Incorporating Driving Knowledge in Deep Learning Based Vehicle Trajectory Prediction: A survey

Zhezhang Ding and Huijing Zhao

***Abstract*—Vehicle Trajectory Prediction (VTP) is one of the key issues in the field of autonomous driving. In recent years, more researchers have tried applying Deep Learning methods and techniques to VTP tasks. However, due to the black-box nature of Deep Learning, it cannot meet the interpretability and safety requirements of autonomous driving systems. Researchers have tried alleviating this problem by introducing driving knowledge in Deep Learning-based VTP. From the perspective of introducing driving knowledge, this paper systematically investigates the research status of DL-based VTP. First of all, this paper summarizes the research on VTP under three different problem formulations; secondly, this paper summarizes the application methods and application stages of driving knowledge in DL-based VTP; finally, this paper investigates and analyzes the VTP datasets and evaluation, and summarizes the knowledge contained in the datasets and its usage. Through the investigation and summary of problem formulation, knowledge usage, datasets, and evaluation of DL-based VTP, this paper analyzes the challenges and open questions of existing VTP research. It puts forward an outlook on future research directions.**

I. Introduction

The past decades have witnessed tremendous developments in autonomous vehicles [1]. In complex traffic scenes, due to a large number of surrounding vehicles and entities, autonomous vehicles must consider and understand the interaction with others to maintain safe and effective driving. Under this circumstance, predicting future trajectories of the driving agents on the scene has been studied broadly as a vital technique for an autonomous vehicle to make real-time decisions and maintain safe and efficient control. A part of the studies aims to predict possible trajectories of the autonomous ego vehicle [2], also known as motion prediction, in related research [3]. Another part focuses on predicting vehicle trajectories from an ego vehicle's perspective [4] or a top-down perspective [5]. Furthermore, studies have been conducted on predicting the trajectories of heterogeneous traffic agents, while vehicles are modeled as the main traffic agents in the scenes [6]. These studies are part of the large body of research on vehicle trajectory prediction (**VTP**), which has also been referred to by trajectory forecasting [6], vehicle behavior prediction [7], etc.

There are several related surveys on the VTP study. Lefevre et al. [3] give a systematic review of motion prediction and risk assessment methods for autonomous vehicles, in which they present a taxonomy to divide motion prediction methods into three subgroups: physic-based method, maneuver-based method, and interaction-aware method. This taxonomy is widely referred to and profoundly impacts the field of trajectory prediction. Schwarting et al. [8] review the planning and decision-making module of autonomous vehicles, in which the differences between traditional planning, behavior-aware planning, and end-to-end planning are discussed. In recent years, using deep learning (**DL**) methods to improve performance in complex traffic scenes has been the primary trend in VTP studies. Mozaffari et al. [7] give a detailed survey on deep learning-based VTP (**DL-based VTP**) methods from the perspective of the input representation, output type, and prediction method (i.e., deep neural networks) in recent years. [9] also gives a survey on trajectory prediction, in which the methods are divided into physics-based, classic machine learning-based, deep learning-based, and reinforcement learning-based methods. The input and output formulation, as well as the dataset for trajectory prediction, are also analyzed. Besides, driving trajectories are governed by the road, driver behavior, and vehicle models and are influenced by interactions with other traffic participants. In order to improve DL models to meet the interpretability and explainability requirements of autonomous driving systems [10], many studies incorporate extra information into DL-based VTP methods for reliable results, such as external environmental factors [11], [12], vehicle behavior [13] or motion constraints [14], and vehicle interaction patterns [15]. In this paper, all these driving-related factors are collectively referred to as **Driving Knowledge**, or **Knowledge** for short. A similar trend has been observed in the study of human trajectory prediction [16]. A recent review [17] compares DL-based and knowledge-based approaches and suggests the importance of combining both approaches.

In Fig.1, we collate the representative DL-based VTP methods in recent years and divide them into two groups, i.e. **knowledge-free** in red and **knowledge-based** in blue, based on whether the knowledge mentioned above has been incorporated into the models. The works reviewed in this paper are obtained by searching for ”Trajectory Prediction” and ”Vehicle Trajectory Prediction” on Google Scholar, IEEE Xplore, and Researchgate. The time range of papers is set from 2017 till now. We performed some filtering to make all the works reviewed in this paper fit the category of DL-based VTP. The methods related to connected vehicles or vehicular networks are not covered in this review. Hereinafter, we omit ”DL-based” and use ”VTP” for simplicity. From Fig.1, it can be seen that most of the representative works are knowledge-based, indicating that more attention has been paid to research

This work was supported in part by the National Natural Science Foundation of China under Grant 61973004 and U22A2061. Correspondence: H. Zhao, zhaohj@cis.pku.edu.cn.

Z. Ding and H. Zhao are with the Key Lab of Machine Perception (MOE) and also with the School of Intelligence Science and Technology, Peking University, Beijing, China. (e-mail: zzding@pku.edu.cn; zhaohj@pku.edu.cn).

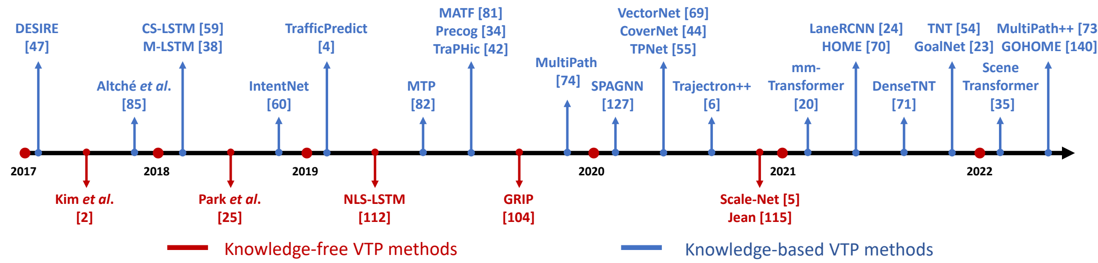


Fig. 1: Representative DL-based VTP Methods in Recent Years.

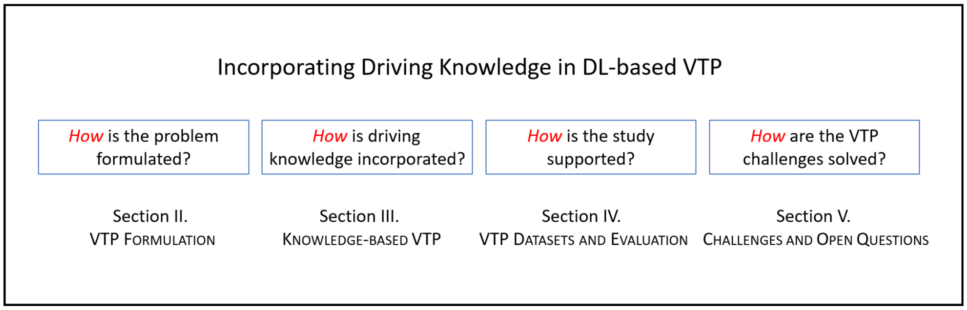

Incorporating Driving Knowledge in DL-based VTP
How is the problem formulated?
How is driving knowledge incorporated?
How is the study supported?
How are the VTP challenges solved?
Section II. VTP Formulation
Section III. Knowledge-based VTP
Section IV. VTP Datasets and Evaluation
Section V. Challenges and Open Questions

Fig. 2: Structure of this Review.

that incorporats driving knowledge into the VTP framework.

However, to the best of our knowledge, none of the existing VTP-related reviews directly answer the following questions:

1) *How is the knowledge-based VTP problem formulated?*
2) *How is driving knowledge incorporated in deep VTP model learning?*
3) *How is knowledge-based VTP supported by dataset and how is it evaluated?*
4) *How are the main VTP challenges solved and what are the open questions?*

To answer the questions, the following steps are taken in the rest parts of this paper as illustrated in Fig.2. First, a taxonomy of literature is presented in Section II from the perspective of how driving knowledge is addressed in VTP problem formulation. The knowledge-based VTP methods are then detailly analyzed in Section III to dig out how driving knowledge is incorporated into deep model learning. Next, the commonly-used VTP datasets and evaluation metrics are reviewed in Section IV to identify foundational support for knowledge-based VTP research. Concerning the main VTP challenges from the interaction at dynamic traffic scenes, Section V reviews how these challenges are solved and discusses the open questions leading to future studies. Finally, the concluding remarks are drawn in Section VI.

We claim our paper to have the following contributions:

1) A taxonomy of the literature is presented from the problem formulation's perspective, in which the studies are divided into the general VTP, incorporating driving knowledge in VTP, and introducing a secondary task to VTP, i.e., intention prediction, based on driving knowledge.
2) A detailed analysis of the knowledge-based VTP approach is presented to reveal which driving knowledge and how they are incorporated into the different stages of deep model learning, such as driving state encoding, trajectory decoding, and intention prediction.
3) A comprehensive review of VTP datasets and evaluation metrics is presented from the perspective of what driving knowledge the datasets provide, how they are used in VTP studies, how VTP results are assessed, and the limitations of the support to the study.
4) An in-depth review of how the main VTP challenges in dynamic traffic scenes, namely interaction-awareness and multimodality issues, are addressed by the literature works. Open questions and other concerns are discussed from the perspectives of driving knowledge, datasets, and evaluation leading to future research.

II. VTP Formulation

From the problem formulation's perspective, the VTP studies can be divided into three groups: the general knowledge-free VTP, incorporating driving knowledge in VTP, and introducing a secondary task to VTP, namely intention prediction, based on driving knowledge. Below we review each VTP formulation and the representative literature works, seeking the answer to the question ”*How is the knowledge-based VTP problem formulated?*”.

A. General Formulation of VTP

The problem of vehicle trajectory prediction is generally formulated in a probabilistic way as estimating a conditional distribution[7]

$$P(Y|X) \tag{1}$$

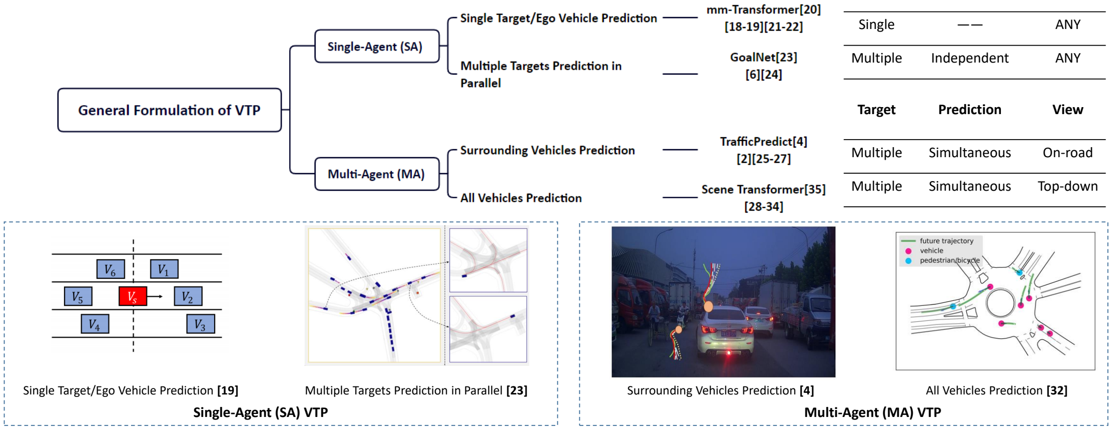

General Formulation of VTP
Single-Agent (SA)
Multi-Agent (MA)
Single Target/Ego Vehicle Prediction
Multiple Targets Prediction in Parallel
Surrounding Vehicles Prediction
All Vehicles Prediction
mm-Transformer[20] [18-19][21-22]
GoalNet[23] [6][24]
TrafficPredict[4] [2][25-27]
Scene Transformer[35] [28-34]
Single —— ANY
Multiple Independent ANY
Target Prediction View
Multiple Simultaneous On-road
Multiple Simultaneous Top-down
V_6 V_1 V_5 V_s V_2 V_4 V_3
future trajectory
vehicle
pedestrian/bicycle
Single Target/Ego Vehicle Prediction [19]
Multiple Targets Prediction in Parallel [23]
Single-Agent (SA) VTP
Surrounding Vehicles Prediction [4]
All Vehicles Prediction [32]
Multi-Agent (MA) VTP

Fig. 3: Single/Multi-Agent Taxonomy in General Formulation of VTP.

where X and Y denote the input and output of the problem. In the scope of VTP, X is the driving state observation of the target vehicle(s) and scene context up to the prediction time T, e.g., $X = \{X_{T-\tau_s}, ..., X_T\}$, while Y is the predicted trajectories of the target vehicle(s) in a future time horizon, e.g., $Y = \{Y_{T+1}, ..., Y_{T+\tau_e}\}$. The output can be deterministic or probabilistic trajectories. τ_s and τ_e represent the length of observation and prediction, respectively.

According to the different definitions of X-Y, VTP methods can be roughly divided into Single-Agent (**SA**) and Multi-Agent (**MA**) VTP methods, as Fig.3 illustrated.

*1) **Single-Agent VTP**:* In a typical Single-Agent VTP method, there is a specific target vehicle TV whose future trajectory is to be predicted. Several surrounding vehicles SV are observed from TV and could potentially affect its driving, while the future trajectories of SV are not to be explicitly predicted. In some VTP methods, the target vehicle TV and the ego vehicle EV refer to the exact vehicle to be predicted. The problem of a single-agent VTP is formulated as $X_t = (X_t^{TV}, X_t^{SV}, X_t^{Add})$, where $t \in [T - \tau_s, T]$, X_t^{TV} and X_t^{SV} describe the states of the target and surrounding vehicles at time t, respectively, and X_t^{Add} represents the additional scene contexts that are relevant to the inference of TV's future trajectory. The output of such formulation is the predicted future trajectory of TV, i.e., $Y = \{Y_{T+1}^{TV}, ..., Y_{T+\tau_e}^{TV}\}$.

According to the specific setting of the prediction task, SA can be further divided into two subgroups. **Single Target/Ego Prediction:** Methods in this subgroup have only one target to be predicted. [18] uses a Recurrent Neural Network (RNN) to build a trajectory prediction framework for a single target vehicle in a roundabout scene. [19] exploits a Long Short-Term Memory network (LSTM) to encode the trajectory input of the target vehicle and its surrounding vehicles in the road scene to predict the future trajectory of the target vehicle. [20] uses multiple Transformers to model the target vehicle's motion information, map information, and interaction information with surrounding vehicles and outputs the probability estimate of the target vehicle's trajectory. [21] builds a cross-model target vehicle trajectory prediction framework, which can effectively predict the target vehicle trajectory under different sensor data. [22] designs a novel 3D spatial-temporal feature representation for the target vehicle's prediction; the different neighborhood agents are treated equally during inference. **Multiple Targets Prediction in Parallel:** Methods in this subgroup have multiple target vehicles to be predicted, but the trajectory of each target vehicle is predicted independently and the potential influence between the future trajectories of different target vehicles is ignored. [23] obtains a unified environment encoding for all target vehicles in the scene and then applies the same prediction model to all target vehicles to generate their trajectories. [6] constructs a topological graph for different types of traffic participants and extracts local map features and interaction information of surrounding participants for each individual target to infer its future trajectory. [24] builds a unified feature representation for all vehicles and roads in the scene and extracts the relevant local information for inference when predicting the trajectory of each vehicle. As the possible conflicts between the future trajectories of different target vehicles are ignored, methods in this formulation can be regarded as the combination of multiple independent single-agent VTP, so we still group it into the single-agent VTP method.

*2) **Multi-Agent VTP**:* In Multi-Agent VTP methods, there is no distinction between TV and SV as in Single-Agent VTP. The future trajectories of all relevant vehicles RV (including other types of traffic participants at mixed traffic scenes) will be predicted jointly. The problem of a multi-agent VTP is formulated as $X_t = (X_t^{RV}, X_t^{Add})$, where $t \in [T - \tau_s, T]$, X_t^{RV} describes the states of all relevant traffic participants at time t. The output of such formulation is the predicted future trajectories of all relevant vehicles RV, i.e., $Y = \{Y^{RV_1}, ..., Y^{RV_m}\}$, where for the jth RV, it predicts a trajectory $Y^{RV_j} = \{Y_{T+1}^{RV_j}, ..., Y_{T+\tau_e}^{RV_j}\}$ during a future time horizon $[T + 1, T + \tau_e]$.

According to the perspective of data acquisition and prediction tasks, the multi-agent methods can be further divided into two subgroups. **Surrounding Vehicles Prediction:** Methods in this subgroup usually obtain information about the environment and other surrounding vehicles from the perspective of the ego vehicle with multimodal sensors. The trajectory of the

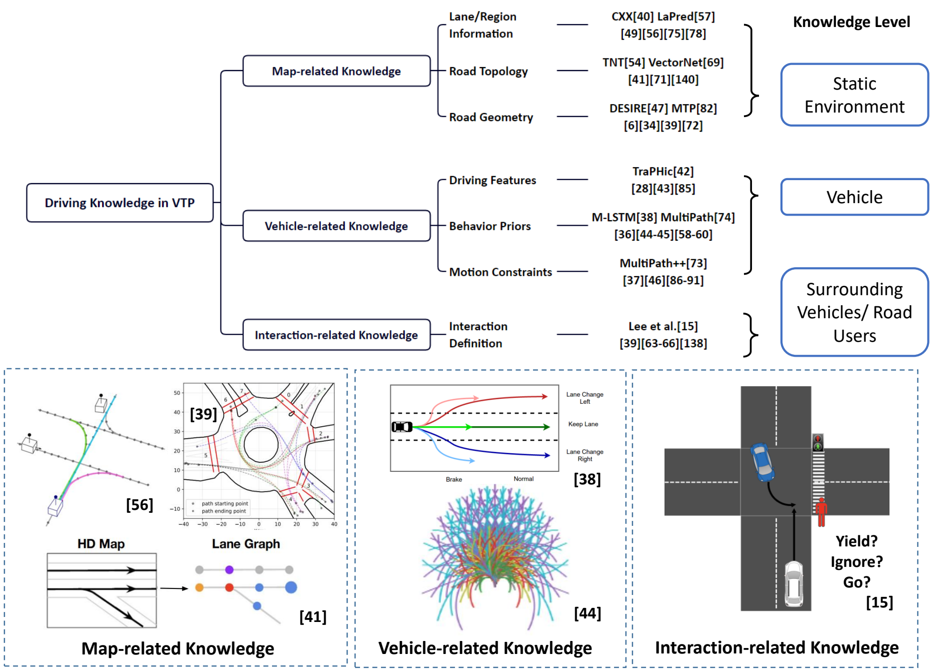


Fig. 4: Driving Knowledge in VTP.

ego vehicle itself is not to be predicted. [2] and [25] build a grid map of the environment from ego vehicle and predict the trajectories of all surrounding vehicles on it. [26] and [4] predict the trajectories of other traffic participants in the scene directly from the ego-view data. [27] projects the monocular camera videos to bird's eye view, then designs a framework to simultaneously predict future instance segmentation and probabilistic trajectory. **All Vehicles Prediction:** Methods in this subgroup usually obtain the observation of the scene from a top-down view (such as aerial video or surveillance camera). In this kind of data, there is no specific target vehicle; all vehicles are of the same intelligence level and have equal priority to be predicted. [28] uses a structural LSTM to encode information and predict the trajectory for each vehicle in the scene, while [29] and [30] extract the environmental information for all vehicle's trajectory prediction. [31] takes the behavior of each agent in the scene into the prediction framework to guarantee effective trajectory prediction in mixed traffic environments with different types of traffic participants. [32] uses a similar method to model the relationship of different types of agents in the scene through a heterogeneous graph, on which the states of all agents are encoded, and their future trajectories are predicted. [33] builds a Social ODE (Ordinary Differential Equations) that models temporal agent dynamics and agent interactions, which leverages Neural ODEs for continuous modeling and output. [34] designs a data-driven likelihood-based multi-agent prediction framework that uses latent variables to describe the ways of reacting under scene observation. The multi-agent trajectories are forecasted conditioned on these latent variables. [35] proposes a unified Scene Transformer that can produce consistent futures between agents. The experiments show its effect on both marginal (SA) and joint (MA) prediction.

B. *Incorporating Driving Knowledge in VTP*

Vehicle driving follows many rules. First, vehicles drive on the lane and follow traffic regulations in most situations. Therefore, maps that contain information relevant to driving, such as lanes, routes, road boundaries, etc., are helpful in predicting the future trajectories of vehicles [36]. Secondly, the vehicle's non-holonomic kinematics constrain its motion. By using the vehicle's kinematic model, the searching space of the future trajectories can be greatly reduced [37]. Thirdly, different driving maneuvers and interactions with other vehicles yield different trajectories. By defining the types of maneuvers such as lane change, lane keeping, car following, etc., [38] and interactions such as going, yielding, ignoring, etc., [15], this kind of driving knowledge can be used to model and infer multimodal trajectories with higher reliability.

Making use of driving knowledge in the problem of vehicle trajectory prediction, formula (1) is converted to the following

$$P(Y|X,M) \tag{2}$$

where M denotes driving knowledge. In this paper, we categorize the driving-related information from three different levels into three groups of knowledge: The information of the static environment level is categorized as Map-related Knowledge; the priors about the vehicle itself, including its behavior, policy, and vehicle models, etc., are grouped into Vehicle-related Knowledge; the interaction patterns considering other vehicles and road users level are categorized into Interaction-related Knowledge, as illustrated in Fig.4.

1) Map-related Knowledge: Whether used explicitly or implicitly, maps can provide rich environment information for autonomous vehicles, such as road geometry [39], lane/region information [40], road topology [41], etc. This information plays a vital role in the VTP process, especially in complex scenarios where the autonomous vehicle has multiple potential directions, such as ramps, intersections, and roundabouts. The incorporation of map-related knowledge is normally accomplished through Graph Neural Networks (GNN) or Convolutional Neural Networks (CNN).

The use of map-related knowledge enables VTP methods to output trajectory prediction results that match the actual structure of the road. The detailed discussion on how to incorporate map-related knowledge into VTP framework is further presented in Sec.III-B.

2) Vehicle-related Knowledge: In the scope of VTP, vehicle-related knowledge can be further divided into three subgroups: **Driving Features:** The classic driving features in traditional ITS/AV research can also be used in DL-based frameworks for trajectory modeling. Except for the vehicle trajectory itself, features such as the relative speed [42], relative distance [28], and time to collision (TTC) [43] are taken as extra features in literature, which are usually modeled via Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM) for temporal feature encoding. **Behavior Priors:** Efficient mining of historical trajectory data enables the acquisition of necessary behavior priors and patterns, which is vital for considering or modeling the behavior of the target vehicle for reliable prediction. This type of knowledge includes maneuver prior [38], anchor trajectory [44], driving policy [45], etc., and is often used for the design of secondary tasks, which will be expanded in Sec.II-C and Sec.III-D. The networks considering behavior as a secondary task are usually designed with multiple branches. **Motion Constraints:** Considering constraints for driving such as vehicle model [46], kinematics [37], traffic rules [36] enables more realistic output. Such knowledge is often used during Trajectory Decoding for temporal modeling, which will be further discussed in Sec.III-C.

The vehicle-related knowledge motivates the model to consider behavior and constraints of trajectory. Thus more realistic and reliable trajectory can be obtained.

3) Interaction-related Knowledge: Interaction with other road users is another crucial factor in VTP. For example, in a merge scenario, whether the ego vehicle chooses to yield or pass a vehicle in the target lane will result in very different driving trajectories. However, the type of interaction is a hidden state in real-time inference, and there is no uniform definition of such events. In the literature, researchers use a posteriori approach to determine the labels of interaction events, i.e., complete trajectories are used to define the labels of interaction events, while the categories of labels are defined using driving knowledge. For example, 'Ignoring', 'Going', and 'Yielding' are defined in [47] for pair-wise interaction labels, while [39] define the interaction according to the conflict between trajectories under the roundabout scenario.

Such knowledge helps VTP method to better understand the underlying interaction via interactive event prediction through a multi-branch network or explicit interaction modeling through a multi-stage network, thus making the prediction results collision-free and safe. A further discussion of interactive events is presented in Sec.II-C, and a detailed introduction to interaction modeling locates in Sec.V-A.

C. Introducing Secondary Task to VTP based on Knowledge

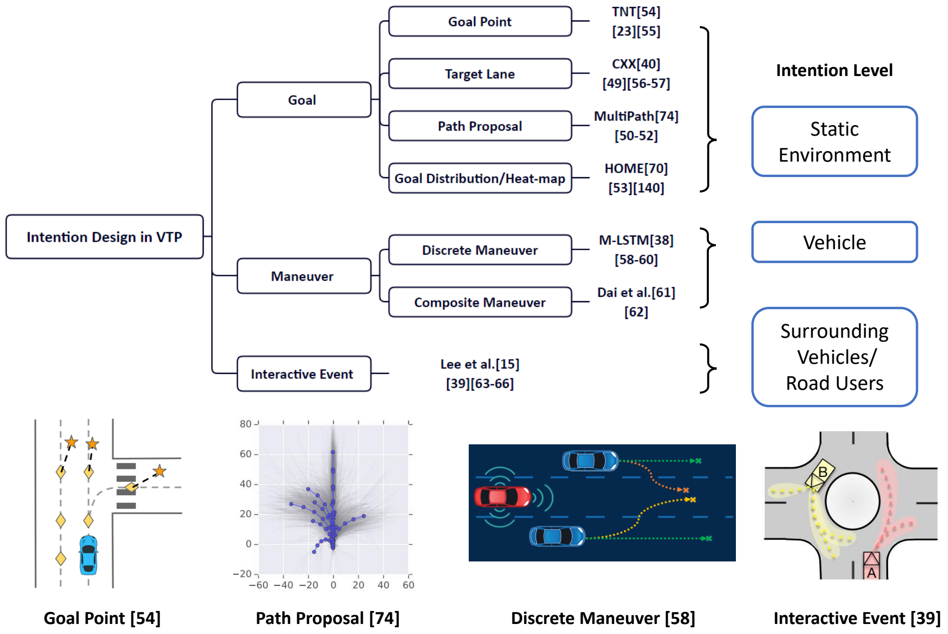
Driving has been identified in a hierarchical framework with three levels of processes, i.e., operational processes that involve manipulating control inputs for stable driving, tactical processes that govern safe interactions with the environment and other vehicles, and strategic processes for route and mission planning [48]. Such driving processes are guided by tasks at different levels, which can not be directly observed by other vehicles. Based on driving knowledge M and the designed task H, formula (1) is expanded to two consecutive inference processes

$$P(Y|X,M)=\sum_{H}P(Y|X,H,M)P(H|X,M) \qquad (3)$$

where $P(Y|X,H)$ is the main task of vehicle trajectory prediction conditioned on a given H, while $P(H|X)$ is the task of inferring H conditioned on the observation X. M represents driving knowledge described in Sec.II-B.

In this paper, we maintain the main task as trajectory prediction itself, and all the processes that helps to output prediction are regarded as the secondary task. In the scope of VTP, the secondary task are often modeled as inferring the target vehicle's driving intention for reliable trajectory prediction. Hereinafter, we use the keyword **Intention Prediction** instead. To match the knowledge level in II-B, we divide the most relevant intention design for vehicle trajectory prediction into three groups: Goal for environmental level, Maneuver for vehicle level, and Interactive Event for other vehicles level, as illustrated in Fig.5.

1) ***Goal:*** In recent VTP methods, goal-based prediction has become one of the most important VTP formulations. Goal-based VTP methods first generate goal proposals based on the environment information, then forecast possible trajectories to those predicted goals. In literature, the definition of goal has taken many forms, such as goal point [23], target lane [40], [49], path proposal [50], [51], [52], and goal point in the form of distribution/heat map [53]. **Goal Point:** [23] and [54] generate multimodal goal points based on the environment inputs, [55] further formulate several trajectory proposals around the predicted endpoint. **Target Lane:** [40] and [56] design a secondary task to predict the lane of the target vehicle, [57] uses a similar approach to encode each lane's information independently, while [49] predicts the lane segment that is more likely to be the destination of trajectory. **Path Proposal:** In [50], several feasible traversals are planned on the topological scene graph. [51] uses reachable paths of the target vehicle as the candidate set for prediction, while [52] exploits a lightweight radial grid-based and kinematics-based representation to generate the potential path set. The final outputs of these works are calculated based on the sample



Goal Point [54]

Path Proposal [74]

Discrete Maneuver [58]

Interactive Event [39]

Fig. 5: Driving Intention Design in VTP.

paths. **Goal Distribution/Heat map:** [53] changes the task of VTP to evaluate a heatmap for the input scenario. The trajectory of the target vehicles is sampled from the estimated probabilistic heatmap.

Due to the existence of visible intermediate results, goal-based VTP brings more interpretable output. However, it relies on a more complex system design.

*2) **Maneuver**:* Maneuver is the specific behavior performed by the vehicle, such as lane-keeping, lane-change, etc. As different maneuvers lead to diverse trajectories, the well-predicted maneuver can guide trajectory prediction as a vital reference. Under this circumstance, a lot of works attempt to define maneuvers in a discrete way or a composition way for trajectory prediction. **Discrete Maneuver:** [58] considers three different maneuvers, {Left Lane Change, Right Lane Change, and Lane Keeping}, while [38] and [59] propose a division of maneuvers into six discrete categories according to the lateral and longitudinal movements of the ego vehicle. [60] defines 8-way maneuver as {Keep Lane, Turn Left, Turn Right, Left Lane Change, Right Lane Change, Stopping/Stopped, Park, and Other}. **Composite Maneuver:** [61] consider an initial maneuver division of 5 categories, which is {Going straight, Left lane change, Merge into the left lane, Right lane change, and Merge into the right lane}. The authors then design a semi-supervised And-or-Graph (AOG) to learn a detailed representation of these roughly designed maneuvers, called sub-maneuvers. [62] regard complex driving behavior as the combination of basic maneuvers. The range of basic maneuvers is defined as {Lane Change, Turn, Stop, Acceleration, Deceleration, Others / Following Lane} and can be automatically labeled by rules.

Their prediction results indicate that the hierarchical workflow of trajectory prediction with maneuver recognition improves the interpretability of the VTP process.

*3) **Interactive Event**:* Interactive event is the specific definition of interaction behavior between vehicles and other road users. In frequently interacting scenarios, the accurate modeling of interaction is crucial for VTP methods. Therefore, many researchers try to design driving intention as Interactive Event to improve the reliability and safety of VTP results. [15] and [63] use discrete types to describe the underlying interaction in the intersection. The pair-wise interaction is labeled as ’Ignoring’, ’Going’, or ’Yielding’ according to the target vehicle’s behavior during the interaction procedure. [64] defines two types of directed pair-wise interaction, {Follow, Yield}, the model has to simultaneously predict the interaction label if such kind of interaction exists. [39] defines the interaction type based on the reference paths of two interacting vehicles in the roundabout. The interact pairs are taken as guidance to generate conflict-free trajectories for all vehicles in the scene. [65] builds a hierarchical VTP framework to first predict the interaction type of different traffic participants from the ego vehicle’s view, then forecast their future trajectory according to the interaction type. [66] believes that the vehicle acceleration is directly reflected by the interaction. The author uses a sub-network to predict the acceleration information of the target vehicle in the current scene and obtains the final speed and trajectory by integration.

The results of the above work show that introducing interactive events as intention can achieve more reliable predictions in interactive scenarios.

In order to investigate the development trend of VTP methods in different formulations, we denote the general knowledge-free VTP methods as ***Fo1***; the methods that introduce driving knowledge but do not include a secondary task is marked as ***Fo2***; the methods that introduce a secondary task based on knowledge is denoted as ***Fo3***, as illustrated

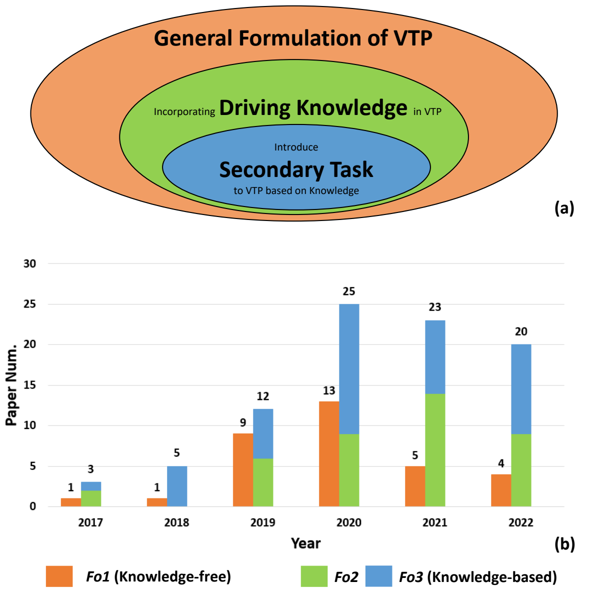


Fig. 6: Proportion of Different Fomulation VTP Methods.

in Fig.6(a). The development trend and proportion of VTP methods in different formulations are shown in Fig.6(b).

From the figure, it can be seen that the proportion of knowledge-based VTP methods (***Fo2*** & ***Fo3***) have been on the rising trend in the literature in recent years, indicating that more researches tend to introduce a variety of driving knowledge into the VTP framework.

III. Knowledge-based VTP

Seeking the answer to the question ”*How is driving knowledge incorporated in deep VTP model learning?*”, the general workflows of the three VTP formulations are first reviewed in this section, followed by a detailed analysis of the knowledge-based VTP methods to dig out which driving knowledge and how they are incorporated into the different stages of deep VTP model learning, such as Driving State Encoding, Trajectory Decoding, and Intention Prediction.

A. General Workflow of VTP

The traditional non-DL-based VTP methods usually follow a two-stage manner: first, extract hand-crafted driving features; then, model the output in a generative way (sampling from the estimated distribution) or a discriminative way (choosing optimal trajectory from candidate set) [3]. When it comes to the DL-based methods, the modules are replaced by neural-network-based modules, respectively.

As illustrated in Fig.7, the general VTP workflow of ***Fo1*** can be straightly divided into two sequential modules, one is denoted as Driving State Encoding and the other is called Trajectory Decoding. The Driving State Encoding module takes all the sensory input to formulate a latent encoding that represents the driving state of the current input scenario. Then the latent encoding is fed to the Trajectory Decoding module to generate a future trajectory as output. [5] build a scene graph based on the location of vehicles in the scene, then use an edge-enhanced GCN together with an LSTM as the encoding module. An LSTM-based decoder is used to generate trajectory prediction results.

TABLE I: Knowledge Used in Different Modules

What \ When	Driving State Encoding	Trajectory Decoding	Intention Prediction
Map-related Knowledge	DESIRE [47] VectorNet [69] Trajectron++ [6] CXX [40] HOME [70] TNT [54] DenseTNT [71] Scene Transformer [35] **(48)**		DROGON [72] **(3)**
Vehicle-related Knowledge	TraPHic [42] **(4)**	MultiPath++ [73] Social ODE [33] **(14)**	CS-LSTM [59] M-LSTM [38] MultiPath [74] CoverNet [44] IntentNet [60] **(26)**
Interaction-related Knowledge			Lee *et al.* [15] **(7)**
Paper Sum	**88** [*Fo2*+*Fo3*]	**88** [*Fo2*+*Fo3*]	**48** [*Fo3*]

# The values in this table are the numbers of VTP works incorporating the corresponding knowledge in each stage.

Fo2 extends ***Fo1*** by adding driving knowledge into the workflow. In [67], the author designed several agent encoders to obtain the state encoding of the target vehicle and surrounding vehicles, respectively, and the map knowledge was incorporated into the state encoding through an independent branch. In the trajectory decoding stage, the authors use two sub-modules to simultaneously generate and evaluate the final multimodal trajectory prediction results. [68] uses a similar structure, which samples the map as waypoints and uses the attention mechanism to model the connection between the target vehicle, surrounding vehicles, and road waypoints to generate the predicted trajectory.

Fo3 methods usually design an intermediate module based on driving knowledge to obtain better interpretable prediction results. This module is denoted as Intention Prediction. [38] designs a maneuver prediction branch as the intention prediction task, the predicted maneuver is then combined with decoding outputs to formulate multimodal trajectories. [44] employs an anchor trajectory prediction task, which transforms the regression task of trajectory prediction into a classification task. Thus the independent trajectory decoding module is no longer needed in the whole workflow.

The workflows of ***Fo2*** and ***Fo3*** in Fig.7 belong to knowledge-based VTP, which are composed of Driving State Encoding, Trajectory Decoding, and Intention Prediction modules. Statistics were performed in Tab.I to uncover which driving knowledge was incorporated into which VTP modules, in which the representative model names and the number of works for each stage are presented. In the rest of this section, we analyze in detail how the driving knowledge is incorporated into each VTP module to accomplish deep model-based inference.

B. Knowledge in Driving State Encoding

In the Driving State Encoding module, most knowledge-based VTP methods consider maps as additional knowledge

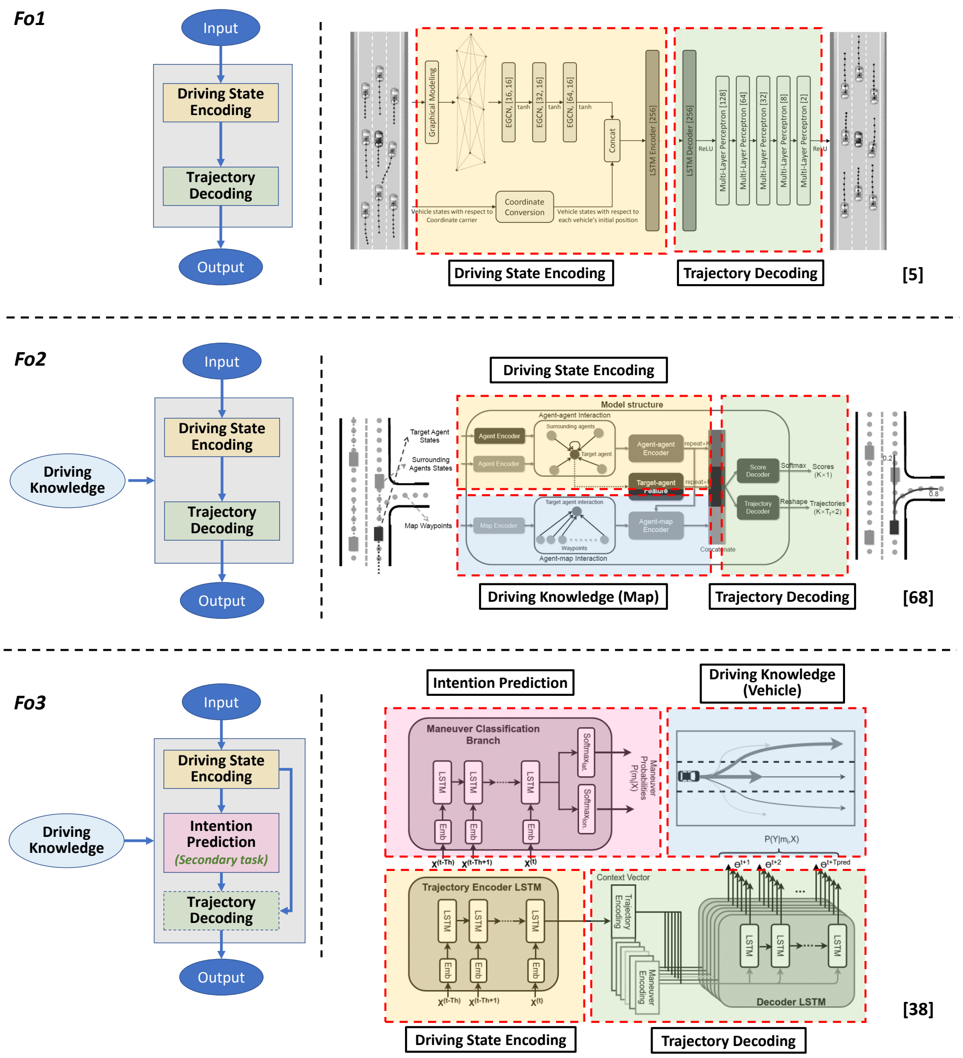


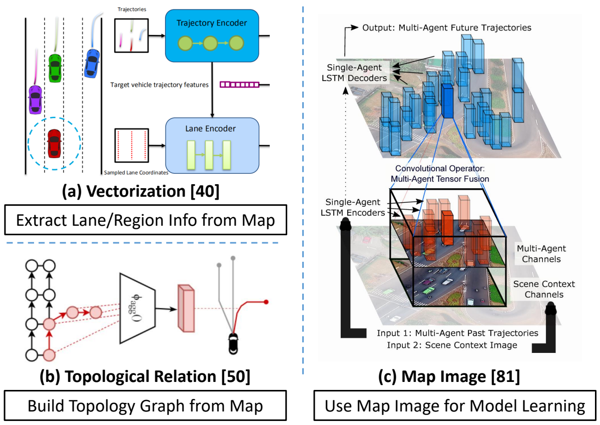
Fig. 7: General VTP Workflow in Different Formulation.

information. The usage of map-related knowledge can be further classified into three groups, including Vectorization, Topological Relation, and Map Image, as Fig.8 shows.

1) Vectorization: One straightforward way to exploit map knowledge is to explicitly vectorize driving-related information, such as lane, region, etc., for neural network learning. [36] extracts information such as the coordinates of the lane centerline and road boundaries from the map, and constructs a traffic cost map of the target vehicle in the current scene as important prior knowledge for target vehicle prediction. [75] uses the road structure information provided by the map to eliminate possible trajectories that are out of the drivable area in the prediction results. This heuristic design greatly reduces the probability of wrong results. [76] encodes the lane information independently, uses a self-attention module to evaluate the weight of different lane encoding, and finally formulates the whole environment encoding. [77] designs a tailored-KF to combine the lane feature and vehicle states, in which the road shape and the distance between the lane centerline and the vehicle are taken as the feature of each lane. [69] heuristically abstracts the map into vectors with different semantic attributes (such as lanes, sidewalks, etc.), which transforms complex and dense rasterized map information into abstract and concise vectorized information to achieve efficient feature encoding and accurate inference. [78] designs an Off-Yaw Rate loss based on lanes as an extra measure of the trajectory's orientation. The proposed loss can help VTP models to learn better result that fits the direction of the road. [79] builds an integrated grid map of different scales for the scene, in which lines and road information are defined as different values.

The extra lane and region information enables the VTP method to capture key driving features of the scenario, thus making it possible to predict trajectories that fit the road geometry.

2) Topological Relation: Due to the topological nature of the traffic scene, many topological graph-based methods have been used to solve the problem of VTP in recent years. Some researchers try to regard road segments in maps and vehicles as nodes with different attributes to build a scene topology



(a) Vectorization [40]

Extract Lane/Region Info from Map

(b) Topological Relation [50]

Build Topology Graph from Map

(c) Map Image [81]

Use Map Image for Model Learning

Fig. 8: Incorporating Map-related Knowledge in VTP

graph so that the interaction between vehicles, as well as the association between vehicles and the static scene, can be modeled at the same time. [53] uses a graph network to model the topological information of each road segment in the scene, so as to predict an independent trajectory heatmap for each target vehicle. Similarly, [41] constructs a lane graph, modeling each lane segment in the scene as a node, and the edges between nodes represent the accessibility between lane segments. In this form, the feature encoding of the scene can be updated on the lane graph by graph neural network technology. [50] further aggregates the vehicle feature vector into the lane graph node encoding. The author samples the reference path on the constructed lane graph, which is used as a priori for trajectory decoding. [80] uses map information to encode vehicle states, the author proposes a rotation-invariant scene representation and convolution method, and verifies the effectiveness of this method in dealing with input changes.

Digging the topological relation in the map enables the effective modeling of the relation between vehicles and the static environment information.

3) Map Image: Benefiting from the development of research related to convolutional neural networks and image processing technology, some researchers try to use neural networks to directly extract the information required for VTP from the raw map input. [47] uses a generative model to predict the trajectory of the target, which uses a separate branch to extract the environmental information contained in the top-down view map. [81] directly uses the submap of each vehicle's location as the feature encoding of the vehicle, and then aggregates the information encoding of all vehicles to obtain the scene encoding. [70] uses a separate branch to process Bird Eye View(BEV) map input, and the output is also organized as a BEV heat map aligned with the original input. [45], [82], [83] combine the input BEV map information and the historical trajectory of the target to predict the trajectory of each target vehicle in the scene. [84] combines the image of the scene with the Social Tensor containing different vehicle information, the image feature encodings and the vehicle state encodings are aggregated to obtain more accurate trajectory prediction results.

Using maps in this way can straightly extract rich information through well-designed CNN and output prediction results in the corresponding coordinates to the input.

In addition to the rich map-related knowledge, a small number of studies have tried to incorporate vehicle-related knowledge (driving features) in the driving state encoding stage. [42] adds the vehicle's turning radius and the relative speed with other traffic participants into the vehicle state encoding. [28] explicitly extracts classic driving features such as the distance to the lane boundary, the relative distance of the preceding vehicle, and the relative speed of the preceding vehicle, and adds them to the vehicle state encoding. Time To Collision (TTC) is an important driving characteristic in the field of intelligent transportation for car-following behavior modeling. [43] regards this feature as additional information into the vehicle state encoding to include the driving behavior pattern information. Similarly, in addition to TTC, [85] also uses vehicle types as extra features, ranging from {motorcycle, car, truck}. Since different types of vehicles behave differently, explicitly considering vehicle types can help accurate trajectory prediction in dense traffic.

These methods attempt to make the learning process of neural networks more efficient by extracting classic driving/traffic features into the model. However, the discussion on the importance of different driving features for dl-based VTP models is limited.

C. Knowledge in Trajectory Decoding

Among the included knowledge-based methods, a small group of works attempt to introduce vehicle-related knowledge in the trajectory decoding module, such as kinematics and vehicle models, etc. [37] and [86] add kinematic constraints during decoding, in which the bicycle model is employed to model the movement of the target vehicle. Similarly, [46] explicitly constructs the vehicle kinematics model in the trajectory decoding module to generate more realistic trajectory results. [87] separately predicts the longitudinal and lateral movements of the target vehicle, in which the lateral position deviation is directly used for output while the longitudinal trajectory is calculated through the predicted acceleration. [73] similarly takes the acceleration and heading change rate as the output of the model, and the final trajectory is integrated based on them. [88], [89], [90] employ the model predictive control (MPC) to increase the feasibility of the prediction result, in which the vehicle model is taken as the constraints of model parameters. [91] combine the predicted behavior with a model-based trajectory decoder, in which a cubic spiral-based geometric path generation and an MPC-based speed profile prediction are exploited for the final trajectory output. Specifically, for ***Fo3*** methods, the predicted intention can be regarded as explicit guidance for decoding. For example, [50] samples reference paths from the scene graph as the prior for trajectory decoding, while [23] outputs the final trajectory based on the reference path proposed from the map.

Although there are few methods incorporating knowledge in the trajectory decoding module, we can still see the effect of

TABLE II: VTP Datasets Overview

Dataset	Overview			Content		
	Year	Scenario	Viewpoint	Duration	Traj. Stat.	Extra Info
NGSIM (I-80 & US-101)	2005	highway	top-down	45min (each)	3k vehicles (I-80)	Lane ID, Vehicle Types
SDD	2016	campus	top-down	5h	19k objects (1.3k cars)	Vehicle Types
HighD	2018	highway	top-down	16.5h	110k vehicles	
Argoverse	2019	urban	on-road	320h	333k segments	HD map
nuScenes	2019	urban	on-road	5.5h	1k segments	HD map, Vehicle Types
Apolloscape	2019	urban	on-road	103min	82k objects (60k vehicles)	Vehicle Types
INTERACTION	2019	roundabout, un-signalized intersection, signalized intersection, merging and lane changing	top-down	16.5h	40k vehicles	HD map
BLVD	2019	urban	on-road	6h	5k instances	Interactive Event
Lyft Level 5	2020	urban	on-road	1118h	170k segments	Road Geometry, HD map,Traffic Signal States
WOMD	2021	urban	on-road	570h	104k segments	HD map, Traffic Signal States

these methods on improving the practicability and feasibility of the prediction results.

D. Knowledge in Intention Prediction

In Section II-C, the scope of the intention design has been discussed. In this part, we clarify the connection between intention design and the knowledge used in the intention prediction module.

From Fig.5 in Section II-C and Tab.I the conclusion can be drawn that:

a) The knowledge used in intention prediction module is basically corresponding and relates to the intention design. [38][59][92] set the intention as maneuvers, and the corresponding knowledge is the vehicle-related knowledge (behavior prior).
b) The driving knowledge used in intention prediction is diverse, as for the same driving intention, different knowledge might be used. [39] sets the Interactive Event as intention, and it first extracts entrance and exit from the map, then generates reference paths based on it, and finally obtains the label of the interactive event according to the relation between vehicles' reference paths. Therefore, in the stage of intention prediction, three kinds of knowledge are all incorporated.
c) The driving knowledge introduced in different intention designs may be similar. Intentions in [50] and [45] are not exactly the same (path proposal v.s. goal point), but both works try to introduce driving policy for intention prediction. Both [74] and [44] design the task of anchor trajectory prediction, and the only difference between their intention prediction modules is that [74] extracts anchor trajectories from historical trajectories clustering, while [44] uses the vehicle model to generate the candidate set of trajectories.

In all, how to use different types of driving knowledge to design rich intention prediction tasks to enhance the feasibility and interpretability of VTP results is still one of the issues worthy of attention and research.

In this section, we summarize how different modules of the VTP method introduce knowledge. Among them, most of the works use map-related knowledge in the Driving State Encoding module; a small number of works use vehicle-related knowledge in the Trajectory Decoding stage as extra guidance or constraints; among works involving intention prediction, methods introduce various types of knowledge to design different tasks for better interpretable prediction results. In all, the current VTP methods that introduce driving knowledge have achieved specific research results, while how to fully use the rich driving knowledge in each prediction module to improve the prediction ability is still one of the important research directions in the field of VTP.

IV. VTP Datasets and Evaluation

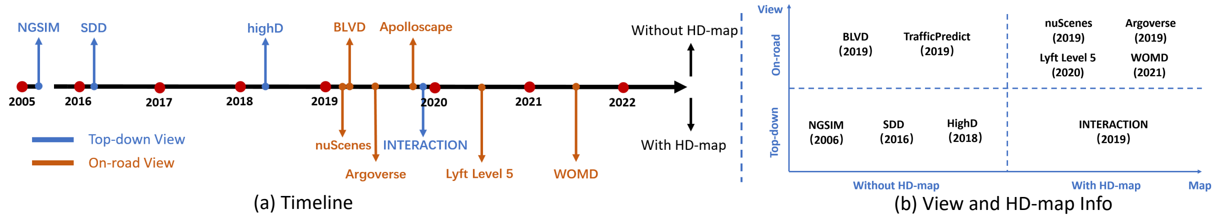
This section answers the question "*How is knowledge-based VTP supported by dataset and how is it evaluated?*". We first overview the public VTP datasets and then analyze what knowledge is contained in the datasets and how are they used in VTP studies. Finally, the primary methods and metrics for VTP evaluation are reviewed and discussed.

A. Overview of VTP datasets

According to the data acquisition view point, the widely-used VTP datasets can be divided into two groups: 1) Top-down View dataset: the data are collected using drones or fixed surveillance cameras from a top-down view. 2) On-road view dataset: the data are collected through an ego vehicle with onboard sensors, such as Lidars and cameras.

1) Top-down View Datasets: The top-down view datasets are widely used in traditional ITS research, such as traffic simulation, behavior modeling, and trajectory analysis. In these datasets, the original data are often captured from drones or fixed surveillance cameras in video format, then the trajectories of all vehicles in the scenario are extracted from the original data. Although these data are collected in fixed scenarios, as a large number of vehicle trajectories with various driving patterns can be extracted from them, they can still be used in DL-based VTP research. The typical top-down view datasets for VTP are NGSIM [94], SDD [95], HighD [96] and INTERACTION [97], as briefly introduced in Tab.II.

2) On-road View Datasets: With the development of onboard sensors and perception techniques for autonomous vehicles, several large-scale on-road view datasets are proposed in the field of autonomous driving. The on-road view datasets



(a) Timeline

(b) View and HD-map Info

Fig. 9: Development Trend of VTP Datasets

TABLE III: Knowledge Usage in VTP Datasets

Dataset	Map-related Knowledge		Vehicle-related Knowledge		Interaction-related Knowledge	
	Knowledge	Used by	Knowledge	Used by	Knowledge	Used by
NGSIM	Lane ID	Kim *et al.* [76] (Lane) Mo *et al.* [32] (Road Geometry)	Vehicle Type	Altché *et al.* [85] Hou *et al.* [28] (Driving Features) CS-LSTM [59] M-LSTM [38] (Behavior Priors) Xin *et al.* [87] (Motion Constraints)	—	Ju *et al.* [66]
SDD	Aerial Photo	TNT [54] (Road Topology) DESIRE [47] (Road Geometry)	Vehicle Type	MultiPath [74] (Behavior Priors) Li *et al.* [37] (Motion Constraints)	—	—
HighD	—	Mozaffari *et al.* [12] (Road Geometry)	—	Tang *et al.* [88] (Behavior Priors) Social ODE [33] (Motion Constraints)	—	—
Argoverse	HD-map	CXX [40] MultiPath++ [73] (Lane) TNT [54] DenseTNT [71] VectorNet [69] (Road Topology) HOME [70] mm-Transformer [20] (Road Geometry)	—	Chandra *et al.* [93] (Behavior Priors) MultiPath++ [73] (Motion Constraints)	—	—
nuScenes	HD-map	Greer *et al.* [78] (Lane) LaPred [57] (Road Topology) Trajectron++ [6] (Road Geometry)	Vehicle Type	CoverNet [44] (Behavior Priors) Ghoul *et al.* [52] (Motion Constraints)	—	Kumar *et al.* [63]
Apolloscape	—	TPNet [55] (Road Geometry)	Vehicle Type	Trafficpredict [4] (Driving Features) Chandra *et al.* [93] (Behavior Priors)	—	—
INTERACTION	HD-map	TNT [54] (Road Topology) Recog [67] (Road Geometry)	—	Bahari *et al.* [89] (Motion Constraints)	—	—
BLVD	—	—	—	—	Interaction Definition	Li *et al.* [65]
Lyft Level 5	HD-map	FIERY [27] (Road Geometry)	—	Chandra *et al.* [93] (Behavior Priors)	—	—
WOMD	HD-map	MultiPath++ [73] (Lane) Scene Transformer [35] (Road Geometry)	—	MultiPath++ [73] (Motion Constraints)	—	—
Unpublic Datasets	—	DRONGON [72] (Road Geometry)	—	Xie *et al.* [46] (Behavior Priors/Motion Constraints)	—	Lee *et al.* [15]

—-Provided by Datasets —-User Extracted from Datasets

are often collected by a sensor-equipped vehicle (usually called as Ego vehicle) in various formats, such as Lidar points, images, IMU data sequences, and so on. For VTP usage, the trajectories of the ego vehicles in these datasets are obtained from the original IMU and GNSS data, while the trajectories of the surrounding vehicles are often extracted through perception techniques such as object detection and tracking. Since these datasets usually have the characteristics of extensive range, long time, and rich data volume compared to the earlier top-down view VTP datasets, their appearance promotes the development of DL-based VTP research. The typical on-road view datasets for VTP are Argoverse [98], nuScenes [99], Apolloscape [4], BLVD [100], Lyft Level 5 [101] and WOMD [102], as introduced in Tab.II.

Fig.9(a) shows the timeline of vehicle trajectory prediction datasets from different views in the form of a timeline. It can be seen that the early vehicle trajectory prediction datasets are usually obtained from a top-down perspective, while the recent vehicle trajectory prediction datasets are mainly collected by the ego vehicle from an on-road view.

B. Knowledge Contained in VTP Datasets and the Usage

The developments of VTP datasets are not only in the viewpoint of data acquisition but also in the scale of datasets and the rich driving knowledge they contain. Especially with the development of automatic driving systems, high-definition maps (HD maps) have been widely developed and used as indispensable and essential driving knowledge for current automatic driving systems. According to the availability of high-definition map in the dataset, Fig.9 is drawn to present the development trend of VTP dataset concerning HD-map support and data collection viewpoint. To further analyze the knowledge usage of knowledge-based methods on VTP datasets, we present detailed statistics on the wildly-used VTP datasets according to their information and knowledge used in representative knowledge-based methods in Tab.III.

Combining the Tab.II, Tab.III and Fig.9, it can be seen that the current datasets have:

a) **Well** support for **map-related knowledge**. Specifically, there are a large number of VTP methods using road environment knowledge, and most of them use the knowledge or information provided by the dataset.
b) **Insufficient** support for **vehicle-related knowledge**. Specifically, only NGSIM, SDD, Apolloscape, and nuScenes datasets provide vehicle types as vehicle-related knowledge; vehicle-related knowledge used by most methods is extracted from the dataset, rather than directly provided by the datasets.
c) **Little** support for **interaction-related knowledge**. Specifically, within the scope of this review, only one public dataset (BLVD) explicitly provides interactive event labels as interaction-related knowledge; Compared with the other two types of knowledge, fewer studies utilize interaction-related knowledge explicitly.

To better examine the relationship between knowledge-based VTP and datasets, we counted the number of knowledge-based and knowledge-free VTP methods on different datasets by year. To present the effect of recently proposed datasets (such as Argoverse and nuScenes), we heuristically set the year after their existence (2020) and divide the timeline

TABLE IV: Metrics of VTP

Focus	Metirc	Overview	Calculation
Trajectory Accuracy	RMSE	Point-wise, Root Mean Square Error	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{\tau_e}\sum_{t=T+1}^{T+\tau_e}\|\|\hat{Y}_t - Y_t^{GT}\|\|_2^2)}$
	ADE	Point-wise, Average Euclidean distance	$ADE = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{\tau_e}\sum_{t=T+1}^{T+\tau_e}\|\|\hat{Y}_t - Y_t^{GT}\|\|_2^2)$
	FDE	End Point, Euclidean distance	$FDE = \frac{1}{n}\sum_{i=1}^{n}(\|\|\hat{Y}_{T+\tau_e} - Y_{T+\tau_e}^{GT}\|\|_2^2)$
Multimodal Output	Min K	K Prediction, Minimun Error	$MIN_K(ADE, FDE) = \min_k(ADE_k, FDE_k), k \in [1, K]$
	NLL	Distribution Similarity	$NLL = \frac{log\sigma^2(\hat{Y})}{2} + \frac{(Y-\mu(\hat{Y}))^2}{2\sigma^2(\hat{Y})}$
Prediction Validity	MR	Ratio of All Prediction Fail	
	DAC	Drivable Area Coverage	

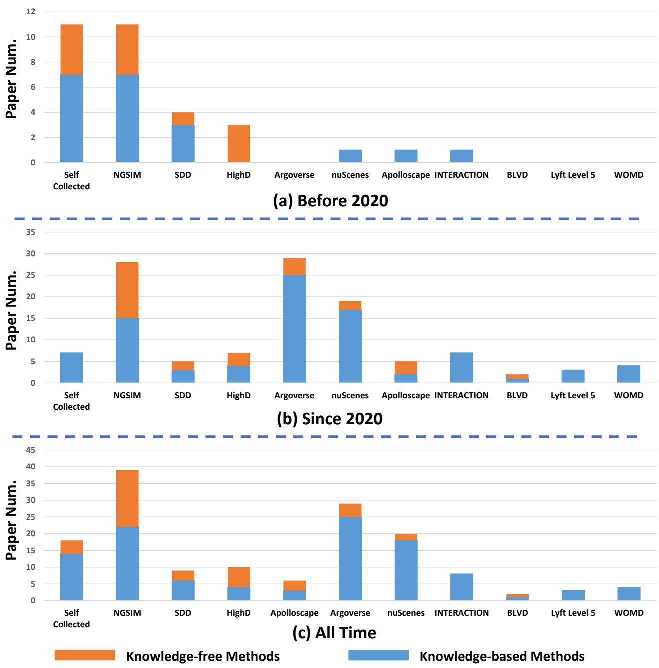


Fig. 10: Knowledge-based VTP in Different Time Periods

into three different periods in Fig.10. It can be concluded from the figure that: **Before 2020**, public VTP datasets cannot provide sufficient support for knowledge-based VTP research. Specifically, the proportion of knowledge-based VTP methods is basically close to knowledge-free methods; some research tends to build small-scale, non-public datasets according to customized needs for knowledge-based research. **Since 2020**, public datasets such as Argoverse, nuScenes, and INTERACTION have facilitated the development of knowledge-based VTP research. Specifically, the proportion of knowledge-based methods has increased significantly, and the works focus on these recent datasets; the proportion of research on VTP using self-collected and non-public datasets has decreased significantly. In **All Time**, the public datasets in recent years are the better choice for Knowledge-based VTP research, while NGSIM is a better choice for general VTP research. Specifically, most of the methods on recent datasets, such as Argoverse and nuScenes are knowledge-based; NGSIM is the most used VTP dataset of all time, with or without driving knowledge.

C. *Evaluation of VTP*

In this part, we will first briefly introduce the current VTP evaluation method, then present an overview of wildly-used VTP metrics.

1) General Evaluation of VTP: Most of the existing works regard the VTP task as an open-loop prediction task in which the evaluation datasets are split into samples of fixed length. For example, works using NGSIM dataset always follow the setting of 8s trajectory segments with 3s as history and 5s to be predicted. On Argoverse, the official datasets are organized into 5s samples with 2s observation and 3s prediction. For nuScenes, the history/prediction horizon is set to be 2s/6s.

The data-driven open-loop evaluation method of VTP has the advantage of a uniform benchmark for comparison. However, the VTP method based on deep learning has not been fully verified in closed-loop and onboard experiments yet. In particular, for the intention prediction task, it is hard to obtain the ground truth of the predicted intention during online inference, making it hard for onboard deployment. Other issues, such as computational efficiency and continuity of prediction results in real-time experiments, have not been systematically explored either.

2) Metrics: In this part, we will list the wildly-used VTP metrics. According to the focus of metrics, the VTP metrics can be divided into Trajectory Accuracy metrics, Multimodal Output metrics, and Prediction Validity metrics, as listed in Tab.IV. **Trajectory Accuracy:** Metrics that focus on the deviation between the predicted trajectory and the ground truth. The mostly used metrics are RMSE (Root Mean Square Error), ADE (Average Displacement Error), and FDE (Final Displacement Error). These metrics can be computed directly from the determined trajectory output or by selecting the most probable prediction from the probabilistic output. The latter case is categorized in [103] as Most-Likely (ML) based metrics. **Multimodal Output:** Metrics that judge the multimodality ability of the model. a) Min K: Calculate the minimum error of K predictions from the model, which is normally combined with ADE or FDE. b) NLL (Negative Log Likelihood): Estimate the similarity between the predicted trajectory distribution and the ground truth. For the convenience of computing, the distribution of predicted results is normally estimated as Gaussian model. **Prediction Validity:** Estimate the validity of model predictions. a) MR (Missing Rate): Compute the percentage of samples that all (K) of the prediction given by the model have higher error than a given threshold. The higher MR indicates that the model cannot handle such scenarios. b) DAC (Drivable Area Compliance): Compute the ratio of predictions that keep in the drivable area.

TABLE V: RMSE of Representative Works on NGSIM

Model	**Year**	**Prediction Horizon**				
		1s	**2s**	**3s**	**4s**	**5s**
CS-LSTM [59]	2018	0.61	1.27	2.09	3.1	4.37
M-LSTM [38]	2018	0.58	1.26	2.12	3.24	4.66
GRIP [104]	2019	0.37	0.86	1.45	2.21	**3.16**
MHA-LSTM [105]	2020	0.41	1.01	1.74	2.67	3.83
GISNet [106]	2020	**0.33**	**0.83**	**1.42**	**2.14**	3.23
SCALE-Net [5]	2020	0.46	1.16	1.97	2.91	-

The lower DAC indicates that the model is more likely to output unsafe results under the input scenario.

From the evaluation criteria summarized above, it can be seen that the current metrics of trajectory accuracy are quite sufficient, while the metrics concerning multimodality and validity are only one aspect to quantify these capabilities. Besides, metrics concerning model reasoning speed, trajectory prediction efficiency, and prediction continuity have not been widely discussed in current VTP research. How to design more comprehensive and effective VTP metrics covering multiple driving attributes is still a topic worthy of discussion and research.

3) Performances of Representative Methods: As shown in Fig.10, NGSIM has been the most popular dataset in VTP study. On the other hand, RMSE/ADE/FDE has never been an absence in evaluating VTP accuracy in literature works. Taking NGSIM and RMSE for example, the accuracy of some representative methods is shown in Tab.V. Obviously, later studies always had relatively better prediction results with smaller errors. However, the average RMSE only provides a rough evaluation of the model performance. The following questions are rarely answered in literature: Under what circumstances has the accuracy of VTP improved? How does VTP perform in rare and critical situations? Does the improvement in VTP accuracy indicate that the algorithm can better cope with interaction and multimodal scenarios? How much has the use of knowledge contributed to the improvement of VTP accuracy? In-depth analysis to understand the performance and challenges of VTP remains the work of the future. We further discuss this issue in Sec.V.

V. Challenges and Open Questions

In complex traffic scenes, vehicle trajectory prediction needs to consider the interaction with road and other traffic participants and the multimodal nature of driving behaviors, which are the main challenges of the VTP task. How to effectively model and address these challenges, namely **Interaction-awareness** and **Multimodality**, is the focus of current research. *How are the main VTP challenges solved and what are the open questions?*

This section reviews current approaches to these challenges, including knowledge-free and knowledge-based approaches, as well as unimodal and multimodal approaches, followed by open questions and other concerns that may lead to future studies.

A. Interaction-awareness

In complex and dense traffic scenes, the ego vehicle completes the driving task while interacting with the road and surrounding vehicles. As a result, the driving trajectory of the ego vehicle is affected by the interactive behavior between the road and the surrounding vehicles, and different interaction patterns may lead to entirely different trajectories, which poses a significant challenge for accurately predicting the vehicle's trajectory. The vehicle trajectory prediction problem considering dynamic scene interaction (**Interaction-aware VTP**) is one of the key issues in current research. According to the reliance on driving knowledge during interaction modeling, the current solution on interaction-aware methods can be divided into Knowledge-free and Knowledge-based modeling, as illustrated in Fig.11. Note that the keywords Knowledge-free Modeling in this section is only used to describe the interaction modeling process. The methods in Knowledge-free Modeling could **still** incorporate driving knowledge for encoding, intention design, or decoding.

1) Knowledge-free Modeling: Utilizes deep learning's feature encoding ability for complex data and modeling ability for nonlinear processes to learn interaction representation in VTP framework. The techniques used in the knowledge-free approach can be further divided into four sub-groups: **Social Tensor:** Model the spatial relation of different traffic participants in the scene through tensors inspired by [107]. The shape of social tensor are normally 2D [59], [108], [109], [105], and can be extended to 3D to include temporal information [110], [111]. **Attention:** Use the Attention mechanism to model the influence of surrounding vehicles or other road users. The attention mechanism can be applied on social tensor [112], [113], [114], feature vector [115], [116], [117], or the grid map [118]. **Graph Neural Networks (GNN):** Use GNN to infer interaction on graph structure. The most important variants of GNN used in VTP are Graph Convolutional Network (GCN) [119], [120], [121] and Graph Attention (GAT) [122], [123], [124]. The construction of the scene graph are normally based on the vehicles and road users [125], [126], [127], [128], [129], and can be extended to include waypoints [130], [71], temporal information [131], or even spectrual information [93], [132]. **Hybrid:** Integrate the above techniques from feature level to model interaction [104], [106], [133].

The Knowledge-free modeling methods usually regard the feature update and aggregation process on different types of neural networks as the process of interaction modeling. This group of methods has the advantage of easily learning from large-scale data. However, as the learned result is a high-dimensional feature vector, the interaction modeling results in knowledge-free modeling methods lack intuitive interpretability and is hard to be independently verified.

2) Knowledge-based Modeling: Effective modeling of interaction is inseparable from the support of driving knowledge. According to the specific usage of knowledge, Knowledge-based modeling methods for interaction can be further divided into two sub-groups: **Implicit Modeling**: Utilize knowledge to design the network structure, reasoning process [134], [51], or constraints of neural network [135], to improve the performance of the model in the interactive scene. **Explicit Modeling**: Utilize the knowledge of interaction mode to explicitly divide different interactive event, then predict trajectory under each event [65], [39], [15]; or explicitly consider the possible

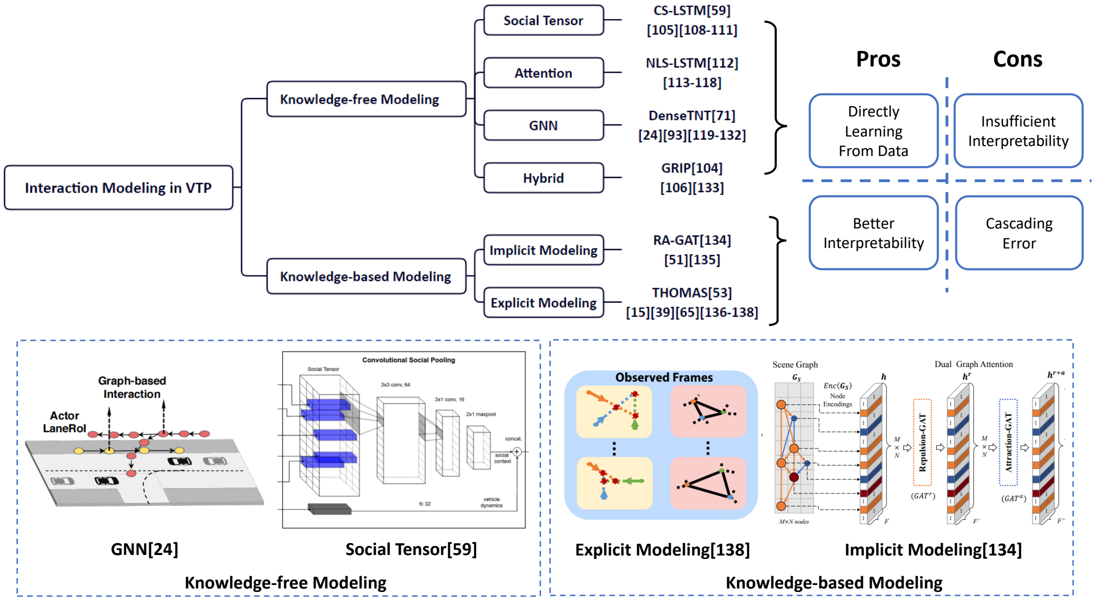


Fig. 11: Interaction Modeling in VTP

conflicts of vehicle trajectories in the scene, then optimize trajectories from a system perspective for collision-free results [53], [136], [137], [138].

The way of introducing knowledge for interaction modeling generally has intermediate results with better interpretability, which can be used to judge whether the model has really ”learned” the given knowledge, and judge whether the current modeling of interaction is effective. However, the interaction itself lacks a unified definition, and introducing knowledge explicitly or implicitly also has the risk of increasing the cascading error of the model, i.e., the wrong prediction of interaction leads to worse trajectory results.

3) Open Questions and Future Directions: From the above analysis, it can be seen that most of the current efforts towards Interaction-aware VTP are in a knowledge-free, data-driven manner, and regard the feature update and aggregation processes of neural networks as the process of interaction modeling, while there are relatively few studies on introducing knowledge in an explicit or implicit way to model interaction.

Combining the conclusions from IV-C, we can conclude that the current interaction-aware VTP research has the following open questions and future research directions: **i)** From the perspective of **driving knowledge**, due to the lack of a unified definition of the interaction behavior itself, most of the current research concerning Interaction-aware VTP is in a data-driven manner, and the research on introducing knowledge to model interaction is relatively scarce. A recent survey on interaction modeling [139] suggests that cognitive models are one of the possible directions which could bring theoretical evidence for interaction modeling. However, further discussions about explicit interaction modeling based on specific knowledge are still needed. **ii)** From the perspective of **dataset support**, in the current public datasets, only a few provide heuristically defined labels related to the interactive event, and cannot support knowledge-based interaction modeling. Human drivers are making decisions in a unified way in which the decision processes could be continuous, and the interaction could be non-discretized. How to effectively and reasonably discretize and define interaction labels to build VTP datasets, like BLVD [100], is still a future direction for interaction-aware VTP research. **iii)** From the perspective of **evaluation methods**, there is currently no unified evaluation method and metrics for interaction modeling, which cannot reflect the effectiveness of Interaction-aware VTP. For now, the interaction modeling results are mostly presented through the accuracy of trajectory prediction or the heuristic attention map [105], [134]. Therefore, more discussions on designing evaluation methods and metrics to effectively examine the quality of interaction modeling could be helpful for related research.

B. Multimodality

Driving behavior is multimodal in nature. Taking the lane-changing behavior of vehicles as an example, when it is necessary to change lanes, the driver may choose to first overtake or pass the vehicle in the target lane, and then complete the lane change maneuver. Such different decisions can lead to entirely different trajectories. Facing the same scene, different drivers may have different behavioral decisions, and the same driver may have different choices under different driving states. If the trajectory prediction model only fits the final trajectory completed by the ego vehicle in the dataset while ignoring the multimodal characteristics of driving behavior, it is hard to guarantee the accuracy of prediction in real applications, and brings disastrous consequences for downstream control tasks. How to consider and model multimodality is also one of the critical challenges in the research of VTP. According to the trajectory output mode in the existing VTP research,

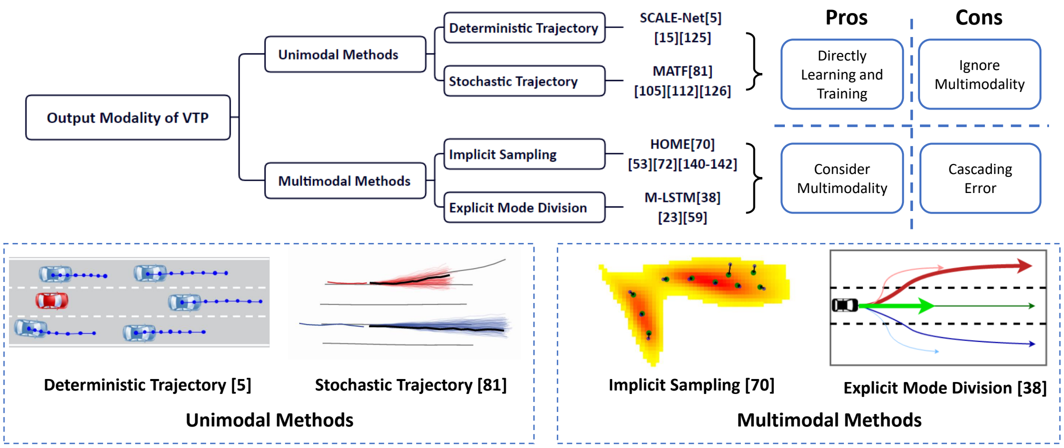


Fig. 12: Output Modality of VTP

these methods can be divided into Unimodal and Multimodal methods, as shown in Fig12.

1) Unimodal Methods: Directly outputs the unimodal trajectory as the prediction result. The trajectory forms output by unimodal methods can be further divided into two types: **Deterministic trajectory**: Outputs the predicted trajectory in the form of deterministic trajectory points [5], [125], [15], which can be directly used for back-propagation and training through metrics such as RMSE. **Stochastic trajectory**: Outputs the parameters of the unimodal trajectory distribution [112], [105], or samples multiple trajectories from unimodal distribution [81], [126].

The Unimodal methods are concise and straightforward, and their outputs error can be directly calculated through various trajectory metrics, which is convenient for model learning and training. However, they actually ignore the multimodal nature of trajectories, the model might fail to output accurate results facing complex scenes with multiple possible trajectories.

2) Multimodal Methods: Output trajectory predictions in the form of multimodal probability distribution. According to the process of output obtaining, multimodal methods can be further divided into two sub-groups: **Implicit Sampling**: Sample trajectories from the trajectory probability density heatmap [70], [53], [72], [140] or from the multimodal distribution such as Gaussian Mixture Models (GMM) [141], [142]. In such a form of multimodal distribution, the parameters (or heatmaps) of the model are obtained by neural network and do not correspond to specific driving behaviors. **Explicit Mode Division**: Explicitly divide the driving modes according to the driving intentions, and output probability combination under different driving intentions (modes) [59], [38], [23]. Such division requires specific driving knowledge.

The Multimodal methods consider the multimodal nature of trajectories and driving behaviors, and can output trajectories with evident modal distinctions. However, they rely on more complex framework design, and might suffer from extra problems such as mode collapse, cascading errors, etc.

3) Open Questions and Future Directions: According to the above analysis and combining the conclusions of Section II-C and IV-C, we believe that the current research on multimodal VTP has the following open questions: **i)** From the perspective of **driving knowledge**, the modal division and definition of multimodal trajectory lack a unified definition. The research of knowledge-based ways for multimodal VTP is not systematic. Under this circumstance, more discussions on the modal design combining different knowledge will lead to various research. For instance, [143] defines the natural language description of vehicle maneuvers as the driving intention the model needs to predict, which combines the VTP task with NLP techniques. We believe such works could prompt the development of multimodal VTP research. **ii)** From the perspective of **dataset support**, as one sequence of real-world data only has one deterministic trajectory ground truth, it cannot support effective multimodal VTP research in this formulation. One possible solution is to organize datasets that provide multiple ground truth trajectories for a single observation, such as the ForkingPath [144] for human trajectory prediction. Datasets in such a formulation will help to model vehicles' multimodal behavior. **iii)** From the perspective of **evaluation methods and metrics**, there is also no unified and effective evaluation method and metrics for multimodal trajectory output, which means the model's ability of multimodality handling cannot be accurately reflected. With the above multimodal-oriented VTP dataset, multimodal outputs can be evaluated at the distribution level [103], i.e., Distribution-aware metrics can be exploited. This could be one of the future directions for related research.

C. Other Concerns

1) Trajectory Prediction for Autonomous Vehicles: As discussed in Sec.IV-C-1), most of the current VTP methods are based on an open-loop data-driven manner, which treats TP as an independent task. However, in the hierarchical autonomous driving system, the trajectory prediction is located between perception and planning. When performing trajectory prediction under such a system, many additional situations must be carefully considered. For example, the tracking sequences of surrounding road users are always complete and well

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3266446

annotated in the public dataset, but the real-time perception results could be discontinuous and noisy. One possible solution is to combine part of the perception task (such as object detection and tracking) with the trajectory prediction using the raw sensor input, as discussed in [145], [127]. For downstream planning and control tasks, it is vital to consider the stability and continuity of the prediction results. Similar concern has been addressed in [146], [147], and we believe this topic deserves further discussion.

2) The Higher Accuracy, the Better?: As described in Sec.IV-C-3), current VTP methods mainly focus on chasing better accuracy in public datasets. However, we address a different concern. Under normal circumstances, the majority of the naturalistic driving data are common cases, while the minority are the critical cases, as discussed in [148], [149]. When the RMSE given by VTP models drops, it is hard to distinguish whether the performance on common cases is slightly better or the hard cases are effectively solved by the model. We believe that the latter form of improvement is of higher significance for trajectory prediction, however, it is rarely discussed in the current study.

3) Dealing with Uncertainty: Deep neural networks have unavoidable uncertainties. The researchers have divided it into aleatoric uncertainty (also known as data uncertainty) and epistemic uncertainty (also known as model uncertainty) [150]. When the model fails to output precise results, the uncertainty analysis could help to find whether it is a brand new case different from the training dataset, or it is because the model did not learn well enough. Some studies have addressed this issue briefly [149], [151]. We believe that further mining of uncertainties can lead to more research insights.

VI. Conclusion

In this paper, we present a systematic literature review on the incorporation of driving knowledge into DL-based VTP research from the perspective of problem formulation, methodology, dataset and evaluation, and open challenge. We firstly present a taxonomy on VTP formulation, in which the studies are divided into the general VTP, incorporating driving knowledge in VTP, and introducing a secondary task to VTP, i.e., intention prediction, based on driving knowledge. We then review the general workflows of the three VTP formulations, followed by a detailed analysis to reveal how driving knowledge is incorporated in each module. We further review the public VTP datasets and evaluation metrics to dig out how knowledge-based VTP studies are supported from the dataset and evaluations' perspective. We finally discuss the main VTP challenges in dynamic traffic scenes, namely interaction-awareness and multimodality issues, to understand how these challenges are addressed by the literature works, and what open questions have remained. Our main findings are:

1) Although incorporating driving knowledge into DL-based methods has attracted significant concerns in VTP research in recent years, methods have been mainly developed to incorporate map-related knowledge in driving state encoding, while vehicle-related knowledge in intention prediction as shown in Tab.I. The literature shows that the existing work is uneven, lacks systematicity, and is far from exhaustive, leaving much room for future research to achieve a systematic study of knowledge-based VTP.
2) The development of public VTP datasets has received more attention, leaving less demand for researchers to use their own developed datasets as revealed in Fig.10 and Tab.III. The public datasets developed in recent years are mainly from an on-road view and place great emphasis on the inclusion of map knowledge, i.e., HD maps. However, vehicle-related knowledge and interaction-related knowledge are rarely included, which poses a significant limitation to the systematic and diverse study of knowledge-based VTP.
3) In complex traffic scenes, vehicle trajectory prediction needs to consider the interaction with the road and other traffic participants, interaction-awareness has been one of the main challenges of the current VTP study. As shown in Fig.11 and discussed in Section V-A, most research efforts have focused on developing knowledge-free methods using deep learning's ability to represent interaction through feature encoding and aggregation, with insufficient research on explicit or implicit incorporation of interaction-related knowledge. Furthermore, current metrics focus on evaluating trajectory accuracy, and cannot directly assess how well the model aware interaction.
4) Driving behavior is multimodal in nature, leading to largely different trajectories in the same situation and even by the same driver. Addressing multimodal driving behaviors in complex traffic scenes has been another main challenge of the current VTP study. As shown in Fig.12 and discussed in Section V-B, the study of multimodal VTP incorporating driving knowledge is insufficient. Furthermore, the datasets provide ground truth trajectories that are unimodal, which means that the current datasets by nature do not support multimodal VTP studies well.

A vehicle trajectory is an integrated result of an ego vehicle considering interaction with the road and surrounding vehicles, the constraints of the vehicle's kinematic and dynamic models, and the preference of driving pattern. As a result, different road structures, different traffic environments, and different vehicle models will lead to different vehicle trajectories. Obviously, we can simply learn a VTP model by regressing on a given VTP dataset without considering the above conditions, however, ignoring the inherent correlation of data often leads to limited and biased results. In this case, introducing driving knowledge into the VTP methods can significantly improve the interpretability and generalization ability of the methods. If the trajectory prediction model only fits the final trajectory completed by the ego vehicle in the dataset while ignoring the multimodal characteristics of driving behavior, it is hard to guarantee the accuracy of prediction in real applications and may bring disastrous consequences for downstream control tasks. This paper provides a thorough review of the literature on DL-based VTP by focusing on incorporating driving knowledge. The findings of this review suggest future works to address the challenges of VTP in complex dynamic scenes.

References

[1] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: a survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 2, pp. 315–329, 2020.

[2] 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 *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 399–404.

[3] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH journal*, vol. 1, no. 1, pp. 1–14, 2014.

[4] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, and D. Manocha, "Trafficpredict: Trajectory prediction for heterogeneous traffic-agents," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 6120–6127.

[5] H. Jeon, J. Choi, and D. Kum, "Scale-net: Scalable vehicle trajectory prediction network under random number of interacting vehicles via edge-enhanced graph convolutional neural network," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2095–2102.

[6] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *European Conference on Computer Vision*. Springer, 2020, pp. 683–700.

[7] 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 Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 33–47, 2020.

[8] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and decision-making for autonomous vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 187–210, 2018.

[9] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, "A survey on trajectory-prediction methods for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, 2022.

[10] É. Zablocki, H. Ben-Younes, P. Pérez, and M. Cord, "Explainability of deep vision-based autonomous driving systems: Review and challenges," *International Journal of Computer Vision*, pp. 1–28, 2022.

[11] S. Casas, C. Gulino, S. Suo, and R. Urtasun, "The importance of prior knowledge in precise multimodal prediction," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2295–2302.

[12] S. Mozaffari, E. Arnold, M. Dianati, and S. Fallah, "Early lane change prediction for automated driving systems using multi-task attention-based convolutional neural networks," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 758–770, 2022.

[13] S. Wang, P. Zhao, B. Yu, W. Huang, and H. Liang, "Vehicle trajectory prediction by knowledge-driven lstm network in urban environments," *Journal of Advanced Transportation*, vol. 2020, 2020.

[14] X. Xu, W. Liu, and L. Yu, "Trajectory prediction for heterogeneous traffic-agents using knowledge correction data-driven model," *Information Sciences*, vol. 608, pp. 375–391, 2022.

[15] D. Lee, Y. Gu, J. Hoang, and M. Marchetti-Bowick, "Joint interaction and trajectory prediction for autonomous driving using graph neural networks," *arXiv preprint arXiv:1912.07882*, 2019.

[16] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[17] R. Korbmacher and A. Tordeux, "Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[18] A. Zyner, S. Worrall, and E. Nebot, "Naturalistic driver intention and path prediction using recurrent neural networks," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 4, pp. 1584–1594, 2019.

[19] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified lstm models for trajectory prediction," *IEEE Access*, vol. 7, pp. 38 287–38 296, 2019.

[20] Y. Liu, J. Zhang, L. Fang, Q. Jiang, and B. Zhou, "Multimodal motion prediction with stacked transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7577–7586.

[21] C. Choi, J. H. Choi, J. Li, and S. Malla, "Shared cross-modal trajectory prediction for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 244–253.

[22] H. He, H. Dai, and N. Wang, "Ust: Unifying spatio-temporal context for trajectory prediction in autonomous driving," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5962–5969.

[23] L. Zhang, P.-H. Su, J. Hoang, G. C. Haynes, and M. Marchetti-Bowick, "Map-adaptive goal-based trajectory prediction," in *Conference on Robot Learning*. PMLR, 2021, pp. 1371–1383.

[24] W. Zeng, M. Liang, R. Liao, and R. Urtasun, "Lanercnn: Distributed representations for graph-centric motion forecasting," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 532–539.

[25] 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 *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1672–1678.

[26] R. Chandra, U. Bhattacharya, C. Roncal, A. Bera, and D. Manocha, "Robusttp: End-to-end trajectory prediction for heterogeneous road-agents in dense traffic with noisy sensor inputs," in *ACM Computer Science in Cars Symposium*, 2019, pp. 1–9.

[27] A. Hu, Z. Murez, N. Mohan, S. Dudas, J. Hawke, V. Badrinarayanan, R. Cipolla, and A. Kendall, "Fiery: Future instance prediction in bird's-eye view from surround monocular cameras," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 273–15 282.

[28] L. Hou, L. Xin, S. E. Li, B. Cheng, and W. Wang, "Interactive trajectory prediction of surrounding road users for autonomous driving using structural-lstm network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4615–4625, 2019.

[29] M. Bhat, J. Francis, and J. Oh, "Trajformer: Trajectory prediction with local self-attentive contexts for autonomous driving," *arXiv preprint arXiv:2011.14910*, 2020.

[30] J. Li, H. Ma, and M. Tomizuka, "Conditional generative neural system for probabilistic trajectory prediction," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 6150–6156.

[31] J. Li, F. Yang, M. Tomizuka, and C. Choi, "Evolvegraph: Multi-agent trajectory prediction with dynamic relational reasoning," *Advances in neural information processing systems*, vol. 33, pp. 19 783–19 794, 2020.

[32] X. Mo, Z. Huang, Y. Xing, and C. Lv, "Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[33] S. Wen, H. Wang, and D. Metaxas, "Social ode: Multi-agent trajectory forecasting with neural ordinary differential equations," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII*. Springer, 2022, pp. 217–233.

[34] N. Rhinehart, R. McAllister, K. Kitani, and S. Levine, "Precog: Prediction conditioned on goals in visual multi-agent settings," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2821–2830.

[35] J. Ngiam, V. Vasudevan, B. Caine, Z. Zhang, H.-T. L. Chiang, J. Ling, R. Roelofs, A. Bewley, C. Liu, A. Venugopal *et al.*, "Scene transformer: A unified architecture for predicting future trajectories of multiple agents," in *International Conference on Learning Representations*, 2022.

[36] W. Ding and S. Shen, "Online vehicle trajectory prediction using policy anticipation network and optimization-based context reasoning," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 9610–9616.

[37] J. Li, H. Ma, Z. Zhang, J. Li, and M. Tomizuka, "Spatio-temporal graph dual-attention network for multi-agent prediction and tracking," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[38] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based lstms," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1179–1184.

[39] Y. Hu, W. Zhan, L. Sun, and M. Tomizuka, "Multi-modal probabilistic prediction of interactive behavior via an interpretable model," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 557–563.

[40] C. Luo, L. Sun, D. Dabiri, and A. Yuille, "Probabilistic multi-modal trajectory prediction with lane attention for autonomous vehicles," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2370–2376.

[41] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun, "Learning lane graph representations for motion forecasting," in *European Conference on Computer Vision*. Springer, 2020, pp. 541–556.

[42] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8483–8492.

[43] X. Feng, Z. Cen, J. Hu, and Y. Zhang, "Vehicle trajectory prediction using intention-based conditional variational autoencoder," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3514–3519.

[44] T. Phan-Minh, E. C. Grigore, F. A. Boulton, O. Beijbom, and E. M. Wolff, "Covernet: Multimodal behavior prediction using trajectory sets," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 074–14 083.

[45] N. Deo and M. M. Trivedi, "Trajectory forecasts in unknown environments conditioned on grid-based plans," *arXiv preprint arXiv:2001.00735*, 2020.

[46] 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 Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5999–6008, 2017.

[47] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. 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.

[48] W. Yao, Q. Zeng, Y. Lin, D. Xu, H. Zhao, F. Guillemard, S. Geronimi, and F. Aioun, "On-road vehicle trajectory collection and scene-based lane change analysis: Part ii," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 206–220, 2016.

[49] J. Wang, T. Ye, Z. Gu, and J. Chen, "Ltp: Lane-based trajectory prediction for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 134–17 142.

[50] N. Deo, E. Wolff, and O. Beijbom, "Multimodal trajectory prediction conditioned on lane-graph traversals," in *Conference on Robot Learning*. PMLR, 2022, pp. 203–212.

[51] H. Song, D. Luan, W. Ding, M. Y. Wang, and Q. Chen, "Learning to predict vehicle trajectories with model-based planning," in *Conference on Robot Learning*. PMLR, 2022, pp. 1035–1045.

[52] A. Ghoul, K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "A lightweight goal-based model for trajectory prediction," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 4209–4214.

[53] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, "Thomas: Trajectory heatmap output with learned multi-agent sampling," *arXiv preprint arXiv:2110.06607*, 2021.

[54] H. Zhao, J. Gao, T. Lan, C. Sun, B. Sapp, B. Varadarajan, Y. Shen, Y. Shen, Y. Chai, C. Schmid *et al.*, "Tnt: Target-driven trajectory prediction," in *Conference on Robot Learning*. PMLR, 2021, pp. 895–904.

[55] L. Fang, Q. Jiang, J. Shi, and B. Zhou, "Tpnet: Trajectory proposal network for motion prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6797–6806.

[56] D. Sierra-Gonzalez, A. Paigwar, O. Erkent, and C. Laugier, "Multilane: Lane intention prediction and sensible lane-oriented trajectory forecasting on centerline graphs," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3657–3664.

[57] B. Kim, S. H. Park, S. Lee, E. Khoshimjonov, D. Kum, J. Kim, J. S. Kim, and J. W. Choi, "Lapred: Lane-aware prediction of multi-modal future trajectories of dynamic agents," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 636–14 645.

[58] A. Benterki, M. Boukhnifer, V. Judalet, and C. Maaoui, "Artificial intelligence for vehicle behavior anticipation: Hybrid approach based on maneuver classification and trajectory prediction," *IEEE Access*, vol. 8, pp. 56 992–57 002, 2020.

[59] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1468–1476.

[60] S. Casas, W. Luo, and R. Urtasun, "Intentnet: Learning to predict intention from raw sensor data," in *Conference on Robot Learning*. PMLR, 2018, pp. 947–956.

[61] S. Dai, Z. Li, L. Li, N. Zheng, and S. Wang, "A flexible and explainable vehicle motion prediction and inference framework combining semi-supervised aog and st-lstm," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[62] H. Berkemeyer, R. Franceschini, T. Tran, L. Che, and G. Pipa, "Feasible and adaptive multimodal trajectory prediction with semantic maneuver fusion," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 8530–8536.

[63] S. Kumar, Y. Gu, J. Hoang, G. C. Haynes, and M. Marchetti-Bowick, "Interaction-based trajectory prediction over a hybrid traffic graph," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 5530–5535.

[64] Y. Ban, X. Li, G. Rosman, I. Gilitschenski, O. Meireles, S. Karaman, and D. Rus, "A deep concept graph network for interaction-aware trajectory prediction," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 8992–8998.

[65] Z. Li, C. Lu, Y. Yi, and J. Gong, "A hierarchical framework for interactive behaviour prediction of heterogeneous traffic participants based on graph neural network," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[66] C. Ju, Z. Wang, C. Long, X. Zhang, and D. E. Chang, "Interaction-aware kalman neural networks for trajectory prediction," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1793–1800.

[67] X. Mo, Y. Xing, and C. Lv, "Recog: A deep learning framework with heterogeneous graph for interaction-aware trajectory prediction," *arXiv preprint arXiv:2012.05032*, 2020.

[68] Z. Huang, X. Mo, and C. Lv, "Multi-modal motion prediction with transformer-based neural network for autonomous driving," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2605–2611.

[69] J. Gao, C. Sun, H. Zhao, Y. Shen, D. Anguelov, C. Li, and C. Schmid, "Vectornet: Encoding hd maps and agent dynamics from vectorized representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 525–11 533.

[70] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, "Home: Heatmap output for future motion estimation," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 500–507.

[71] J. Gu, C. Sun, and H. Zhao, "Densetnt: End-to-end trajectory prediction from dense goal sets," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 303–15 312.

[72] C. Choi, S. Malla, A. Patil, and J. H. Choi, "Drogon: A trajectory prediction model based on intention-conditioned behavior reasoning," in *Conference on Robot Learning*. PMLR, 2021, pp. 49–63.

[73] B. Varadarajan, A. Hefny, A. Srivastava, K. S. Refaat, N. Nayakanti, A. Cornman, K. Chen, B. Douillard, C. P. Lam, D. Anguelov *et al.*, "Multipath++: Efficient information fusion and trajectory aggregation for behavior prediction," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7814–7821.

[74] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction," *arXiv preprint arXiv:1910.05449*, 2019.

[75] Y. Yoon, T. Kim, H. Lee, and J. Park, "Road-aware trajectory prediction for autonomous driving on highways," *Sensors*, vol. 20, no. 17, p. 4703, 2020.

[76] H. Kim, D. Kim, G. Kim, J. Cho, and K. Huh, "Multi-head attention based probabilistic vehicle trajectory prediction," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1720–1725.

[77] A. Kawasaki and A. Seki, "Multimodal trajectory predictions for urban environments using geometric relationships between a vehicle and lanes," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9203–9209.

[78] R. Greer, N. Deo, and M. Trivedi, "Trajectory prediction in autonomous driving with a lane heading auxiliary loss," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4907–4914, 2021.

[79] L. Hou, S. E. Li, B. Yang, Z. Wang, and K. Nakano, "Integrated graphical representation of highway scenarios to improve trajectory prediction of surrounding vehicles," *IEEE Transactions on Intelligent Vehicles*, 2022.

[80] R. Walters, J. Li, and R. Yu, "Trajectory prediction using equivariant continuous convolution," *arXiv preprint arXiv:2010.11344*, 2020.

[81] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 126–12 134.

[82] 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 *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2090–2096.

[83] J. Strohbeck, V. Belagiannis, J. Müller, M. Schreiber, M. Herrmann, D. Wolf, and M. Buchholz, "Multiple trajectory prediction with deep temporal and spatial convolutional neural networks," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 1992–1998.

[84] K. Messaoud, N. Deo, M. M. Trivedi, and F. Nashashibi, "Trajectory prediction for autonomous driving based on multi-head attention with joint agent-map representation," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 165–170.

[85] F. Altché and A. de La Fortelle, "An lstm network for highway trajectory prediction," in *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)*. IEEE, 2017, pp. 353–359.

[86] H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, J. Schneider, D. Bradley, and N. Djuric, "Deep kinematic models for kinematically feasible vehicle trajectory predictions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 563–10 569.

[87] L. Xin, P. Wang, C.-Y. Chan, J. Chen, S. E. Li, and B. Cheng, "Intention-aware long horizon trajectory prediction of surrounding vehicles using dual lstm networks," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 1441–1446.

[88] X. Tang, K. Yang, H. Wang, J. Wu, Y. Qin, W. Yu, and D. Cao, "Prediction-uncertainty-aware decision-making for autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 849–862, 2022.

[89] M. Bahari, I. Nejjar, and A. Alahi, "Injecting knowledge in data-driven vehicle trajectory predictors," *Transportation research part C: emerging technologies*, vol. 128, p. 103010, 2021.

[90] K. Cho, T. Ha, G. Lee, and S. Oh, "Deep predictive autonomous driving using multi-agent joint trajectory prediction and traffic rules," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 2076–2081.

[91] Y. Na, J. Lee, and K. Jo, "Interaction-aware trajectory prediction of surrounding vehicles based on hierarchical framework in highway scenarios," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 1060–1065.

[92] H. Zhang, Y. Wang, J. Liu, C. Li, T. Ma, and C. Yin, "A multi-modal states based vehicle descriptor and dilated convolutional social pooling for vehicle trajectory prediction," *arXiv preprint arXiv:2003.03480*, 2020.

[93] R. Chandra, T. Guan, S. Panuganti, T. Mittal, U. Bhattacharya, A. Bera, and D. Manocha, "Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4882–4890, 2020.

[94] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale, "The next generation simulation program," *Institute of Transportation Engineers. ITE Journal*, vol. 74, no. 8, p. 22, 2004.

[95] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *European conference on computer vision*. Springer, 2016, pp. 549–565.

[96] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2118–2125.

[97] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle *et al.*, "Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," *arXiv preprint arXiv:1910.03088*, 2019.

[98] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8748–8757.

[99] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[100] J. Xue, J. Fang, T. Li, B. Zhang, P. Zhang, Z. Ye, and J. Dou, "Blvd: Building a large-scale 5d semantics benchmark for autonomous driving," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6685–6691.

[101] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, L. Chen, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska, "One thousand and one hours: Self-driving motion prediction dataset," in *Conference on Robot Learning*. PMLR, 2021, pp. 409–418.

[102] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou *et al.*, "Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9710–9719.

[103] R. Huang, H. Xue, M. Pagnucco, F. Salim, and Y. Song, "Multimodal trajectory prediction: A survey," *arXiv preprint arXiv:2302.10463*, 2023.

[104] X. Li, X. Ying, and M. C. Chuah, "Grip: Graph-based interaction-aware trajectory prediction," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3960–3966.

[105] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Attention based vehicle trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 175–185, 2020.

[106] Z. Zhao, H. Fang, Z. Jin, and Q. Qiu, "Gisnet: Graph-based information sharing network for vehicle trajectory prediction," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–7.

[107] 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.

[108] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Relational recurrent neural networks for vehicle trajectory prediction," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 1813–1818.

[109] Y. Wang, S. Zhao, R. Zhang, X. Cheng, and L. Yang, "Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 236–248, 2020.

[110] S. Wang, Y. Huang, M. Kang, B. Chen, and N. Zheng, "3d-mbnet: Intention based multimodal vehicle trajectory prediction with 3d social convolution," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 880–887.

[111] S. Mukherjee, S. Wang, and A. Wallace, "Interacting vehicle trajectory prediction with convolutional recurrent neural networks," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4336–4342.

[112] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Non-local social pooling for vehicle trajectory prediction," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 975–980.

[113] T. Yang, Z. Nan, H. Zhang, S. Chen, and N. Zheng, "Traffic agent trajectory prediction using social convolution and attention mechanism," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 278–283.

[114] L. Lin, W. Li, H. Bi, and L. Qin, "Vehicle trajectory prediction using lstms with spatial–temporal attention mechanisms," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 2, pp. 197–208, 2021.

[115] J. Mercat, T. Gilles, N. El Zoghby, G. Sandou, D. Beauvois, and G. P. Gil, "Multi-head attention for multi-modal joint vehicle motion forecasting," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9638–9644.

[116] N. Kamra, H. Zhu, D. K. Trivedi, M. Zhang, and Y. Liu, "Multi-agent trajectory prediction with fuzzy query attention," *Advances in Neural Information Processing Systems*, vol. 33, pp. 22 530–22 541, 2020.

[117] Z. Huang, X. Mo, and C. Lv, "Recoat: A deep learning-based framework for multi-modal motion prediction in autonomous driving application," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 988–993.

[118] X. Ren, T. Yang, L. E. Li, A. Alahi, and Q. Chen, "Safety-aware motion prediction with unseen vehicles for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 731–15 740.

[119] D. Xu, X. Shang, Y. Liu, H. Peng, and H. Li, "Group vehicle trajectory prediction with global spatio-temporal graph," *IEEE Transactions on Intelligent Vehicles*, 2022.

[120] J. Schmidt, J. Jordan, F. Gritschneder, and K. Dietmayer, "Crat-pred: Vehicle trajectory prediction with crystal graph convolutional neural networks and multi-head self-attention," *arXiv preprint arXiv:2202.04488*, 2022.

[121] E. M. Rella, J.-N. Zaech, A. Liniger, and L. Van Gool, "Decoder fusion rnn: Context and interaction aware decoders for trajectory prediction,"

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3266446

in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 5937–5943.

[122] K. Zhang, L. Zhao, C. Dong, L. Wu, and L. Zheng, “Ai-tp: Attention-based interaction-aware trajectory prediction for autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, 2022.

[123] X. Mo, Z. Huang, and C. Lv, “Stochastic multimodal interaction prediction for urban driving,” in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 1000–1005.

[124] S. Carrasco, D. F. Llorca, and M. Sotelo, “Scout: Socially-consistent and understandable graph attention network for trajectory prediction of vehicles and vrus,” in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 1501–1508.

[125] F. Diehl, T. Brunner, M. T. Le, and A. Knoll, “Graph neural networks for modelling traffic participant interaction,” in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 695–701.

[126] J. Su, P. A. Beling, R. Guo, and K. Han, “Graph convolution networks for probabilistic modeling of driving acceleration,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–8.

[127] S. Casas, C. Gulino, R. Liao, and R. Urtasun, “Spagnn: Spatially-aware graph neural networks for relational behavior forecasting from sensor data,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9491–9497.

[128] X. Mo, Y. Xing, and C. Lv, “Graph and recurrent neural network-based vehicle trajectory prediction for highway driving,” in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 1934–1939.

[129] B. Ivanovic and M. Pavone, “The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2375–2384.

[130] L. Zhang, P. Li, J. Chen, and S. Shen, “Trajectory prediction with graph-based dual-scale context fusion,” *arXiv preprint arXiv:2111.01592*, 2021.

[131] J. Pan, H. Sun, K. Xu, Y. Jiang, X. Xiao, J. Hu, and J. Miao, “Lane-attention: Predicting vehicles’ moving trajectories by learning their attention over lanes,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 7949–7956.

[132] D. Cao, J. Li, H. Ma, and M. Tomizuka, “Spectral temporal graph neural network for trajectory prediction,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 1839–1845.

[133] Z. Sheng, Y. Xu, S. Xue, and D. Li, “Graph-based spatial-temporal convolutional network for vehicle trajectory prediction in autonomous driving,” *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[134] Z. Ding, Z. Yao, and H. Zhao, “Ra-gat: Repulsion and attraction graph attention for trajectory prediction,” in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 734–741.

[135] S. Narayanan, R. Moslemi, F. Pittaluga, B. Liu, and M. Chandraker, “Divide-and-conquer for lane-aware diverse trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 799–15 808.

[136] J. Roh, C. Mavrogiannis, R. Madan, D. Fox, and S. Srinivasa, “Multi-modal trajectory prediction via topological invariance for navigation at uncontrolled intersections,” in *Conference on Robot Learning*. PMLR, 2021, pp. 2216–2227.

[137] N. Deo, A. Rangesh, and M. M. Trivedi, “How would surround vehicles move? a unified framework for maneuver classification and motion prediction,” *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 2, pp. 129–140, 2018.

[138] X. Xie, C. Zhang, Y. Zhu, Y. N. Wu, and S.-C. Zhu, “Congestion-aware multi-agent trajectory prediction for collision avoidance,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 693–13 700.

[139] W. Wang, L. Wang, C. Zhang, C. Liu, L. Sun *et al.*, “Social interactions for autonomous driving: A review and perspectives,” *Foundations and Trends® in Robotics*, vol. 10, no. 3-4, pp. 198–376, 2022.

[140] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, “Gohome: Graph-oriented heatmap output for future motion estimation,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9107–9114.

[141] M. N. Azadani and A. Boukerche, “A novel multimodal vehicle path prediction method based on temporal convolutional networks,” *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[142] X. Huang, S. G. McGill, B. C. Williams, L. Fletcher, and G. Rosman, “Uncertainty-aware driver trajectory prediction at urban intersections,” in *2019 International conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 9718–9724.

[143] Y.-L. Kuo, X. Huang, A. Barbu, S. G. McGill, B. Katz, J. J. Leonard, and G. Rosman, “Trajectory prediction with linguistic representations,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2868–2875.

[144] J. Liang, L. Jiang, K. Murphy, T. Yu, and A. Hauptmann, “The garden of forking paths: Towards multi-future trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 508–10 518.

[145] M. Biparva, D. Fernández-Llorca, R. I. Gonzalo, and J. K. Tsotsos, “Video action recognition for lane-change classification and prediction of surrounding vehicles,” *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 569–578, 2022.

[146] X. Wang, K. Tang, X. Dai, J. Xu, J. Xi, R. Ai, Y. Wang, W. Gu, and C. Sun, “Safety-balanced driving-style aware trajectory planning in intersection scenarios with uncertain environment,” *IEEE Transactions on Intelligent Vehicles*, 2023.

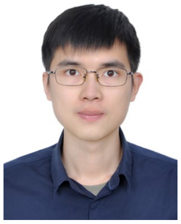
[147] R. McAllister, B. Wulfe, J. Mercat, L. Ellis, S. Levine, and A. Gaidon, “Control-aware prediction objectives for autonomous driving,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 01–08.

[148] Z. Li, Y. Wang, and Z. Zuo, “Interaction-aware prediction for cut-in trajectories with limited observable neighboring vehicles,” *IEEE Transactions on Intelligent Vehicles*, 2023.

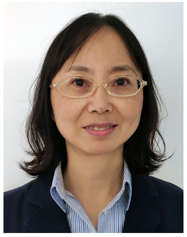
[149] W. Zhou, Z. Cao, Y. Xu, N. Deng, X. Liu, K. Jiang, and D. Yang, “Long-tail prediction uncertainty aware trajectory planning for self-driving vehicles,” in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 1275–1282.

[150] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya *et al.*, “A review of uncertainty quantification in deep learning: Techniques, applications and challenges,” *Information Fusion*, vol. 76, pp. 243–297, 2021.

[151] G. Li, Z. Li, V. Knoop, and J. van Lint, “Uqnet: Quantifying uncertainty in trajectory prediction by a non-parametric and generalizable approach,” *Social Science Research Network (online)*, 2022.



Zhezhang Ding received the B.S. degree in computer science (intelligent science and technology) from Peking University, Beijing, China, in 2018. He is currently pursuing the Ph.D. degree in intelligent vehicles with the Key Laboratory of Machine Perception (Ministry of Education), School of Intelligence Science and Technology, Peking University, Beijing, China. His research interests include intelligent vehicles and machine learning.



Huijing Zhao received B.S. degree in computer science from Peking University in 1991. She obtained M.E. degree in 1996 and Ph.D. degree in 1999 in civil engineering from the University of Tokyo, Japan. From 1999 to 2007, she was a postdoctoral researcher and visiting associate professor at the Center for Space Information Science, University of Tokyo. In 2007, she joined Peking University as a tenure-track professor at the School of Electronics Engineering and Computer Science and became an associate professor with tenure on 2013. She is now a full professor with tenure at the School of Intelligence Science and Technology, Peking University. She has research interest in several areas in connection with intelligent vehicle and mobile robot, such as machine perception, behavior learning and motion planning, and she has special interests on the studies through real world data collection.