2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Vehicle Trajectory Prediction with Soft Behavior Constraints

Ke Ye, Sanping Zhou, Miao Kang, Jingwen Fu, Nanning Zheng

Abstract— **Trajectory prediction plays a crucial role in autonomous driving, but it is challenging due to the multimodal nature of future trajectories. Behavior information is frequently employed to capture more diverse modalities of future trajectories. Traditional behavior information is typically hard-encoded, which is often inaccurate and inadequate for reflecting future multimodality. Therefore, we introduce the concept of soft vehicle behavior, which is represented as a probability distribution over a predefined comprehensive set of behaviors. This approach allows for a more rational depiction of vehicle behavior and captures potential future driving modalities. Based on it, we propose a new soft-behavior-constrained vehicle trajectory prediction framework. The framework consists of a backbone and a lightweight and plug-and-play behavior prediction module, which is used to imbue soft behavior constraints to assist in representation learning. We integrated the behavior prediction module into five representative trajectory predictors and achieved improvements of at least 4.2% in minFDE(K=5) on the nuScenes dataset and 0.5% in minFDE(K=6) on the Argoverse 1 motion forecasting dataset. These universal increments prove the effectiveness and generalizability of soft behavior constraints in vehicle trajectory prediction.**

I. INTRODUCTION

Trajectory prediction is crucial for autonomous driving vehicles to maintain safe and efficient driving. Researchers have invested significant effort into trajectory prediction and have made considerable improvements in this field over the past few years [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16].

However, this problem is not well-solved due to the multimodality of trajectory prediction, which means that there exist several unobserved reasonable future choices besides the ground truth. There are numerous approaches to address this problem, such as GANs [17], [7], CVAEs [18], [19], Goal-based methods [20], [21] and so on. Another approach for multi-modal trajectory predictions utilizes a regression framework with a classification module [22], [23], [24], [25], [26], [27] by incorporating agents' behavior information, which is usually obtained through clustering or manual design. Clusters [22] in embedding space suffers from semantic interpretability, and previous rule-based methods [25], [26], [27] limited to freeway scenes. More importantly, these behavior labels are hard-encoded, which means that each trajectory has a single, specific behavior label. However,

This work is supported by the National Natural Science Foundation of China (Grant No. 62088102). Ke Ye, Sanping Zhou, Miao Kang, Jingwen Fu and Nanning Zheng are all with with the National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Shaanxi 710049, China.

Corresponding author: Nanning Zheng, E-mail: nnzheng@xjtu.edu.cn.

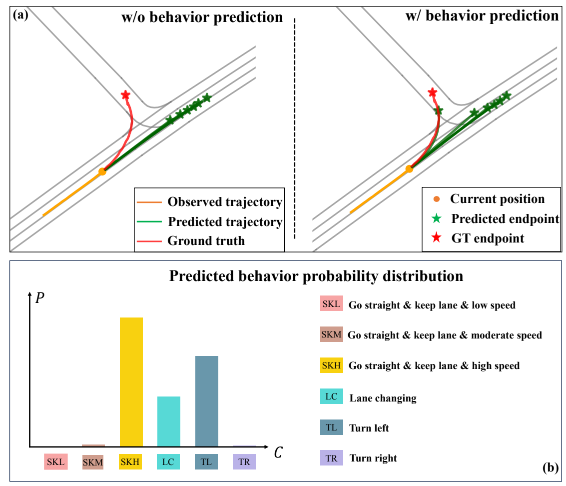


Fig. 1. A complex T-junction sample demonstrating the value of behavior prediction constrains. The top 2 picture (a) shows the predictions from a identical base trajectory prediction model without/with behavior prediction constrains. The bottom picture (b) shows future behavior probability distribution from the behavior prediction module. With the behavior prediction constrains, trajectory prediction model captures the previous ignored turning left modality and covers more possibility, such as changing lane to left.

hard behavior labels cannot describe all vehicle trajectories in real world precisely. For example, for two trajectories which are near a classification boundary and belong to different behavior categories, the significant difference in hard-encoded behavior space does not accurately reflect the relatively small distance variation in the real physical world, which can cause instability during model training and may further damage the performance of the predictive model.

Based on the above thoughts, we introduce soft behavior in vehicle trajectory prediction. Specifically, we devise a novel set of vehicle behaviors firstly, to avoid issue of interpretability from clustering in an embedding space, and each specific behavior includes three independent factors: direction, lane changing, and speed. Then a total of six vehicle behaviors are defined, which are comprehensive, relatively balanced, and simple to use for describing realistic vehicle behaviors. Furthermore, in contrast to the hard behavior labels used in previous works [22], [23], [26], [27], we design *soft behavior labels*, which are probability distributions in behavior space. Compared to hard labels, soft behavior labels keep more original information when compressing trajectories to discrete behavior pattern space, and allow for a more reasonable description of a trajectory. They also reflect latent possibilities of future modalities to some extent, which aids in modeling the nature of multimodality. With the help of soft behaviors, we propose a vehicle trajectory prediction framework incorporating soft behavior constraints to make multi-modal and accurate predictions. Specifically, soft behavior constrains are implemented as an auxiliary behavior prediction task, which is undertaken by a lightweight and -plug-and-play module. Soft behavior constrains imbue semantic information to base trajectory prediction framework to help make more accurate predictions and cover more diverse potential modalities. For instance, as shown in Fig. 1, our framework can capture a diverse range of potential modalities, such as going straight and keeping lanes at high speed, turning left, and changing lanes to the left at a T-shaped junction, which is a complex traffic scenario for an autonomous vehicle to manage, while the base model ignores the possibility of turning left.

We conduct extensive experiments on five prediction backbones and achieve *at least* a 4.2% improvement in minFDE (K=5) in the nuScenes dataset and a 0.5% improvement in minFDE (K=6) in the Argoverse 1 dataset.

Specifically, the contributions of this paper are:

- We introduce soft behaviors for vehicles to describe movements more accurately and reflect different possibilities of future various choices.
- We propose a novel and easy-to-use trajectory prediction framework with soft behavior constraints. This framework consists of a base trajectory predictor and a lightweight, plug-and-play behavior prediction module to assist in representation learning.
- We conduct extensive experiments on five popular models over two large-scale benchmarks, nuScenes and Argoverse 1. Consistent improvements demonstrate the value of the soft behavior constraints.

II. RELATED WORK

Multi-modal Trajectory Prediction. Early works in trajectory prediction were based on kinematic and dynamic models [11], [12], [13], [14], [15], [16]. However, there are multiple reasonable trajectories aside from the observed one, necessitating the use of stochastic trajectory prediction models to capture the inherent multimodality. Researchers have been exploring various methods to pursue more accurate and multi-modal predictions. Generative adversarial networks [17], [7] have been applied to model future trajectory distributions. Conditional variational auto-encoders [18], [19] are employed to capture diverse modalities by sampling latent variables in a learned latent space. Both of these approaches suffer from a lack of interpretability and a potential risk of mode collapse. Another methods leverages the uncertainty of endpoints to model the multimodality of future trajectories. These methods predict probability heatmaps for future goal distributions, from which sampled goals are used to generate full future trajectories [21], [20]. PGP [28] utilizes sampled likely traversed paths to model future diverse possibility. Interaction between agents and environment is also an important factor to estimate future traffic states. LaneGCN [3] and LaPred [29] utilize vectorized HD maps to efficiently model interactions between agents and lanes. FFINet [30] incorporates a feedback module to improve the consistency of multi-agent, multi-modal trajectory prediction. Moreover, numerous transformer-based methods [5], [31], [6] have emerged, showing robust performance in multimodal trajectory prediction. Additionally, behavior patterns can enhance the accuracy and diversity of predictions by constraining representation learning or serving as motion anchors [23], [24], [22].

Multi-modal Trajectory Prediction with Behavior Prediction. Vehicle behaviors reflect intentions of drivers, and help understand and capture the multimodality of future trajectories. Researchers have explored how to effectively combine behavior and trajectory prediction to enhance performance. The coarse-to-fine paradigm is commonly employed in this framework, where future likely motion behaviors are predicted first, followed by the regression of fine-grain trajectories based on behavior prediction results. However, there are three critical problems that degrade the effectiveness and flexibility of behavior prediction.

The first aspect is the completeness of the behavior ground truth space. There are two main approaches for generating behavior labels: clustering and handcrafting. Clustering methods [22], [24], [23] suffer from issues of interpretability and potential mode collapse. On the other hand, existing handcrafted behavior label spaces [25], [27], [26] are limited to freeway scenes.

The second aspect is the accuracy of behavior ground truth. Both clustered pseudo behavior labels [22] and rule-based behavior labels [26] employ one-hot-encoding hard labels. For example, longitudinal behaviors such as normal driving and braking are classified by a speed change ratio threshold of 0.8 [25]. This can lead to inaccuracies for vehicles with speed change ratios near the threshold. Hard-encoded behavior labels fail to express the inherent multimodality in real-world trajectories.

The third aspect is the flexibility of behavior prediction. The coarse-to-fine framework tightly integrates behavior prediction with corresponding works [27]. A more flexible behavior prediction module can provide benefits to more prediction models, rather than a few limited frameworks.

As a solution, we introduce soft behavior labels for vehicles based on our predefined complete behavior set. Based on soft behaviors, we propose a new multi-modal trajectory prediction framework, incorporating a lightweight and plug-and-play behavior module.

III. METHOD

This section is divided into three parts. Firstly, we present the problem formulation of trajectory prediction in Section III-A. Secondly, we introduce our soft-behavior-constrained vehicle trajectory prediction framework and provide a detailed description of the behavior prediction module in Section III-B. Thirdly, we offer an in-depth explanation of soft behaviors, including the definition and algorithm for soft vehicle behavior labels in Section III-C.

A. Problem Formulation

Following previous work [2], [32], [20], we need to construct a model f, which takes the target agent's observed

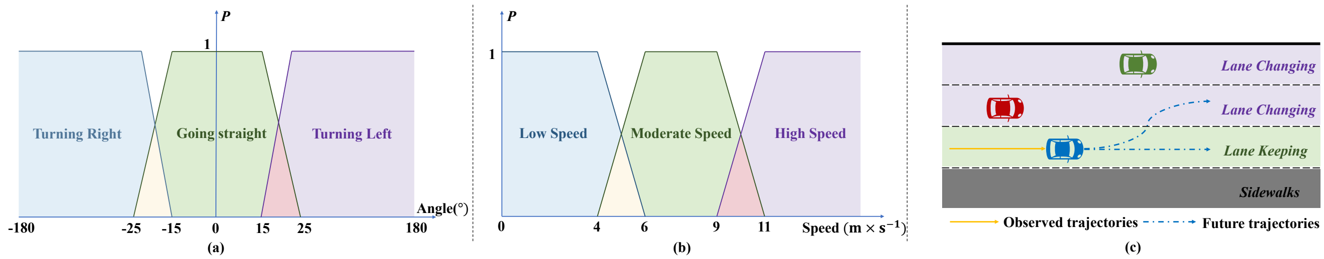


Fig. 2. **(a)/(b)** shows probability distributions of direction/speed categories; **(c)** shows the standard of identifying lane changing.

trajectory $X = [s_1, s_2, \dots, s_{T_{obs}}]$ (s_k stands for the 2D coordinates of the target agent in the k-th frame), neighboring observed trajectories $\mathbb{N} = [X_{n_1}, X_{n_2}, \dots, X_{n_k}]$, and nearby environmental information C as inputs, to predict K future trajectories $\mathbb{Y} = [Y_1, Y_2, \dots, Y_K]$, where $Y_k = [s^k_{T_{obs}+1}, s^k_{T_{obs}+2}, \dots, s^k_{T_{obs}+T_{pred}}]$ with the corresponding probabilities $\mathbb{P} = [p_1, p_2, \dots, p_K]$, as follows:

$$(\mathbb{Y}, \mathbb{P}) = f(X, \mathbb{N}, C). \tag{1}$$

B. Soft-Behavior-Constrain Prediction Framework

Soft behavior constrains is implemented as an auxiliary behavior prediction task, which is undertaken by an lightweight and plug-and-play behavior prediction module. This module can be easily integrated into existing trajectory prediction models with various structures [30], [3], [31], [29], [28]. Specifically, the behavior prediction module takes embedding features that incorporate agent, environment, and interaction information as input and outputs a behavior probability distribution to estimate future driving modalities. We use soft cross-entropy to compute the behavior prediction loss:

$$L_{behavior} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{C} p_{i,k}\log(\hat{p}_{i,k}), \tag{2}$$

where N is the number of trajectories in a batch, C is the number of behavior categories, $p_{i,k}$ is the ground truth probability distribution, and $\hat{p}_{i,k}$ is the predicted behavior likelihood of the i-th sample. Thus, the module can be trained to predict a future behavior distribution that represents possible driver intentions.

C. Soft Behavior Labels

1) Definition of Soft Behavior Labels: Inspired by the previous work [25], [26], [27], we design comprehensive behavior set for a wider scope of application. Firstly, we propose a new set of vehicle behaviors. For the behavior of a vehicle, we describe it from three independent aspects: 1) direction, 2) lane changing, and 3) speed. There are three subcategories for direction: going straight, turning left, and turning right; two subcategories for lane changing: keeping the lane, changing lanes; three subcategories for speed: low speed, moderate speed, and high speed. There are a total of $18 = 3 \times 2 \times 3$ categories for vehicle behaviors. The probability for a particular behavior is given by:

$$P_b = P_d \times P_l \times P_s, \tag{3}$$

where P_b, P_d, P_l, P_s represent probabilities for a specific category of behavior, direction, lane changing, speed of a vehicle. However, compared to other behaviors, going straight always dominates current datasets, so we have divided it into more detailed categories and merged the small categories. Ultimately 18 categories are narrowed down to 6 diverse and relatively balanced categories as shown in Tab. I.

TABLE I
THE DISTRIBUTION OF BEHAVIOR IN ARGOVERSE 1 DATASET

Category	Proportion
Going straight and lane keeping with low speed	10.53%
Going straight and lane keeping with moderate speed	32.15%
Going straight and lane keeping with high speed	33.45 %
Going straight and lane changing	6.3%
Turning left	10.94%
Turning right	6.63%

Then we show the process for identifying the soft behavior of a vehicle. There are many probability functions to determine soft behavior labels, but we choose the simple-to-use trapezoid-shape probability function to obtain soft behavior labels, illustrated in Fig. 2. The difference in directions $\Delta\theta$ between the initial and final direction of a vehicle is the key to identify turns. When $|\Delta\theta|$ is small enough, *i.e.*, less than θ_s in our design, the vehicle is identified as going straight (the green part in Fig. 2 (a)); it will be identified as turning when $|\Delta\theta| > \theta_t$ (the blue/purple part represents turning right/left in Fig. 2 (a)). For moderate $|\Delta\theta|$, where $\theta_s < |\Delta\theta| < \theta_t$, we devise uncertain zones.

$$P(c = turning) = \frac{||\Delta\theta| - \theta_s|}{|\theta_t - \theta_s|} \tag{4}$$

With a slight abuse of the notion of probability distribution, we use trapezoidal probability distributions over turning and going straight by linearly assigning the probabilities (equation 4) based on $|\Delta\theta|$. The soft behavior labels not only describe vehicle behavior more reasonably, and keep original trajectory information as much as possible, but also reflect future diverse driving choices to some extent. For example, the turning process can be represented by the probability of turning. As the turning angle increases, the probability of turning also increases. However, hard behavior labels cannot capture this relationship.

Similarly, there are three levels for vehicle speed S: low speed, moderate speed, high speed, illustrated in the blue part, green part, purple part in Fig. 2 (b), respectively. In the uncertain zones of speed, that is, $S_{u1} < S < S_{u2}$ or $S_{u3} < S < S_{u4}$ (the light yellow/light red part in Fig. 2 (b)), the linear probability assignment principle is also used to capture the speed diversity.

We make a judgment for lane changing according to the lanes where trajectory points locate. For example, in Fig. 2 (c), if the trajectory points of the blue car all lie in the initial green lane, it is identified as keeping the lane; otherwise, its behavior is recognized as lane changing. Since the possibility of lane changing can not simply be reflected by the ratio of the number of trajectory points lying in the original lane or by noisy deviations from lane centerlines, so we do not design uncertainty zones for lane changing, but it is worth to explore in the future.

2) Algorithm for Soft Behavior Labels: In this subsection, we detail the algorithm for identifying the soft behavior label for a vehicle. For a trajectory, we compute the mean speed S, the steering angle $\Delta\theta$, and the most likely lane for each trajectory point, to determine whether the vehicle leaves its initial lane LC or not. Finally, we compute the probabilities based on $S, \Delta\theta, LC$, to obtain the final soft behavior label.

The algorithm 1 representing the rough process of identifying the behavior of a vehicle in pseudo codes is following:

Algorithm 1 Identifying vehicle soft behavior label

Require: observed trajectory X, future ground truth Y, corresponding map M

Ensure: soft behavior label $B \in \mathbb{R}^{6\times 1}$

1. $Traj_{whole} = [X, Y]$
2. $v = Traj_{whole}[1:] - Traj_{whole}[:-1]$
3. $v_s = Mean(v[:t_1])$, $v_e = Mean(v[-t_1:])$
4. $cos(\Delta\theta) = \frac{v_s \cdot v_e}{|v_s||v_e|}$
5. $S = Mean(Norm(v))$
6. Find the most likely lane L_i in M for each point $P_i = (x_i, y_i)$ in $Traj_{whole}$ and make a set of these lanes $L_p = [L_{p_1}, L_{p_2}, L_{p_3}, ...]$
7. Find initial lanes and successors set $L_s = [L_{s_1}, L_{s_2}, L_{s_3}, ...]$
8. $LC = False$ if $L_p \subseteq L_s$ else $LC = True$
9. Calculate P_d, P_s, P_l based on $\Delta\theta, S, LC$, respectively.
10. Calculate probabilities for final six interesting behaviors as the B

IV. EXPERIMENTS

A. Benchmarks

We conduct experiments on two popular and distinct large-scale trajectory prediction benchmarks.

nuScenes[33]: We use the standard split from the nuScenes software kit, which is 32186, 8560, 9041 samples for the training, validation, and test sets, respectively. Each trajectory covers 8 seconds sampled at 2 Hz, with the first 2 seconds used as observations, and the next 6 seconds leveraged to supervise the predictions.

Argoverse 1 Motion Forecasting Dataset[34]: These trajectories are divided into training, validation, and test sets, with 205942, 39472, 78143 samples, respectively. Every trajectory is sampled at 10 Hz with 5 seconds long. Models are required to predict future 3-second trajectories, given the 2-second observed trajectory and high definition maps.

B. Metrics

We follow previous works [30], [31], [29] to apply ADE, FDE, and MR to evaluate prediction performance of the proposed model. Specifically, ADE/FDE measures the average/final displacement error of predicted trajectory and ground truth. Best-of-K means that the minimum error among all K predictions is reported. Miss Rate (MR) is defined as the proportion of scenarios in which the minimum final displacement error of all predictions for the target agent exceeds a specified threshold (2.0 meters in our experiments) relative to the total number of scenarios. It's worth noting that brier-minFDE is unique metric in Argoverse, which is calculated by adding a $(1.0 - p)^2$ penalty to the endpoint L_2 distance, where p corresponds to the probability of the best-forecasted trajectory.

$$brier - minFDE(K) = \min_k ||s_T - s_T^k||_2 + (1-p)^2 \quad (5)$$

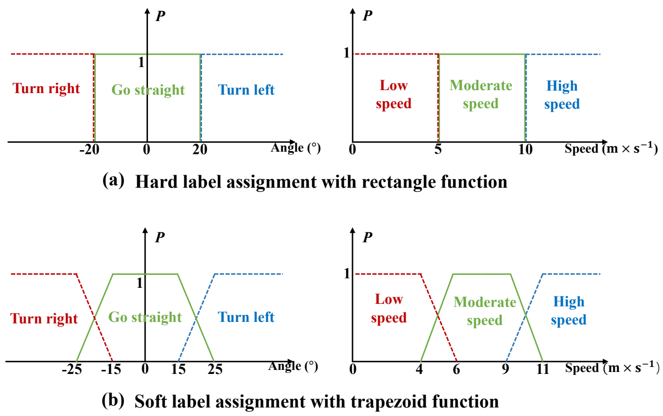
C. Base Models

To comprehensively evaluate the effectiveness and generalizability of soft vehicle behaviors, we select five models with diverse structures, low cost and high performance as baselines. *LaneGCN*[3]: A GNN-based model captures rich interactions between agents and environments. *FFINet*[30]: A GNN-based model introduces interactions among future trajectories. *HiVT*[31]: A transformer-based model can make fast and effective trajectory forecasting. *PGP*[28]: A trajectory prediction framework utilizes sampling in a learned policy and latent space for multi-modal forecasting. *LaPred*[29]: A framework focuses on future most likely lane candidates for scene-compliant predictions.

The final losses for above base models with behavior prediction module are the sum of original losses of base models and behavior prediction loss $L_{behavior}$.

D. Implementation Details

For FFINet+BP, we set the batch size, learning rate, the coefficient of behavior prediction loss to 192, 7.5e-4, 0.1, respectively. And we conduct these experiments on four RTX 3090 Ti GPUs. The next four base models experiments are all conducted on one RTX 3090 Ti GPU. For LaneGCN+BP, the coefficients of regression, classification, behavior prediction loss are set to 1.2, 2.0, 0.8, respectively; For HiVT+BP, the coefficient of behavior prediction loss is 0.01, and the dimension of embedding space is 64; For LaPred+BP, the coefficient of behavior prediction loss is 1, learning rate is set to 7.5e-4; For PGP+BP, we use a behavior prediction loss coefficient of 0.05. Due to sampling randomness, we present the mean results from five experiments. θ_s and θ_t is set to $15°$ and $25°$; $S_{u1}, S_{u2}, S_{u3}, S_{u4}$ is set to 4, 6, 9, 11 m/s, which are determined from a analysis for steering angle and speed distribution of datasets. Unless otherwise specified, the behavior prediction module is implemented with two-layer MLP with a hidden state dimension of 128.



(a) Hard label assignment with rectangle function

(b) Soft label assignment with trapezoid function

Fig. 3. Different vehicle behavior label assignments.

E. Main Results

We will analyze and discuss the value of soft vehicle behavior quantitatively. Results in Tab. II and III clearly show that with the soft behavior prediction module, there are consistent improvements for all five base models on the nuScenes and Argoverse 1 datasets in most metrics, compared with corresponding original models. For FFINet and PGP, which already achieve impressive prediction results in Argoverse 1 and nuScenes dataset, respectively, the auxiliary behavior prediction task brings 2% boost in brier-minFDE(K=6) for FFINet, and bring 4.2% improvement in minFDE(K=5) for PGP. These results indicate that introducing soft behavior constraints is beneficial for predicting more accurate and multi-modal trajectories.

F. Analysis

Soft Behavior Label. We take direction and speed subcategories as examples to discuss the effectiveness of hard and soft behaviors, as shown in Fig. 3. We obtain hard and soft behavior labels for a vehicle by the rectangle-shape and trapezoid-shape probability distribution functions, respectively. As shown in Tab. IV, with behavior prediction module, the predictor can achieve superior performance compared with the original model consistently, which demonstrates the effectiveness of behavior prediction. Furthermore, compared with hard behavior labels, soft behavior labels provide a 2% performance improvement in brier-minFDE6 and minFDE6, highlighting the advantages of soft behavior constraints. Soft behavior labels can be obtained by many other probability functions, such as continuous gaussian-like functions, the best form of probability functions can be explored in the future.

TABLE IV

COMPARISON ON VEHICLE BEHAVIOR LABEL. (K=6)

Method	Behavior	b-minFDE	minFDE	minADE	MR
FFINet	-	1.602	0.923	0.655	0.078
FFINet+BP	Hard	1.591	0.912	0.651	0.078
FFINet+BP	Soft	**1.582**	**0.903**	**0.649**	**0.077**

Structures of Behavior Prediction Module. In this section, we conduct experiments on different implementations of the behavior prediction module in Argoverse 1 dataset. There are three structures: standard self-attention, a residual block consisting of a two-layer MLP with a residual connection, and a standard two-layer MLP. As shown in Table V, the behavior predictor with all three structures leads to improved performance in most metrics. The residual block provides the greatest boost in multi-modal prediction metrics, while the

TABLE II

COMPARISON WITH THE STATE-OF-THE-ART METHODS IN NUSCENES MOTION FORECASTING VALIDATION SET. ↓ MEANS THE LOWER THE METRIC IS, THE BETTER PREDICTIONS ARE. 'BP' MEANS BEHAVIOR PREDICTION MODULE. **BOLD** INDICATES THE BEST PERFORMANCE.

Method	minADE(K=5) ↓	minFDE(K=5) ↓	minADE(K=10) ↓	minFDE(K=10) ↓	minADE(K=1) ↓	minFDE(K=1) ↓
LaPred	1.583	3.235	**1.220**	**2.253**	3.550	8.369
LaPred+BP	**1.537**	**3.130**	1.240	2.324	**3.528**	**8.355**
Impr.%	*+ 4.6*	*+ 10.5*	*- 2.0*	*- 7.1*	*+ 2.2*	*+ 1.4*
PGP	1.280	2.502	0.954	1.580	3.162	7.362
PGP+BP	**1.264**	**2.460**	**0.936**	**1.536**	**3.124**	**7.250**
Impr.%	*+ 1.6*	*+ 4.2*	*+ 1.8*	*+ 4.4*	*+ 3.8*	*+ 11.2*

TABLE III

EXPERIMENTS FOR BEHAVIOR PREDICTION MODULE IN ARGOVERSE 1. B-MINFDE STANDS FOR BRIER-MINFDE.

Method	b-minFDE(K=6) ↓	minFDE(K=6) ↓	minADE(K=6) ↓	MR(K=6)↓	minFDE(K=1) ↓	minADE(K=1) ↓	MR(K=1) ↓
FFINet	1.602	0.923	0.655	0.078	2.618	1.210	0.443
FFINet+BP	**1.582**	**0.903**	**0.649**	**0.077**	**2.604**	**1.209**	**0.439**
Impr.%	*+ 2*	*+ 2*	*+ 0.6*	*+ 0.1*	*+ 1.4*	*+ 0.1*	*+ 0.4*
LaneGCN	1.756	1.087	0.710	**0.102**	3.071	1.382	0.505
LaneGCN+BP	**1.752**	**1.081**	**0.709**	0.103	**2.881**	**1.311**	**0.486**
Impr.%	*+ 0.4*	*+ 0.6*	*+ 0.1*	*- 0.1*	*+ 19*	*+ 7.1*	*+ 1.9*
HiVT	1.689	1.026	0.686	0.103	3.893	1.758	0.678
HiVT+BP	**1.683**	**1.021**	**0.685**	**0.100**	**3.865**	**1.750**	**0.670**
Impr.%	*+ 0.6*	*+ 0.5*	*+ 0.1*	*+ 0.3*	*+ 2.8*	*+ 0.8*	*+ 0.8*

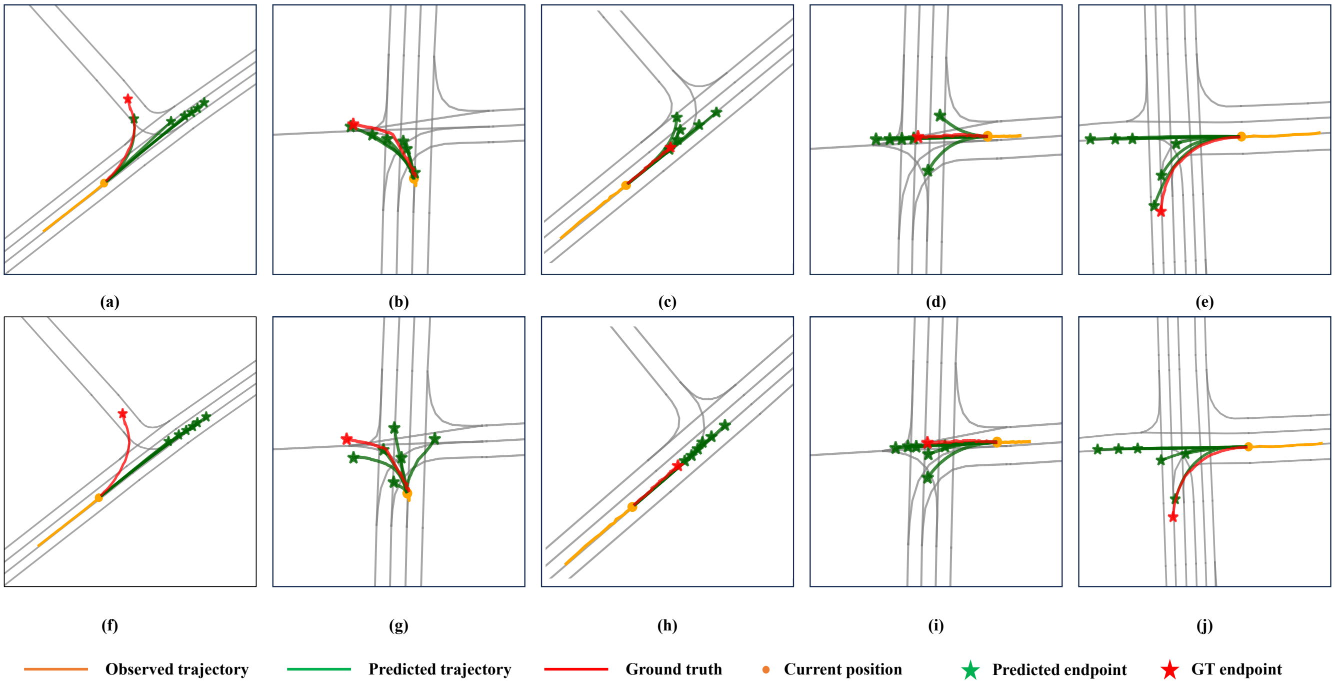


Fig. 4. Visualized prediction results comparison on Argoverse 1 validation set. The top row((a)-(e)) shows the predictions of base model with behavior prediction module; while the bottom row((f)-(j)) presents the results of original base model. The grey lines in the pictures presents lane centerlines. FFINet is chosen as the compared base model. Best view in colors.

standard two-layer MLP performs best in unimodal metrics.

Memory And Time Cost. According to Table VI, the lightweight behavior prediction module enhances prediction performance of the base model by 2% in terms of brier-minFDE (K=6), with a few additional memory usage and negligible inference time cost.

TABLE VI

COMPARISON ON COMPUTATIONAL COST. (K=6)

Method	b-minFDE	Params(M)	FLOPs(G)	Inference Time(ms)
FFINet	1.602	6.1998	2.6698	5.5472
FFINet+BP	1.582	6.2171	2.6701	5.5652

G. *Qualitative Analysis*

There are several typical and challenging scenes in Fig. 4, such as cross intersections and T-junctions. The topology of lane networks reflects many feasible choices (e.g. going straight, turning) for drivers in these scenes, which makes it difficult to predict accurate and multi-modal trajectories. By comparing (a), (c) with (f), (h), it is obvious that soft behavior constraints can alleviate the problem about the lack of diversity deficiency and inaccuracies of original predictions. On the other hand, the comparison between (b), (e) and (g), (j) shows that, beneficial from the soft behavior constraints, our model can predict more plausible and accurate trajectories, which are more consistent with the topology of lanes. Furthermore, our model strikes a good balance between accuracy and multimodality, as shown in the comparison between (d) and (i).

V. CONCLUSIONS

In this paper, we introduce soft vehicle behavior, allowing for a more reasonable description of a trajectory in a comprehensive behavior space and capturing the latent diverse intentions of drivers for multi-modal trajectory prediction. Building on soft behaviors, we propose a soft-behavior-constrained vehicle behavior prediction framework for multi-modal vehicle trajectory prediction. This framework consists of a base trajectory predictor and a lightweight, plug-and-play behavior predictor to assist in representation learning. Extensive experiments are conducted on five representative models over two large-scale benchmarks: nuScenes and

TABLE V

COMPARISON ON THE STRUCTURE OF BEHAVIOR PREDICTION MODULE IN ARGOVERSE 1 MOTION FORECASTING VALIDATION DATASET.

Method	Structure	b-minFDE(k=6)	minFDE(K=6)	minADE(K=6)	MR(K=6)	minFDE(K=1)	minADE(K=1)	MR(K=1)
LaneGCN	-	1.756	1.087	0.710	**0.102**	3.071	1.382	0.505
LaneGCN+BP	MLP	1.752	1.081	0.709	0.103	**2.881**	**1.311**	0.486
LaneGCN+BP	Residual Block	**1.747**	**1.075**	**0.708**	0.104	2.908	1.323	**0.483**
LaneGCN+BP	Self-Attention	1.763	1.091	0.717	0.103	2.960	1.346	0.493

Argoverse 1 datasets, and the superior prediction results demonstrate the generalizability and effectiveness of soft behavior constraints in vehicle trajectory prediction.

Limitations and Future Work. Firstly, while we focus on soft behavior constraints specifically for vehicles, this philosophy can be easily applied to other important traffic participants, such as pedestrians. Secondly, despite consistent improvements across many models, the effect of soft behavior constraints is not significant. In the future, we will explore wider applications and more effective ways to apply soft behavior constraints in trajectory prediction.

REFERENCES

[1] N. Ye, Y. Zhang, R. Wang, and R. Malekian, "Vehicle trajectory prediction based on hidden markov model," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 10, no. 7, pp. 3150–3170, 2016.

[2] 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.

[3] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun, "Learning lane graph representations for motion forecasting," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 541–556.

[4] Y. Sun, T. He, J. Hu, H. Huang, and B. Chen, "Socially-aware graph convolutional network for human trajectory prediction," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2019, pp. 325–333.

[5] S. Shi, L. Jiang, D. Dai, and B. Schiele, "Motion transformer with global intention localization and local movement refinement," *Advances in Neural Information Processing Systems*, vol. 35, pp. 6531–6543, 2022.

[6] Z. Zhou, J. Wang, Y.-H. Li, and Y.-K. Huang, "Query-centric trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 863–17 873.

[7] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, "Sophie: An attentive gan for predicting paths compliant to social and physical constraints," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1349–1358.

[8] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," 2021.

[9] L. Shi, L. Wang, C. Long, S. Zhou, M. Zhou, Z. Niu, and G. Hua, "Sgcn: Sparse graph convolution network for pedestrian trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8994–9003.

[10] X. Zhang, G. Liu, C. Hu, and X. Ma, "Wavelet analysis based hidden markov model for large ship trajectory prediction," in *2019 Chinese Control Conference (CCC)*. IEEE, 2019, pp. 2913–2918.

[11] S. Ammoun and F. Nashashibi, "Real time trajectory prediction for collision risk estimation between vehicles," in *2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing*, 2009, pp. 417–422.

[12] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *2008 11th International Conference on Information Fusion*, 2008, pp. 1–6.

[13] A. Polychronopoulos, M. Tsogas, A. J. Amditis, and L. Andreone, "Sensor fusion for predicting vehicles' path for collision avoidance systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 549–562, 2007.

[14] P. Lytrivis, G. Thomaidis, and A. Amditis, "Cooperative path prediction in vehicular environments," in *2008 11th International IEEE Conference on Intelligent Transportation Systems*, 2008, pp. 803–808.

[15] A. Barth and U. Franke, "Where will the oncoming vehicle be the next second?" in *2008 IEEE Intelligent Vehicles Symposium*, 2008, pp. 1068–1073.

[16] T. Batz, K. Watson, and J. Beyerer, "Recognition of dangerous situations within a cooperative group of vehicles," in *2009 IEEE Intelligent Vehicles Symposium*, 2009, pp. 907–912.

[17] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2255–2264.

[18] K. Mangalam, H. Girase, S. Agarwal, K.-H. Lee, E. Adeli, J. Malik, and A. Gaidon, "It is not the journey but the destination: Endpoint conditioned trajectory prediction," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 759–776.

[19] 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.

[20] 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.

[21] 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.

[22] J. Sun, Y. Li, H.-S. Fang, and C. Lu, "Three steps to multimodal trajectory prediction: Modality clustering, classification and synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 250–13 259.

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

[24] 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.

[25] 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 (CVPR) Workshops*, June 2018.

[26] R. Wang, S. Wang, H. Yan, and X. Wang, "Wsip: Wave superposition inspired pooling for dynamic interactions-aware trajectory prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, 2023, pp. 4685–4692.

[27] X. Chen, H. Zhang, F. Zhao, Y. Hu, C. Tan, and J. Yang, "Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19 471–19 483, 2022.

[28] 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.

[29] 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.

[30] M. Kang, S. Wang, S. Zhou, K. Ye, J. Jiang, and N. Zheng, "Ffinet: Future feedback interaction network for motion forecasting," *arXiv preprint arXiv:2311.04512*, 2023.

[31] Z. Zhou, L. Ye, J. Wang, K. Wu, and K. Lu, "Hivt: Hierarchical vector transformer for multi-agent motion prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8823–8833.

[32] 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 (CVPR)*, June 2020.

[33] 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.

[34] 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.