29]
		Map-aware	RNN	✗	[49]
		Map+Interaction-aware	CNN	✗	[62]
			CNN	✗	[60]
	Multimodal	Scene+Interaction-aware	LSTM+CNN	✓	[26]
			CNN+LSTM+GAN	✓	[27]
		Interaction-aware	CNN	✓	[51]
	Occupancy-Map	Map-aware	CNN	✓	[38]
			CNN	✗	[61]
		Map+Interaction-aware	LSTM	✗	[56]
			LSTM	✗	[57]
			RNN	✗	[58]

pooling is used to capture scene context. Another example is [24] where a top-down image is given to the model that captures scene level features of the scene. A heterogeneous multi-agent motion model in [27] encodes scene context on top of individual LSTM layers that capture the context of the scene for both pedestrians and vehicles.

4) MAP AWARE

Map aware models are an extension of the scene aware models. They take semantic information of the map as a contextual cue. The semantic information includes HD maps which consist of lanes, road structures, traffic lights and road signs etc. This rich information guides the model's predictions to

TABLE 2. Summary table of physics-based models showing existing works w.r.t proposed taxonomy categorization. Note: All physics-based methods discussed in this survey have Unimodal output type.

Road Agent	Situational Awareness	Method	Code Availability	Works
Pedestrian	Unaware	CV	✓	[12]
		CA	✗	[16]
	Map-aware	Particle filter	✗	[15]
Vehicle	Unaware	CA	✗	[13]
		Kalman filter	✗	[14]
		CTRA	✗	[17]
		CTRV	✗	[18]
	Map-aware	IMM + DBN	✗	[19]

give logical trajectory outputs following semantic rules of road and traffic. A good example of a map-aware behaviour prediction model for pedestrians is [20] where pedestrian navigation map is fed to the model. Here, a navigation map of the scene essentially encodes the navigation patterns of similar class agents over a collection of feature patches. Similar to pedestrians, [38], [60] use road semantics as contextual information to refine AV's future motion estimates. The work presented in [62] and [63] passes image rasters of road semantics to the model and outputs future trajectories of different traffic actors. In comparison to scene-awareness, the map-aware models lack information about the environment or weather conditions that could be vital in extreme weather.

D. SUMMARY

Research in motion prediction for pedestrians and vehicles has moved from naive physics-based methods to machine learning-based models over the last decade. Physics-based models are simpler to implement and have been in use for a long time but they usually lack in terms of enriching the model with contextual information, thus most of these models predict motion estimates of an agent independent of other agents and environmental constraints. Modern approaches in motion prediction heavily rely on data. They handle complex long term dependencies quite well compared to physics-based models. These models extract contextual and situational information from the data and refine motion estimates accordingly.

In order to build a fail-safe and reliable navigation system for an AV, the motion model should have an in-depth understanding of the surrounding environment. A robust and reliable model should give a complete picture of possible future trajectories of the agent. Consequently, the output type and situational awareness of the model are very crucial for good behavioural estimates of the road agents. A reliable motion model for both pedestrians and vehicles should be situationally aware of road agent interactions, environmental context and road semantics. Similarly, having unimodal output does not give a complete picture of all possible future motion estimates; a good motion model should make use of multiple predictions using a combination of outputs e.g. multimodal trajectories with intentions etc.

The studies discussed in this paper are classified into the proposed taxonomy, shown in Figure-5. There is a possibility that a discussed study may belong to more than one category of the proposed taxonomy classification. The classification we make here may not fall under one category of taxonomy especially those studies which make use of more than one technique. For example, [23], [24] show models which are scene and interaction-aware. Similarly, [19] uses an IMM that predicts vehicular trajectory using Kalman filter and estimates vehicle's manoeuvre using DBN which are two different modelling approaches according to our proposed taxonomy. Another example is work done in [27] that presents a learning-based model which uses LSTM (sequential) + CNN (non-sequential) neural network to extract temporal and spatial features from the environment. We classify each study according to the modelling approach, output type and situational awareness that fits best to respective paper. Table 1 and Table 2 classify the above studies on the basis of different dimensions of our taxonomy.

IV. CONCLUSION

This work presents a novel taxonomy that classifies motion prediction literature of pedestrians and vehicles for autonomous driving. The studies discussed in this work are distinguished on the basis of three dimensions i.e. modelling approach, output type and situational awareness. For this, we reviewed the related literature and summarized the classification in the end. This work brings motion modelling methods for pedestrians and vehicles under one umbrella and shows how similar motion modelling methods are used for motion estimates of two different road agent classes. The work presented here gives an insight into developing and employing the existing motion models for pedestrians and vehicles in autonomous driving applications with the possibility of expanding this study to all road agent classes in future.

REFERENCES

[1] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, "Understanding pedestrian behavior in complex traffic scenes," *IEEE Trans. Intell. Vehicles*, vol. 3, no. 1, pp. 61–70, Mar. 2018.

[2] W. Zhan, A. de La Fortelle, Y.-T. Chen, C.-Y. Chan, and M. Tomizuka, "Probabilistic prediction from planning perspective: Problem formulation, representation simplification and evaluation metric," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1150–1156.
[3] *Global Status Report on Road Safety 2018: Summary*, World Health Org., Geneva, Switzerland, 2018.
[4] J. F. P. Kooij, F. Flohr, E. A. Pool, and D. M. Gavrila, "Context-based path prediction for targets with switching dynamics," *Int. J. Comput. Vis.*, vol. 127, no. 3, pp. 239–262, Mar. 2019.
[5] M. McFarland. (2017). *Robots Hit the Streets. The Streets Hit Back*. Accessed: Mar. 20, 2021. [Online]. Available: http://money.cnn.com/2017/04/28/technology/robot-bullying/
[6] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 900–918, Mar. 2020.
[7] N. Brouwer, H. Kloeden, and C. Stiller, "Comparison and evaluation of pedestrian motion models for vehicle safety systems," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 2207–2212.
[8] D. Ridel, E. Rehder, M. Lauer, C. Stiller, and D. Wolf, "A literature review on the prediction of pedestrian behavior in urban scenarios," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3105–3112.
[9] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH J.*, vol. 1, no. 1, pp. 1–14, 2014.
[10] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Trans. Intell. Transp. Syst.*, early access, Aug. 4, 2020, doi: 10.1109/TITS.2020.3012034.
[11] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *Int. J. Robot. Res.*, vol. 39, no. 8, pp. 895–935, 2020.
[12] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll, "What the constant velocity model can teach us about pedestrian motion prediction," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1696–1703, Apr. 2020.
[13] E. Coelingh, A. Eidehall, and M. Bengtsson, "Collision warning with full auto brake and pedestrian detection—A practical example of automatic emergency braking," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 155–160.
[14] A. Elnagar, "Prediction of moving objects in dynamic environments using Kalman filters," in *Proc. IEEE Int. Symp. Comput. Intell. Robot. Autom.*, Jul./Aug. 2001, pp. 414–419.
[15] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund, "Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 330–335.
[16] I. Batkovic, M. Zanon, N. Lubbe, and P. Falcone, "A computationally efficient model for pedestrian motion prediction," in *Proc. Eur. Control Conf. (ECC)*, Jun. 2018, pp. 374–379.
[17] N. Kaempchen, K. Weiss, M. Schaefer, and K. C. J. Dietmayer, "IMM object tracking for high dynamic driving maneuvers," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2004, pp. 825–830.
[18] P. Lytrivis, G. Thomaidis, and A. Amditis, "Cooperative path prediction in vehicular environments," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2008, pp. 803–808.
[19] G. Xie, H. Gao, L. Qian, B. Huang, K. Li, and J. Wang, "Vehicle trajectory prediction by integrating physics-and maneuver-based approaches using interactive multiple models," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5999–6008, Jul. 2018.
[20] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese, "Knowledge transfer for scene-specific motion prediction," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 697–713. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-46448-0_42
[21] T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 1625–1631.
[22] 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.*, Jun. 2016, pp. 961–971.
[23] F. Bartoli, G. Lisanti, L. Ballan, and A. D. Bimbo, "Context-aware trajectory prediction," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 1941–1946.
[24] H. Xue, D. Q. Huynh, and M. Reynolds, "SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1186–1194.
[25] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified LSTM models for trajectory prediction," *IEEE Access*, vol. 7, pp. 38287–38296, 2019.
[26] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 1468–1476.
[27] T. Zhao, Y. Xu, M. Monfort, W. Choi, C. Baker, Y. Zhao, Y. Wang, and Y. N. Wu, "Multi-agent tensor fusion for contextual trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 12126–12134.
[28] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2255–2264.
[29] L.-W. Kang, C.-C. Hsu, I.-S. Wang, T.-L. Liu, S.-Y. Chen, and C.-Y. Chang, "Vehicle trajectory prediction based on social generative adversarial network for self-driving car applications," in *Proc. Int. Symp. Comput., Consum. Control (IS3C)*, Nov. 2020, pp. 489–492.
[30] A. Bera, S. Kim, T. Randhavane, S. Pratapa, and D. Manocha, "GLMP-realtime pedestrian path prediction using global and local movement patterns," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 5528–5535.
[31] C. G. Keller, C. Hermes, and D. M. Gavrila, "Will the pedestrian cross? Probabilistic path prediction based on learned motion features," in *Proc. Joint Pattern Recognit. Symp.* Berlin, Germany: Springer, 2011, pp. 386–395. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-23123-0_39
[32] S. Atev, G. Miller, and N. P. Papanikolopoulos, "Clustering of vehicle trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 647–657, Sep. 2010.
[33] F. S. Hoseini, S. Rahrovani, and M. H. Chehreghani, "A generic framework for clustering vehicle motion trajectories," 2020, *arXiv:2009.12443*. [Online]. Available: http://arxiv.org/abs/2009.12443
[34] J. Martinsson, N. Mohammadiha, and A. Schliep, "Clustering vehicle maneuver trajectories using mixtures of hidden Markov models," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3698–3705.
[35] N. Nikhil and B. T. Morris, "Convolutional neural network for trajectory prediction," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCV)*, 2018, pp. 1–11.
[36] O. Styles, A. Ross, and V. Sanchez, "Forecasting pedestrian trajectory with machine-annotated training data," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 716–721.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[38] S. Srikanth, J. A. Ansari, R. K. Ram, S. Sharma, J. K. Murthy, and K. M. Krishna, "INFER: Intermediate representations for future prediction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 942–949.
[39] D. Lee, Y. P. Kwon, S. Mcmains, and J. K. Hedrick, "Convolution neural network-based lane change intention prediction of surrounding vehicles for ACC," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
[40] S. Casas, W. Luo, and R. Urtasun, "IntentNet: Learning to predict intention from raw sensor data," *CoRR*, vol. abs/2101.07907, pp. 1–10, Jan. 2021. [Online]. Available: https://arxiv.org/abs/2101.07907
[41] W. Luo, B. Yang, and R. Urtasun, "Fast and furious: Real time end-to-end 3D detection, tracking and motion forecasting with a single convolutional net," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3569–3577.
[42] P. Kumar, M. Perrollaz, S. Lefèvre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 797–802.
[43] W. Ding, J. Chen, and S. Shen, "Predicting vehicle behaviors over an extended horizon using behavior interaction network," *CoRR*, vol. abs/1903.00848, pp. 1–7, Mar. 2019. [Online]. Available: http://arxiv.org/abs/1903.00848
[44] 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.*, vol. 20, no. 5, pp. 1803–1814, May 2019.
[45] J.-Y. Kwak, B. C. Ko, and J.-Y. Nam, "Pedestrian intention prediction based on dynamic fuzzy automata for vehicle driving at nighttime," *Infr. Phys. Technol.*, vol. 81, pp. 41–51, Mar. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350449516304935

[46] A. Rasouli, I. Kotseruba, T. Kunic, and J. Tsotsos, "PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6261–6270.
[47] Z. Fang and A. M. López, "Is the pedestrian going to cross? Answering by 2D pose estimation," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1271–1276.
[48] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, and D. Manocha, "TrafficPredict: Trajectory prediction for heterogeneous traffic-agents," *CoRR*, vol. abs/1811.02146, pp. 1–8, Nov. 2018. [Online]. Available: http://arxiv.org/abs/1811.02146
[49] W. Ding and S. Shen, "Online vehicle trajectory prediction using policy anticipation network and optimization-based context reasoning," *CoRR*, vol. abs/1903.00847, pp. 1–7, Mar. 2019. [Online]. Available: http://arxiv.org/abs/1903.00847
[50] A. Dosovitskiy, G. Ros, F. Codevilla, A. M. López, and V. Koltun, "CARLA: An open urban driving simulator," *CoRR*, vol. abs/1711.03938, pp. 1–16, Nov. 2017. [Online]. Available: http://arxiv.org/abs/1711.03938
[51] H. Cui, V. Radosavljevic, F.-C. Chou, T.-H. Lin, T. Nguyen, T.-K. Huang, J. Schneider, and N. Djuric, "Multimodal trajectory predictions for autonomous driving using deep convolutional networks," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 2090–2096.
[52] S. Eiffert, K. Li, M. Shan, S. Worrall, S. Sukkarieh, and E. Nebot, "Probabilistic crowd GAN: Multimodal pedestrian trajectory prediction using a graph vehicle-pedestrian attention network," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5026–5033, Oct. 2020.
[53] A. Poibrenski, M. Klusch, I. Vozniak, and C. Müller, "M2P3: Multimodal multi-pedestrian path prediction by self-driving cars with egocentric vision," in *Proc. 35th Annu. ACM Symp. Appl. Comput.*, Mar. 2020, pp. 190–197.
[54] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, S. H. Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks," 2019, *arXiv:1907.03395*. [Online]. Available: http://arxiv.org/abs/1907.03395
[55] D. Nuss, S. Reuter, M. Thom, T. Yuan, G. Krehl, M. Maile, A. Gern, and K. Dietmayer, "A random finite set approach for dynamic occupancy grid maps with real-time application," *Int. J. Robot. Res.*, vol. 37, no. 8, pp. 841–866, 2018.
[56] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1672–1678.
[57] M. Schreiber, V. Belagiannis, C. Gläser, and K. Dietmayer, "Dynamic occupancy grid mapping with recurrent neural networks," 2020, *arXiv:2011.08659*. [Online]. Available: http://arxiv.org/abs/2011.08659
[58] B. Kim, C. M. Kang, J. Kim, S. H. Lee, C. C. Chung, and J. W. Choi, "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 399–404.
[59] P. Kaniarasu, G. C. Haynes, and M. Marchetti-Bowick, "Goal-directed occupancy prediction for lane-following actors," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 3270–3276.
[60] F.-C. Chou, T.-H. Lin, H. Cui, V. Radosavljevic, T. Nguyen, T.-K. Huang, M. Niedoba, J. Schneider, and N. Djuric, "Predicting motion of vulnerable road users using high-definition maps and efficient ConvNets," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 1655–1662.
[61] E. Rehder and H. Kloeden, "Goal-directed pedestrian prediction," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, Dec. 2015, pp. 50–58.
[62] N. Djuric, V. Radosavljevic, H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, N. Singh, and J. Schneider, "Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Mar. 2020, pp. 2095–2104.

MAHIR GULZAR received the bachelor's degree in computer science from GC University Lahore, Pakistan, and the master's degree in computer science from the University of Tartu, Estonia, where he is currently pursuing the Ph.D. degree in computer science. He also worked as a Scientific Programmer for applied research in autonomous driving with the University of Tartu. He is currently working as a Junior Research Fellow of autonomous driving with the Autonomous Driving Laboratory, University of Tartu.

YAR MUHAMMAD (Senior Member, IEEE) received the master's degree in computer engineering from Mid Sweden University, Sweden, in 2009, and the Ph.D. degree in information communication technology (ICT) from Tallinn University of Technology, Estonia, in 2015. He taught at the University of Tartu, Estonia. He is currently working as a Senior Lecturer (Assistant Professor) with Teesside University, U.K., where he is also a part of the Centre for Digital Innovation. He received a Young Investigator Award, which was awarded by Springer and IFMBE at 16th Nordic-Baltic Conference on Biomedical Engineering & Medical Physics and Medicinteknikdagarna 2014, Sweden, and he was runner-up for the Best Paper Award in the 26th ISSC, Ireland, in 2015.

NAVEED MUHAMMAD received the Ph.D. degree in robotics from INSA de Toulouse (research stay at LAAS-CNRS), France, in 2012. He has had postdoctoral stays at Tallinn University of Technology and Halmstad University, Sweden, and has taught at the National University of Sciences and Technology, Pakistan, and the Asian Institute of Technology, Thailand. He is currently working as an Assistant Professor in autonomous driving with the University of Tartu, Estonia, where he is also a part of the Autonomous Driving Laboratory. His research interests include autonomous driving, perception, and behavior modeling.