BIPP: Bidirectional Interactive Prediction and Planning Toward Safe and Efficient Autonomous Driving

Xiang Zhao, Keqiang Li, Jianqiang Wang, Yunhao Hu, Hanxiao Ren, and Yugong Luo

***Abstract*—Safe and efficient autonomous driving in dense urban environments requires both accurate modeling of multi-agent interaction and robust closed-loop execution. However, many existing methods still use a one-pass sequential predict-then-plan implementation, in which the current plan does not feed back to the current surrounding-agent prediction and the fixed update order introduces a structural asymmetry between the autonomous vehicle (AV) and surrounding agents. In addition, purely learning-based planners often suffer from distribution shift and error accumulation during closed-loop execution, and consequently rely on heavy post-processing to maintain safety. To address these limitations, we propose BIPP, a Bidirectional Interactive Prediction and Planning framework for autonomous driving. BIPP reformulates conventional sequential decision making as a parallel-iterative process, in which prediction and planning are performed in parallel within each stage and progressively refined across stages using shared future context. A unified bidirectional interaction modeling mechanism jointly leverages historical motion evidence and iteratively updated future intentions to capture mutual adaptation between the AV and surrounding agents. Moreover, a hybrid planning strategy combines rule-based lateral trajectory generation with a learning-based longitudinal policy to improve feasibility, adaptability, and closed-loop performance. Closed-loop experiments on nuPlan and interPlan show that BIPP achieves the best overall performance among the compared baselines in the evaluated benchmark settings, with notable gains in highly interactive and long-tail scenarios.**

***Index Terms*—Autonomous driving, trajectory planning, prediction, bidirectional interactions, closed-loop performance.**

I. Introduction

A. Motivation and Background

ACHIEVING both safety and efficiency in traffic remains a fundamental challenge for autonomous driving [1], [2], [3], [4], [5], [6]. In complex scenarios such as signalized intersections, unprotected turns, merging, and yielding, the ego vehicle must continuously adapt to multiple surrounding agents whose intentions are coupled, evolving, and inherently multimodal [7], [8], [9]. In such environments, safe driving requires the autonomous vehicle (AV) to avoid hazardous interactions, while efficient driving requires it to avoid overly conservative behaviors that unnecessarily sacrifice progress. Balancing these two objectives is particularly difficult because traffic participants mutually influence one another, and even small errors in interaction reasoning may lead either to unsafe aggressiveness or to excessive hesitation [8], [10], [11].

Despite the substantial progress from traditional modular pipelines to learning-based and end-to-end frameworks [10], [11], many existing methods still rely on a sequential predict-then-plan pipeline. In its common one-pass implementation, prediction is first performed for surrounding agents, and planning is then conditioned on the predicted futures [12], [13]. As a result, reciprocal interaction is only weakly captured, which may lead to either overly conservative or unsafe decisions in complex urban negotiations. Moreover, many recent methods still formulate future prediction or planning over the horizon in a one-shot manner, which limits their ability to capture the iterative evolution of scene context and agent intentions over time [14], [15], [16]. This further limits their ability to handle dynamically changing interactions in complex urban environments. A representative example is illustrated in Fig. 1(a). When the AV attempts to merge right to bypass a construction zone, the traditional sequential paradigm predicts that the adjacent vehicle will maintain its motion and therefore induces a conservative deceleration strategy, ultimately causing the AV to miss a feasible merging opportunity.

Furthermore, although recent learning-based methods have achieved impressive open-loop performance on metrics such as average displacement error (ADE) and final displacement error (FDE) [17], [18], strong open-loop accuracy does not necessarily translate into robust closed-loop driving performance [19], [20], [21]. Purely learning-based planners are often vulnerable to distribution shift and compounding errors once deployed in interactive environments, where small deviations may propagate over time and eventually degrade safety, feasibility, or route progress [19], [20]. To compensate for these issues, many methods depend on expensive post-processing or optimization-based refinement [8], [11], [21], which may improve safety but can also weaken behavioral consistency and adaptability.

To address these limitations, we propose *BIPP*, a novel *Bidirectional Interactive Prediction and Planning* framework toward safe and efficient autonomous driving. BIPP explicitly models the reciprocal interactions between prediction and planning, enabling both the AV and surrounding agents

Received 25 March 2026; revised 24 July 2026; accepted 30 August 2026. This work was supported by the National Natural Science Foundation of China under Grant 52221005 and Grant 52472434. The Associate Editor for this article was Y. Wiseman. *(Corresponding author: Yugong Luo.)*

The authors are with the School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China (e-mail: lyg@mail.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TITS.2026.3730243

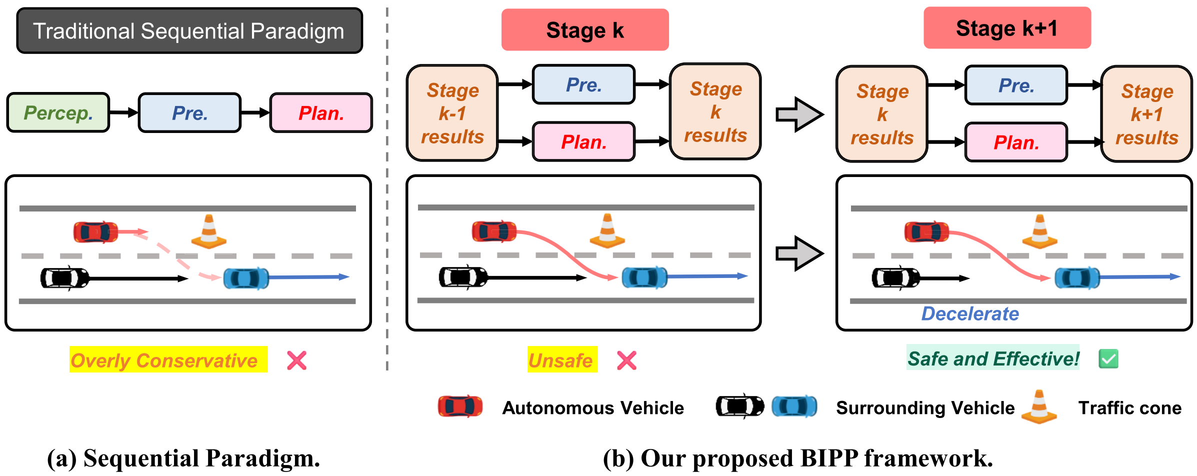


Fig. 1. Illustration of the proposed BIPP framework in a challenging interactive scenario. The AV attempts to merge right to bypass a construction zone, but dense traffic in the target lane prevents direct insertion. In the traditional sequential paradigm, the AV predicts surrounding agents before planning and, assuming the adjacent vehicle (black surrounding vehicle) will maintain speed, adopts a conservative deceleration strategy that fails to complete the merge. In contrast, the proposed parallel-iterative paradigm iteratively refines prediction and planning with explicit bidirectional interaction modeling, allowing the adjacent vehicle to respond to the AV's intention by yielding and enabling a safe and efficient passing maneuver.

to be reasoned about in a unified bidirectional manner, as shown in Fig. 1(b). Different from system-level parallel architectures such as PARA-Drive [22], BIPP uses stage-wise prediction–planning parallelism to remove order-sensitive dependencies and enable reciprocal decision interaction in closed-loop planning. In addition, we introduce a hybrid planning strategy that combines the strengths of rule-based priors and learning-based generation to improve closed-loop performance and mitigate error accumulation. By jointly enhancing interaction modeling and execution stability, the proposed framework aims to achieve a better balance between safety and efficiency in complex urban driving scenarios.

B. Related Work and Limitations

1) Integration of Prediction and Planning: In conventional autonomous driving systems, prediction and planning are typically organized as two decoupled modules in a sequential pipeline. Although such a decomposition simplifies system design and improves interpretability, it limits the ability to capture the reciprocal influence between the AV and surrounding agents, especially in dense urban environments where interactions are highly coupled and continuously evolving. To better account for this coupling, recent studies have attempted to integrate prediction and planning more tightly. Existing approaches can be broadly grouped into three paradigms.

a) Sequential integration: Sequential integration organizes prediction and planning as causally ordered modules, resulting in a unidirectional leader-follower interaction pattern. One line of work first predicts the future behaviors of surrounding agents and then plans the AV trajectory accordingly, corresponding to a *human-leader* formulation [12], [23]. Another line first plans the AV behavior and then predicts how other agents may respond, which can be viewed as a *robot-leader* formulation [24], [25], [26]. However, such methods preserve an explicit order between prediction and planning, which makes it difficult to model the reciprocal influence between the AV and surrounding agents in a symmetric and adaptive manner.

b) Joint integration: Joint integration employs a unified model to jointly reason about the future trajectories of the AV and surrounding agents [16], [27], [28], [29], [30]. Representative methods include InteractionNet [31], which uses local and global Transformers to capture individual motion patterns and inter-agent interactions, and GameFormer [8], which incorporates game-theoretic reasoning to iteratively refine multi-agent futures. SafePathNet [32] also models AV-agent interactions within a Transformer-based framework and outputs probabilistic multimodal trajectories. However, by treating prediction and planning too homogenously within a unified formulation, these methods may underemphasize the stronger safety, feasibility, and closed-loop requirements of planning. Moreover, AV–agent interactions are often modeled only implicitly from shared historical context, which makes it difficult to capture the explicit and progressively evolving mutual adaptation between the ego vehicle and surrounding agents.

c) Co-leader integration: Co-leader integration retains separate prediction and planning modules, while introducing iterative information exchange between them to explicitly model how the AV influences surrounding agents and how those agents, in turn, react to the AV [33], [34], [35], [36]. PPAD [37], for example, alternates between prediction and planning over time to simulate spatiotemporal interaction. Tree-based approaches [34] perform multi-stage interaction reasoning through tree expansion. DTPP [38] generates conditional predictions based on the AV's multimodal plans, and then selects high-scoring trajectories for subsequent tree-policy planning and conditional prediction. Compared with one-shot methods, these approaches better reflect the iterative nature of real-world interaction. However, prediction and planning

TABLE I
COMPARISON OF REPRESENTATIVE PREDICTION–PLANNING ARCHITECTURES

Integration Paradigm	Representative Method	Coupling Scheme	Stage Order	Sequential Dependency	Interaction Context	Planning-Specific
Sequential	P3 [12]	One-shot	Pre-then-Plan	Yes	Pred. SA	Yes
	CMP [26]	One-shot	Plan-then-Pre	Yes	No	Yes
Joint	GameFormer [8]	Iterative	Unified	No	Pred. SA+AV	No
	SafePathNet [32]	One-shot	Unified	No	Hist. traj.	No
Co-leader	DTPP [38]	Iterative	Plan-then-Pre	Yes	Pred. SA	Yes
	BIPP	Iterative	Parallel	No	Pred. SA+AV and hist. traj.	Yes

Notes: Pred. = predicted; hist. traj. = historical trajectories; SA = surrounding agents; AV = autonomous vehicle. Sequential Dependency denotes a fixed prediction–planning order within a stage; Interaction Context lists only the dynamic trajectory features used by planning; Map Context is excluded. Planning-Specific denotes dedicated safety and feasibility designs.

are still executed sequentially within each stage, leaving the inherent limitations of unidirectional modeling unresolved and potentially leading to progressive error amplification across stages.

As summarized in Table I, BIPP retains task-specific prediction and planning branches, updates them in parallel within each stage, and conditions both branches on historical trajectories together with the joint future context propagated across stages. This design grounds iterative interaction refinement in observed motion evidence, preserves planning-specific modeling, and avoids a fixed within-stage execution order.

2) Trajectory Planning: Trajectory planning is a core component of autonomous driving, responsible for generating safe and efficient vehicle motions under dynamic environmental constraints. Existing planning approaches can be broadly categorized into rule-based and learning-based methods.

a) Rule-based: Traditional rule-based planners typically rely on manually designed heuristics, graph search over lane networks, or constrained optimization to produce feasible trajectories that satisfy vehicle dynamics and traffic rules [39], [40], [41], [42], [43]. Their main advantages lie in interpretability, controllability, and the ability to explicitly enforce safety constraints. However, these methods are usually embedded in modular pipelines. Such a decoupled design can cause information loss and error propagation across modules [44]. While rule-based planners offer strong interpretability and explicit feasibility control [19], they often require extensive manual engineering and may struggle to scale to highly interactive urban scenarios.

b) Learning-based: With the rapid development of deep learning, learning-based planning has attracted increasing attention. By learning planning policies directly from large-scale driving data, these methods can generate future trajectories conditioned on scene context and route goals, thereby reducing the reliance on hand-crafted rules [15], [45], [46]. However, despite their promising results in open-loop evaluation, recent benchmarks emphasize that conventional displacement-based metrics cannot fully reflect actual driving quality during closed-loop execution [47]. A substantial gap persists between open-loop training and closed-loop deployment, where distribution shifts, feedback-induced error accumulation, and insufficient reactivity severely degrade both safety and efficiency [19], [21], [48]. Although techniques such as data augmentation [49], [50] and auxiliary losses [50], [51] are employed to mitigate these issues, many state-of-the-art learning-based planners still rely heavily on optimization-based post-processing to guarantee safety [8], [11], [21].

In summary, existing approaches still face two unresolved challenges. First, they have not fully captured the bidirectional and dynamically evolving interaction between the AV and surrounding agents, either due to asymmetric interaction modeling with sequential information exchange at each stage or due to inferring interactions solely from static historical behaviors. Second, they remain susceptible to closed-loop distribution shift and error accumulation, often requiring conservative post-processing to guarantee safety. These limitations motivate the need for a framework that can explicitly model bidirectional interaction while mitigating error accumulation, thereby achieving a better balance between safety and efficiency in complex urban driving.

C. Contributions

The main contributions of this work are summarized as follows:

1) We propose a bidirectional parallel-iterative prediction–planning framework that removes the order-sensitive dependency of sequential pipelines. Prediction and planning are executed in parallel within each stage and refined across stages using the preceding joint future context, enabling reciprocal interaction reasoning while reducing the conservative or aggressive biases caused by a fixed execution order.
2) We develop a unified bidirectional interaction mechanism that models the evolving intentions of the AV and surrounding agents. By combining historical motion evidence with iteratively updated future intentions in both prediction and planning, the mechanism supports mutual adaptation across stages while retaining consistency with observed behaviors.
3) We introduce a multimodal hybrid planning strategy that combines rule-based lateral trajectory generation

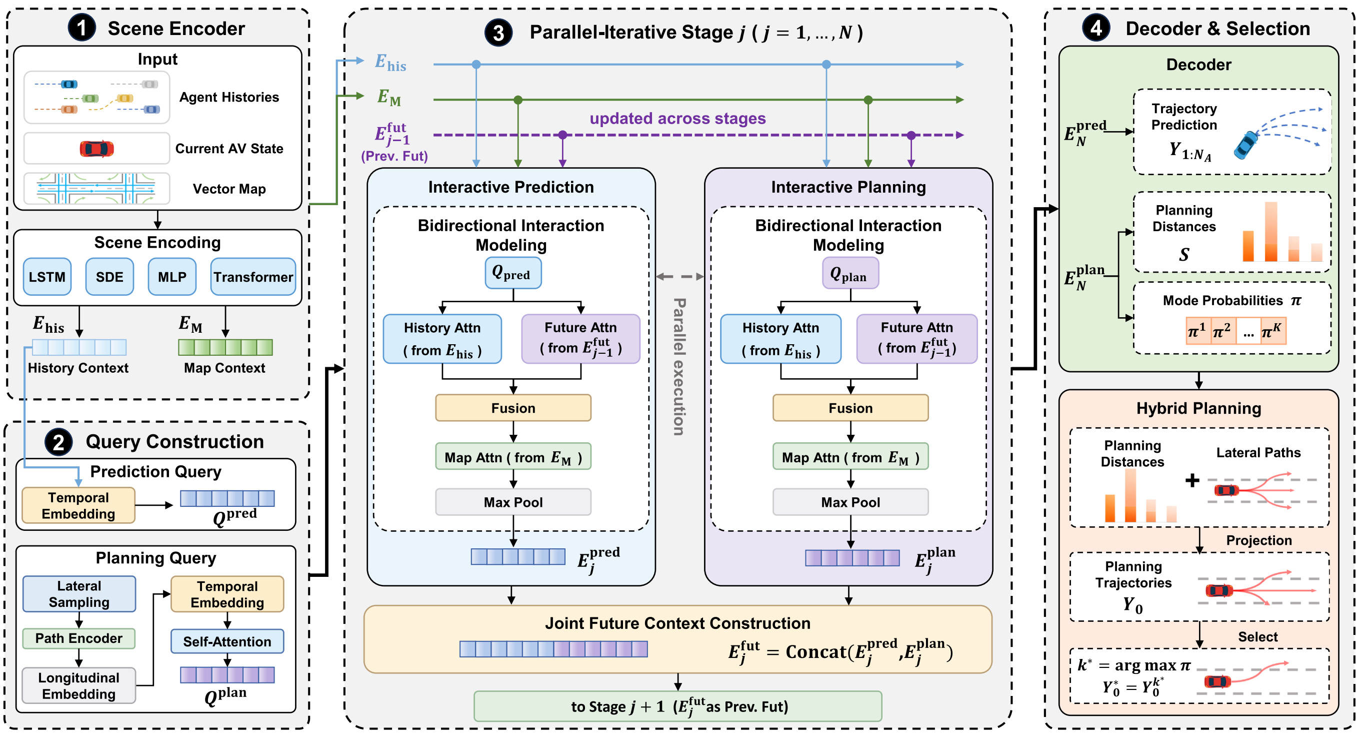


Fig. 2. Overall architecture of BIPP, consisting of scene encoding, query construction, parallel-iterative interaction, and trajectory decoding/selection. At each stage, prediction and planning use task-specific queries to attend to shared historical, map, and previous-stage future contexts; their outputs are concatenated and propagated to the next stage.

with an interaction-conditioned learning-based longitudinal policy. This design preserves lane-level geometric feasibility while enabling longitudinal decisions to adapt to evolving interactions, improving closed-loop stability without relying on heavy optimization-based post-processing.

D. Organization of the Paper

The remainder of the paper is organized as follows. Section II formally presents the overall problem formulation. Section III details the proposed BIPP framework, including the parallel-iterative architecture, bidirectional interaction modeling mechanism, and hybrid planning strategy. Section IV describes the experimental setup. The main results are discussed in Section V. Section VI concludes the paper and discusses future work.

II. Problem Statement

We consider a multi-agent driving scenario where A_0 denotes the AV, $\{A_i\}_{i=1}^{N_A}$ denote the surrounding agents (SAs), $\mathcal{M}$ denotes the map context, and $X_{0:N_A}$ denotes the historical observation trajectories. Our goal is to generate K candidate future trajectories $\boldsymbol{Y}_0$ for the AV and $\boldsymbol{Y}_{1:N_A}$ for all SAs over a planning horizon t_f. The process is mathematically formulated as:

$$\left\{\left(\boldsymbol{Y}_0^{(k)}, \boldsymbol{Y}_{1:N_A}^{(k)}, \pi^{(k)}\right)\right\}_{k=1}^{K} = f_\phi\left(X_{0:N_A}, \mathcal{M}\right), \tag{1}$$

where f_ϕ denotes the neural network of BIPP parameterized by ϕ and $\boldsymbol{\pi} = \left\{\pi^{(k)} \mid k = 1, \ldots, K\right\}$ denotes the confidence probabilities of the K joint future modes.

The objective is to generate a final ego trajectory that remains close to expert driving behavior while achieving a favorable balance between safety and efficiency in interactive driving. BIPP generates a finite set of joint future modes and selects the final ego plan according to the learned confidence score:

$$k^* = \arg \max_{k \in \{1,\ldots,K\}} \pi^{(k)}, \qquad \boldsymbol{Y}_0^* = \boldsymbol{Y}_0^{(k^*)}. \tag{2}$$

Here, $\pi^{(k)}$ denotes the learned confidence of the k-th joint future mode. The selected ego trajectory $\boldsymbol{Y}_0^*$ is expected to balance safety and efficiency, such as favoring collision avoidance, sufficient time-to-collision (TTC) margins, and steady progress along the route.

III. Methodology

We propose BIPP, a parallel-iterative framework that models bidirectional relationships between prediction and planning modules, as shown in Figure 2.

A. Architecture Overview

Figure 2 illustrates the overall architecture of BIPP using four connected components: scene encoding, query construction, parallel-iterative interaction, and decoding/selection. The Scene Encoder first converts agent histories, the current AV state, and vectorized map elements into the historical context E_{his} and map context E_{M}. For the planning branch, Query Construction encodes sampled lateral paths as geometric priors and combines them with longitudinal embeddings to form Q^{plan}. At each stage j, the Interactive Prediction and Interactive

Planning branches are executed in parallel, and each branch instantiates the Bidirectional Interaction Modeling procedure with its own task-specific query. Each task query attends to E_{his}, E^{fut}_{j-1}, and E_{M}, linking history, evolving future intention, and map topology in the tensor flow. The resulting prediction context E^{pred}_{j} and planning context E^{plan}_{j} are concatenated into the joint future context E^{fut}_{j}, which is propagated to stage $j+1$ for iterative refinement. After N stages, the decoder produces multimodal trajectories for surrounding agents, planning distances, and mode probabilities, and the hybrid planning module projects the decoded distances onto lateral paths to select the final AV trajectory.

B. Input Representation and Scene Encoding

1) Agent History Encoding: We represent each agent's observational state at time t as $s^t_i = (p^t_i, \theta^t_i, v^t_i, \omega^t_i, b^t_i, \mathbf{e}_{c_i})$, where (p, θ) denote position and heading, (v, ω) capture velocity and yaw rate, b defines bounding box dimensions, and $\mathbf{e}_{c_i} \in \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ indicates the category of agent i (vehicle, pedestrian, or bicycle). Historical state sequences $S_{\text{p}} \in \mathbb{R}^{N_{\text{A}} \times t_{\text{h}} \times d_{\text{s}}}$ are processed through an LSTM encoder, yielding agent embeddings $E_{\text{A}} \in \mathbb{R}^{N_{\text{A}} \times D}$, where D is the consistent hidden dimension throughout our architecture.

2) AV State Encoding: Recent studies indicate that imitation learning models are often prone to causal confusion by exploiting shortcuts in historical trajectory data rather than reacting to dynamic environments [52]. To mitigate this, we utilize only the current state of the AV, denoted as $s_{\text{AV}} = [p, \theta, v, a]$ (position, heading, velocity and acceleration). We employ an attention-based State Dropout Encoder (SDE) to process this state following [52]. By randomly masking input dimensions during training, the SDE prevents the model from extrapolating kinematic states, yielding a robust AV embedding $E_{\text{AV}} \in \mathbb{R}^{D}$.

3) Vectorized Map Encoding: The map information $\mathcal{M}$ consists of map elements such as lanes, routes, and crosswalks, which are represented as a set of vectorized map polylines. Each polyline is described by a sequence of points with associated attributes (*e.g.*, position, direction). The traffic light status is embedded into the lane features as a one-hot encoded vector. We identify a total of N_{m} nearby map elements, each containing N_{p} waypoints with d_{m}-dimensional feature. Accordingly, the map information is represented as a matrix $\mathcal{M} \in \mathbb{R}^{N_{\text{m}} \times N_{\text{p}} \times d_{\text{m}}}$. The map polylines are encoded using the multi-layer perception (MLP) network, resulting in map feature tensor $E_{\text{p}} \in \mathbb{R}^{N_{\text{m}} \times N_{\text{p}} \times D}$. To reduce the number of map tokens, we group waypoints associated with the same map element and apply max-pooling to aggregate their features. The resulting feature tensor is reshaped to $E_{\text{M}} \in \mathbb{R}^{N_{\text{mp}} \times D}$, where N_{mp} is the number of map tokens after pooling.

4) Scene Encoding: To effectively capture the interactions among the AV, SAs, and map elements, we concatenate the agent encoding E_{A}, AV encoding E_{AV}, and map feature tensor E_{M} to form a unified scene encoding tensor $E_{\text{S}} \in \mathbb{R}^{(N_{\text{A}} + N_{\text{mp}} + 1) \times D}$. This tensor is then fed into a Transformer encoder with N_{e} layers, producing in a final scene representation $E_{\text{env}} \in \mathbb{R}^{(N_{\text{A}} + N_{\text{mp}} + 1) \times D}$, which captures the contextual relationships among all entities within the scene. For notational simplicity, the agent, AV, and map encodings processed by the Transformer encoder will be referred to as E_{SA}, E_{AV}, and E_{M}, respectively, in the subsequent sections.

C. Parallel-Iterative Framework

To effectively model the complex bidirectional interactions between the AV and surrounding agents, we propose a Parallel-Iterative Framework that fundamentally shifts from the conventional sequential paradigm. Instead of a fixed predict-then-plan order, our framework treats prediction and planning as coupled tasks that are updated in parallel within each stage and progressively refined across stages using the previous-stage joint future context.

1) Multi-Stage Iteration: BIPP decomposes the planning horizon t_{f} into a sequence of sub-horizons t^j_{f} $(j = 1, \ldots, N)$. This hierarchical structure enables the modules at stage j to condition on both the trajectories from the previous stage $(j-1)$ and the history context, facilitating bidirectional interaction between prediction and planning. This approach effectively models both the AV and SAs as adaptive agents that dynamically respond to each other's future intentions.

Formally, the multi-stage refinement propagates a joint future context across stages:

$$E^{\text{fut}}_{j} = \text{Concat}\left(E^{\text{plan}}_{j}, E^{\text{pred}}_{j}\right), \quad j = 1, \ldots, N, \tag{3}$$

where E^{plan}_{j} denotes the compact AV planning context produced at stage j, E^{pred}_{j} denotes the compact SA prediction context produced at stage j, and E^{fut}_{j} denotes their concatenated joint future context, which serves as the future context for the following refinement stage when $j < N$.

2) Parallel Execution Architecture: Unlike sequential paradigms that impose a fixed prediction–planning order, BIPP updates prediction and planning in parallel within each sub-horizon. Neither branch takes the current-stage output of the other as its input; instead, both branches operate on the same previous-stage joint future context together with the shared scene context. Thus, there is no directed dependency between prediction and planning within the current stage, and neither branch is assigned a leader role or an information advantage. This parallel information flow reduces the order-induced asymmetry of sequential updates and supports a more consistent estimation of the joint AV–SA future.

Let C denote the shared scene context including historical motion and map information. At stage j, the two branches are updated in parallel:

$$E^{\text{pred}}_{j} = \mathcal{F}_{\text{pred}}\left(E^{\text{fut}}_{j-1}, C\right), \quad E^{\text{plan}}_{j} = \mathcal{F}_{\text{plan}}\left(E^{\text{fut}}_{j-1}, C\right). \tag{4}$$

Thus, there is no current-stage dependency edge $E^{\text{pred}}_{j} \to E^{\text{plan}}_{j}$ or $E^{\text{plan}}_{j} \to E^{\text{pred}}_{j}$.

3) Bidirectional Interaction Modeling: BIPP models reciprocal influence between future prediction and future planning by coupling the two tasks through our Interactive Prediction and Interactive Planning modules. Both modules attend to the shared scene context and the propagated joint future context, allowing prediction and planning to progressively adapt to each other across stages. By combining (3) and (4), the cross-stage bidirectional conditioning for stage $j+1$ can be summarized as

$$\mathcal{F}\big(E_j^{\text{fut}}, C\big) \rightarrow \Big\{E_{j+1}^{\text{plan}}, E_{j+1}^{\text{pred}}\Big\}, \quad j = 1, \ldots, N-1. \tag{5}$$

This means that the next-stage AV planning context and SA prediction context both condition on the future context provided by the previous stage, enabling bidirectional interaction between them.

D. Bidirectional Interaction Modeling Mechanism

To enable explicit and reciprocal influence between the AV and surrounding agents, we design a unified Bidirectional Interaction Modeling Mechanism. This mechanism is shared by both the prediction and planning modules, ensuring that all agents, including the ego vehicle and traffic participants, are modeled as interactive entities that reason about each other's future behaviors. As shown in the parallel-iterative stage of Fig. 2, this mechanism is embedded in both the Interactive Prediction and Interactive Planning branches. Its inputs are explicitly separated into the task query, the historical context E_{his}, the previous-stage future context E_{j-1}^{fut}, and the map context E_{M}.

Formally, let Q^{task} denote the task-specific query input, which can be instantiated as either the prediction query or the planning query and is reused at each interaction stage (details will be explained in following sections). The core of this mechanism follows a four-step process:

1. Bi-Context Conditioning. Historical trajectories record previously observed motion patterns and therefore provide reliable evidence for interaction reasoning. However, history alone cannot fully capture the evolving mutual adaptation that emerges after the AV's future maneuver becomes relevant to surrounding agents. On the other hand, future intention context propagated across stages reflects the model's latest estimate of interactive evolution, but it is still inference-dependent and may contain accumulated errors if used in isolation. To balance reliability and adaptability, BIPP jointly conditions both planning and prediction on historical motion evidence and updated future intention context. This design enables the model to reason about mutual adaptation while remaining grounded in observed behavior, thereby improving interaction modeling in complex urban scenarios.

- *Historical Interaction Stream:* This branch preserves the stable intent prior and kinematic constraints from past behaviors. By attending to the historical state encodings E_{his} of all agents, the model extracts motion patterns and kinematic constraints, providing a foundational prior for trajectory generation.

$$Q_j^{\text{his}} = \text{CrossAttn}(Q^{\text{task}}, E_{\text{his}}, E_{\text{his}}) \tag{6}$$

- *Future Interaction Stream:* This branch captures intention evolution by conditioning on a future context E_{j-1}^{fut} that is iteratively updated from stage $j-1$. This context aggregates the AV plan and the predicted trajectories of all SAs. Through cross-attention (7), the planning and prediction branches attend to the same previous-stage joint future context.

$$Q_j^{\text{fut}} = \text{CrossAttn}(Q^{\text{task}}, E_{j-1}^{\text{fut}}, E_{j-1}^{\text{fut}}) \tag{7}$$

Across stages, each branch reuses its task query, while only the future context E_{j-1}^{fut} is updated. The previous-stage output therefore enters as context rather than replacing the query.

2. Interaction Fusion. We concatenate the outputs of the historical and future interaction streams and project them through a linear layer, as shown in (8). This fusion integrates stable, history-derived kinematic priors with stage-dependent adjustments introduced by future-intention conditioning, yielding an agent-centric feature representation Q_j^{agent}.

$$Q_j^{\text{agent}} = \text{Linear}(\text{Concat}(Q_j^{\text{his}}, Q_j^{\text{fut}})) \tag{8}$$

3. Environmental Contextualization. To ensure physical feasibility and traffic-rule compliance, the fused agent features Q_j^{agent} then attend to the map encoding E_{M} via cross-attention. This step injects road topology, drivable-area boundaries, and traffic-signal states into the interaction-aware representation.

$$Q_j^{\text{ctx}} = \text{CrossAttn}(Q_j^{\text{agent}}, E_{\text{M}}, E_{\text{M}}) \tag{9}$$

4. Temporal Aggregation. To propagate refined intentions to the next stage, we extract the features within the current stage horizon t_{f}^j and apply max pooling. This operation compresses the temporal sequence into a compact context embedding E_j, which serves as the future prior for the subsequent iteration.

$$E_j = \text{MaxPool}(Q_j^{\text{ctx}}[:, : t_{\text{f}}^j, :]) \tag{10}$$

At each stage, the planning and prediction branches condition on the preceding joint future context. This reciprocal update aligns the two future representations across stages; the final planning representation is then decoded to select the AV trajectory with the highest confidence.

E. Interactive Prediction

Interactive Prediction is the prediction instantiation of our Bidirectional Interaction Modeling Mechanism. Applying the same bidirectional interaction update process to prediction queries yields an interaction-aware predictor.

We construct the prediction query Q_{SA} by combining the surrounding-agent history embeddings E_{SA} with temporal (positional) embeddings and K mode-specific query branches, resulting in $Q_{\text{SA}} \in \mathbb{R}^{N_{\text{A}} \times K \times T \times D}$, where T denotes the number of discrete future time steps. At each stage j, the prediction query Q_{SA} is used as Q^{task} in the Bidirectional Interaction Modeling Mechanism. Concretely, we instantiate the mechanism with:

- **Query:** The prediction query Q_{SA} serves as the task query Q^{task}.
- **Historical Context:** The historical encoding E_{his} is composed of both SA and AV embeddings, $E_{\text{his}} = [E_{\text{SA}}, E_{\text{AV}}]$.
- **Future Context:** The future context E_{j-1}^{fut} is the concatenation of the AV's planning context and SA prediction context from the previous stage, denoted as $E_{j-1}^{\text{fut}} = \text{Concat}[E_{j-1}^{\text{plan}}, E_{j-1}^{\text{pred}}]$.

After the four-step interaction update in Section III-D, we obtain the stage-j prediction context E_j^{pred}, which summarizes refined SA intentions conditioned on both history and the evolving joint future context. The resulting E_j^{pred} is then used to form the joint future context for the next stage.

F. Interactive Planning

Planning in dense urban environments requires both strict geometric feasibility and sufficient behavioral adaptability. A purely rule-based planner may be overly rigid in complex interactions, whereas a fully learning-based planner may lack sufficient structural priors for safe closed-loop execution. To preserve geometric feasibility and route consistency, BIPP first generates lateral trajectory candidates using rule-based path construction constrained by the road topology. Conditioned on each candidate path, a learning-based longitudinal planner then predicts speed profiles under the current interaction context, allowing the AV to adapt its progress behavior according to surrounding-agent responses. By combining rule-based lateral generation with learning-based longitudinal control, the proposed planner reduces the search space to feasible candidates while retaining sufficient flexibility for interactive decision making.

(a) Lateral Planning. Our lateral module generates N_{lat} candidate paths $\mathcal{P}_{\text{lat}}$ by sampling terminal points along lane centerlines and fitting third-order polynomial splines from the current state. For a sampled terminal point, the m-th raw path candidate is parameterized by arc length as

$$\mathbf{p}_m(s) = [x_m(s), y_m(s)]^\top, \quad s \in [0, L_m],$$
$$x_m(s) = \sum_{r=0}^{3} a_{m,r} s^r, \quad y_m(s) = \sum_{r=0}^{3} b_{m,r} s^r. \tag{11}$$

where $m \in \{1, \ldots, M\}$ indexes the raw sampled candidates, L_m is the path length, and $a_{m,r}, b_{m,r}$ are spline coefficients determined by the current AV pose and the terminal lane-centerline pose.

Since the sampled endpoints lie on lane centerlines, road-topology consistency is mainly introduced by the sampling process. The raw candidates are then scored by handcrafted rules that penalize collision risk, lane-change distance, and curvature:

$$J_m^{\text{lat}} = w_{\text{col}} J_m^{\text{col}} + w_{\text{lc}} J_m^{\text{lc}} + w_\kappa \max_s |\kappa_m(s)|,$$
$$\mathcal{P}_{\text{lat}} = \text{TopN}_{N_{\text{lat}}}\left(\{\mathbf{p}_m\}_{m=1}^{M}, -J_m^{\text{lat}}\right). \tag{12}$$

Here, J_m^{col} denotes the collision-risk cost, J_m^{lc} denotes the lane-change-distance cost, $\max_s |\kappa_m(s)|$ measures the maximum path curvature, and w_{col}, w_{lc}, and w_κ are non-negative rule weights. $\text{TopN}_{N_{\text{lat}}}(\cdot)$ keeps the N_{lat} candidates with the lowest lateral cost. The selected paths provide compact geometric priors for the subsequent longitudinal planner.

(b) Longitudinal Planning. We employ a learning-based longitudinal planner that adapts to the dynamic environment, instead of relying on predefined speed profiles. To ensure controllability and tracking feasibility, we explicitly condition longitudinal planning on trajectory geometry.

Instead of using reference lines as lateral queries [53], [54], we encode the N_{lat} candidate trajectories (each with n_{p} points) using a PointNet-like [55] architecture to obtain lateral queries $Q_{\text{lat}} \in \mathbb{R}^{N_{\text{lat}} \times D}$. These are combined with N_{lon} learnable longitudinal query vectors $Q_{\text{lon}} \in \mathbb{R}^{N_{\text{lon}} \times D}$ representing multimodal behavior. The concatenated features are linearly projected to form $Q_{\text{init}}^{\text{plan}'} \in \mathbb{R}^{N_{\text{lat}} \times N_{\text{lon}} \times D}$. A learnable temporal embedding $E_{\text{T}} \in \mathbb{R}^{N_{\text{lat}} \times N_{\text{lon}} \times T \times D}$ is then added after broadcasting $Q_{\text{init}}^{\text{plan}'}$ along the temporal dimension:

$$\widetilde{Q}_{\text{init}}^{\text{plan}} = \text{Expand}_T(Q_{\text{init}}^{\text{plan}'}) + E_{\text{T}},$$
$$\widetilde{Q}_{\text{init}}^{\text{plan}} \in \mathbb{R}^{N_{\text{lat}} \times N_{\text{lon}} \times T \times D}. \tag{13}$$

The time-indexed query is then refined by self-attention along the lateral-candidate, longitudinal-mode, and temporal dimensions:

$$Q_{\text{init}}^{\text{plan}} = \text{SA}_{\text{T}}\Big(\text{SA}_{\text{lon}}\big(\text{SA}_{\text{lat}}(\widetilde{Q}_{\text{init}}^{\text{plan}})\big)\Big). \tag{14}$$

where SA_{lat}, SA_{lon}, and SA_{T} denote self-attention after reshaping the tensor along the lateral-candidate, longitudinal-mode, and temporal dimensions, respectively. The final planning query therefore has shape $Q_{\text{init}}^{\text{plan}} \in \mathbb{R}^{N_{\text{lat}} \times N_{\text{lon}} \times T \times D}$, ensuring dimensional compatibility with the prediction module ($N_{\text{lat}} \times N_{\text{lon}} = K$).

At each stage j, we apply the Bidirectional Interaction Modeling Mechanism in Section III-D by using the planning query $Q_{\text{init}}^{\text{plan}}$ as Q^{task} and conditioning it on the shared historical context E_{his} and the propagated future context E_{j-1}^{fut}. This allows the AV plan to be refined in response to both the observed history and the predicted reactions of surrounding agents under the evolving joint future. The resulting planning context E_j^{plan} summarizes the stage-wise planning intent and is concatenated with the prediction context to form the joint future context E_j^{fut} for the next stage.

G. Trajectory Decoder

Upon completing N iterative refinement stages, we obtain the final context embeddings for prediction and planning, denoted as E_N^{pred} and E_N^{plan}, respectively. These embeddings are then decoded to generate the final trajectory outputs.

For the prediction task, the embedding E_N^{pred} is processed by an MLP head to generate multimodal predicted trajectories for each surrounding agent. For the planning task, the query E_N^{plan} is passed through two MLP heads: one to regress the longitudinal motion profile S along the candidate paths, and another to predict the confidence score $\boldsymbol{\pi}$ for each planning mode. The decoding process is formulated as:

$$Y_i = \text{MLP}(E_N^{\text{pred}}) \tag{15}$$
$$S = \text{MLP}(E_N^{\text{plan}}), \quad \boldsymbol{\pi} = \text{MLP}(E_N^{\text{plan}}) \tag{16}$$

Here, $Y_i \in \mathbb{R}^{K \times T \times 2}$ represents the predicted trajectory for agent i ($i \in \{1, \ldots, N_{\text{A}}\}$); $S \in \mathbb{R}^{N_{\text{lat}} \times N_{\text{lon}} \times T}$ denotes the planned longitudinal state profile; T denotes the number of discrete future time steps; and $\boldsymbol{\pi} \in \mathbb{R}^{N_{\text{lat}} \times N_{\text{lon}}}$ is the associated planning confidence score. The complete spatiotemporal trajectory for the r-th longitudinal mode on the l-th lateral path is then obtained by projecting the decoded longitudinal progress onto the corresponding path:

$$\mathbf{Y}_{l,r}(t) = \mathbf{p}_l\big(S_{l,r}(t)\big), \quad l = 1, \dots, N_{\text{lat}}, \quad r = 1, \dots, N_{\text{lon}}. \tag{17}$$

Here, $\mathbf{p}_l$ is the l-th selected lateral path, and $\mathbf{Y}_{l,r}$ is the resulting complete AV trajectory candidate. Finally, this projection yields a total of $N_{\text{lat}} \times N_{\text{lon}}$ spatiotemporal trajectories for the AV.

H. Model Training

1) Planning Loss: To prevent mode collapse during training, we adopt a teacher-forcing approach inspired by [54] with specific improvements. Specifically, we project the endpoint of the ground truth trajectory onto the N_{lat} candidate paths and select the path with the minimum lateral deviation as the target lateral path ℓ. For the longitudinal dimension, to ensure trajectory consistency, the reachable distance range is first estimated according to the vehicle's initial speed and kinematic limits. Given the initial speed v_0, maximum acceleration $a_{\text{pos}}^{\max}$, maximum braking deceleration $a_{\text{neg}}^{\max}$, speed limit v_{lim}, and planning horizon t_{f}, the reachable longitudinal interval is defined as $[s_{\min}, s_{\max}]$. Let the maximum-braking duration within the horizon be

$$t_{\text{stop}} = \min\left(t_{\text{f}}, \frac{v_0}{a_{\text{neg}}^{\max}}\right). \tag{18}$$

The lower reachable bound is then

$$s_{\min} = v_0 t_{\text{stop}} - \frac{1}{2} a_{\text{neg}}^{\max} t_{\text{stop}}^2, \tag{19}$$

whereas the time required to reach the speed limit under maximum acceleration is

$$t_{\text{lim}} = \max\left(0, \frac{v_{\text{lim}} - v_0}{a_{\text{pos}}^{\max}}\right), \tag{20}$$

$$s_{\max} = \begin{cases} v_0 t_{\text{f}} + \frac{1}{2} a_{\text{pos}}^{\max} t_{\text{f}}^2, & t_{\text{lim}} \geq t_{\text{f}}, \\ v_0 t_{\text{lim}} + \frac{1}{2} a_{\text{pos}}^{\max} t_{\text{lim}}^2 + v_{\text{lim}}(t_{\text{f}} - t_{\text{lim}}), & t_{\text{lim}} < t_{\text{f}}. \end{cases} \tag{21}$$

Here, $s_{\min}$ is the distance traveled before either the horizon ends or the vehicle comes to rest under maximum braking. The two cases in $s_{\max}$ correspond, respectively, to accelerating throughout the horizon and to accelerating until the speed limit is reached and then cruising at that limit.

This range is then divided into N_{lon} equal intervals, and the target longitudinal interval index q is determined based on the projected longitudinal distance of the ground truth endpoint. The query corresponding to the target lateral path ℓ and longitudinal interval index q is then designated as the target query.

To make the supervision compatible with the longitudinal-query design, BIPP uses longitudinal distance and velocity losses rather than Cartesian trajectory-point losses, as shown in Eq. (22). To satisfy the comfort requirements, we also include a comfort loss comprising acceleration and jerk terms, as shown in Eq. (23). For the score classification loss, we use cross-entropy to measure the difference between the predicted and target scores, as shown in Eq. (24).

TABLE II

BIPP Model Hyperparameters

Notation	Hyperparameter	Value
D	Hidden dimension	256
H	Attention heads / head dimension	8 / 32
d_{ff}	Transformer FFN dimension	1024
p_{drop}	Attention and FFN dropout	0.1
p_{state}	AV state-dropout probability	0.75
N_{e}	Scene encoder layers	3
N	Interaction stages	3
t_{h}	Historical horizon	2.0 s
t_{f}	Planning horizon	8.0 s
Δt	Sampling interval	0.1 s
$t_{\text{f}}^1, t_{\text{f}}^2, t_{\text{f}}^3$	Stage horizons	3, 5, 8 s
N_{A}	Predicted surrounding agents	10
N_{lat}	Lateral candidates	3
N_{lon}	Longitudinal queries	10
K	Joint future modes	30
$\omega_1, \omega_2, \omega_3$	Planning loss weights	1, 0.1, 1

$$\mathcal{L}_{\text{reg}} = \text{SmoothL1}(S^{\text{gt}}, S^*) + \text{SmoothL1}(v^{\text{gt}}, v^*) \tag{22}$$

$$\mathcal{L}_{\text{com}} = \text{SmoothL1}(a^*) + \text{SmoothL1}(\text{jerk}^*) \tag{23}$$

$$\mathcal{L}_{\text{cls}} = \text{CrossEntropy}(\boldsymbol{\pi}, \boldsymbol{\pi}^*) \tag{24}$$

Here, S^{gt} denotes the ground truth travel distance of the AV, with v representing the first derivatives of S. $\boldsymbol{\pi}^*$ is the one-hot distribution for the planned trajectory derived from S^*. The final planning loss is then computed as:

$$\mathcal{L}_{\text{plan}} = \omega_1 \mathcal{L}_{\text{reg}} + \omega_2 \mathcal{L}_{\text{com}} + \omega_3 \mathcal{L}_{\text{cls}} \tag{25}$$

where ω_1, ω_2, and ω_3 are the weighting factors. To balance different objectives, we set $\omega_1 = 1$, $\omega_2 = 0.1$, and $\omega_3 = 1$, as summarized in Table II. The regression and classification terms are treated as primary planning objectives, while the comfort term is assigned a smaller weight because it acts as a smoothness regularizer.

2) Prediction Loss: As each planned trajectory corresponds to a predicted trajectory, we select the trajectories corresponding to the target supervised planned distance S_0^* as the target predicted trajectories $Y_{1:N_{\text{A}}}^*$. The prediction loss is then computed as the smooth L1 loss between the predicted trajectories and the target predicted trajectories, as shown in Eq. (26).

$$\mathcal{L}_{\text{pred}} = \text{SmoothL1}(Y_{1:N_{\text{A}}}^{\text{gt}}, Y_{1:N_{\text{A}}}^*) \tag{26}$$

The final loss function for the BIPP framework is the weighted sum of the planning loss and prediction loss, as shown in Eq. (27).

$$\mathcal{L}_{\text{BIPP}} = N_{\text{A}} \mathcal{L}_{\text{plan}} + \mathcal{L}_{\text{pred}} \tag{27}$$

where we set the planning weight proportional to agent count to maintain equal importance between the two objectives. All learnable modules are jointly trained end-to-end.

Algorithm 1 Parallel-Iterative Bidirectional Interaction Planning and Prediction (BIPP)

Require: Histories $X_{0:N_\mathrm{A}}$, Map $\mathcal{M}$, Stage horizons $\{t_\mathrm{f}^j\}_{j=1}^N$
Ensure: Optimal AV plan Y_0^* and predicted SA trajectories $Y_{1:N_\mathrm{A}}^*$

$\mathcal{P}_\mathrm{lat} \leftarrow$ LATERALSAMPLING $(X_0, \mathcal{M})$ {Sample N_lat candidate paths}
$E_\mathrm{his}, E_\mathrm{M} \leftarrow$ SCENEENCODING $(X_{0:N_\mathrm{A}}, \mathcal{M})$
$Q^\mathrm{pred}, Q^\mathrm{plan} \leftarrow$ QUERYCONSTRUCT $(E_\mathrm{his}, E_\mathrm{M}, \mathcal{P}_\mathrm{lat})$
for $j = 1$ **to** N **do**
  Parallel Execution:
  $E_j^\mathrm{pred} \leftarrow$ PREDICTION $(Q^\mathrm{pred}, E_{j-1}^\mathrm{fut}, E_\mathrm{his}, E_\mathrm{M})$
  $E_j^\mathrm{plan} \leftarrow$ PLANNING $(Q^\mathrm{plan}, E_{j-1}^\mathrm{fut}, E_\mathrm{his}, E_\mathrm{M})$
  Interaction Update:
  $E_j^\mathrm{fut} \leftarrow \mathrm{Concat}(E_j^\mathrm{plan}, E_j^\mathrm{pred})$
end for
$Y_{1:N_\mathrm{A}}, S, \boldsymbol{\pi} \leftarrow$ TRAJECTORYDECODER $(E_N^\mathrm{pred}, E_N^\mathrm{plan})$
$Y_0 \leftarrow$ SPATIOTEMPORALPROJECTION $(\mathcal{P}_\mathrm{lat}, S)$ {Map 1D profiles to 2D paths}
$k^* \leftarrow \arg\max_{k \in \{1,\ldots,K\}} \pi^{(k)}$
$Y_0^* \leftarrow Y_0[k^*], \quad Y_{1:N_\mathrm{A}}^* \leftarrow Y_{1:N_\mathrm{A}}[k^*]$
return $Y_0^*, Y_{1:N_\mathrm{A}}^*$

I. Planning Process

The overall inference and planning process of the BIPP framework is summarized in Algorithm 1. Given the historical states of the AV and surrounding agents $X_{0:N_\mathrm{A}}$ and the map information $\mathcal{M}$, the framework first samples a set of candidate lateral paths $\mathcal{P}_\mathrm{lat}$ for the AV and encodes the historical motion and map context into E_his and E_M. Based on these scene features and the sampled paths, prediction queries Q^pred and planning queries Q^plan are constructed to initialize the subsequent interaction-aware reasoning process. BIPP then performs prediction and planning in parallel over multiple refinement stages. After the final stage, the decoder outputs the predicted SA trajectories $Y_{1:N_\mathrm{A}}$, the longitudinal motion profiles S, and the planning scores $\boldsymbol{\pi}$. The longitudinal profiles are projected onto the sampled lateral paths to generate complete ego trajectory candidates Y_0, and the final AV plan is selected as the highest-scored candidate,

IV. EXPERIMENTS

A. Experimental Setup

1) NuPlan Dataset: Our model is trained on nuPlan [56], a large-scale dataset comprising 1,282 hours of real-world driving data spanning 75 distinct scenario types. Specifically, we randomly sample 300,000 scenarios, employing an 80/20 split for training and validation. Evaluation is conducted on both the nuPlan and interPlan [57] using closed-loop metrics: the non-reactive score (NR-score) and the reactive score (R-score). The two settings differ in whether the surrounding agents react to the AV's actions (reactive) or not (non-reactive). Both scores range from 0 to 100 (higher is better). For each benchmark, the reported total score is computed as the arithmetic average of the scenario-level scores over all evaluated scenarios. All compared planners use structured agent histories and vectorized map inputs under the same planning-level protocol. To comprehensively assess performance, we evaluate on three complementary benchmarks:

(1) **Normal** [38]: We use the Test-10 split as the *Normal* benchmark. It contains 200 scenarios spanning 10 representative scenario types from the nuPlan planning challenge, covering standard urban driving conditions.
(2) **Hard** [52]: We adopt the *Hard* benchmark proposed in [52], which collects the worst-performing scenarios for the rule-based PDM-Closed [19]. It contains 195 challenging scenarios.
(3) **Long-tail** [57]: We use interPlan as our *Long-tail* benchmark. Built upon the nuPlan dataset, it contains 80 challenging scenarios featuring rare events (*e.g.*, jaywalking pedestrians and construction zones) and is designed to test robustness under uncommon and safety-critical scenarios.

2) Evaluation Metrics: Following the nuPlan closed-loop metric definitions, we categorize the reported metrics into four groups:

(1) **Safety:** *No ego at-fault collisions* (*Collision*) and *Time to collision within bound* (*TTC*), which measure collision avoidance and time-to-collision safety margins.
(2) **Efficiency:** *Ego is making progress* (*Progress*) and *Ego progress along expert route* (*Route ratio*), which measure whether the AV makes sufficient progress and how much route progress is achieved relative to the expert.
(3) **Compliance:** *Drivable area compliance* (*Drivable*), *Driving direction compliance* (*Direction*), and *Speed limit compliance* (*Speed*), which assess adherence to road boundaries, travel direction, and speed limits.
(4) **Comfort:** *Ego is comfortable* (*Comfort*), which checks whether the ego trajectory satisfies comfort-related bounds such as acceleration, yaw rate, and jerk.

3) Baseline: We conduct experiments to compare our BIPP framework with several state-of-the-art methods on the nuPlan benchmark to demonstrate the effectiveness of our method. The baseline methods include:

(1) **Intelligent Driver Model (IDM)** [58]: This is a widely used rule-based model for simulating human-like longitudinal driving behavior in traffic scenarios.
(2) **UrbanDriver** [59]: A learning-based policy that leverages imitation learning to generate realistic urban driving behaviors based on expert demonstrations.
(3) **PDM-Closed** [19]: A rule-based planning method that enhances a classical IDM policy with trajectory proposal sampling, simulation, and scoring to enable robust closed-loop decision-making in complex urban driving scenarios; it placed first in the 2023 nuPlan challenge.
(4) **GameFormer (GF)** [8]: A representative *joint prediction–planning* framework that models multi-agent interactions using transformer-based architectures and applies a nonlinear optimizer for trajectory refinement. **GF w/o refine.** denotes its ablation without the nonlinear optimizer.

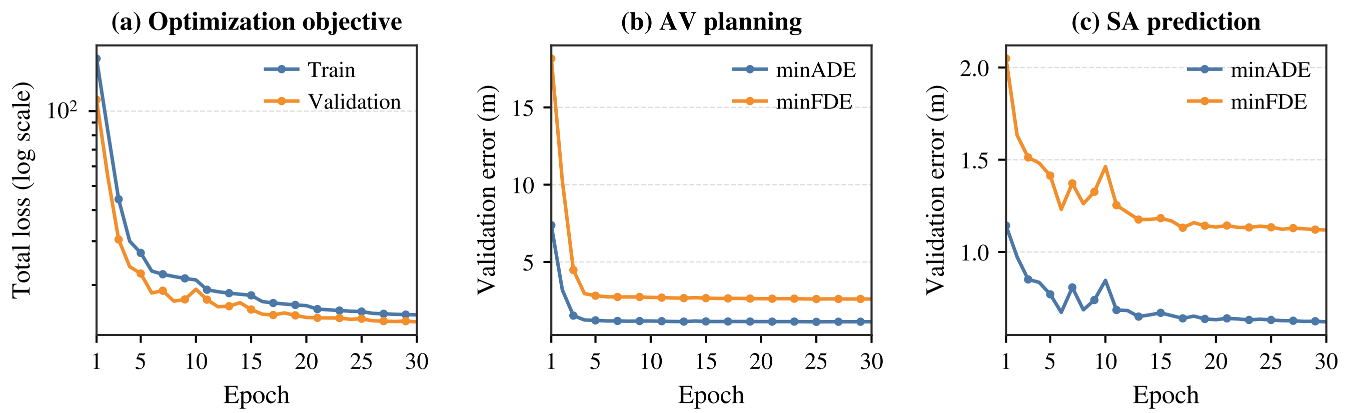


Fig. 3. Learning curves over 30 training epochs: (a) total training and validation loss on a logarithmic scale; (b) validation planning minADE/minFDE; and (c) validation prediction minADE/minFDE.

(5) **DTPP** [38]: A representative *co-leader iterative prediction–planning* framework that alternates ego-plan-conditioned prediction with tree-policy planning, allowing the AV plan to influence surrounding-agent prediction and the predicted agent responses to feed back into ego planning.

B. Implementation Details

We set the planning and historical horizons at 8 seconds and 2 seconds with a time step of 0.1 seconds, consistent with the nuPlan challenge. We set the number of sub-horizons N to 3, with $t_f^1 = 3s, t_f^2 = 5s, t_f^3 = 8s$. The model is trained on an NVIDIA RTX 5090D GPU, with a batch size of 32 over 30 epochs. We use the AdamW optimizer with a weight decay of 1e-2. The learning rate is initialized at 3e-4 and decayed by a factor of 0.5 every 5 epochs. Details of more model parameters can be found in Table II. Project page: https://github.com/zhaox-thu/BIPP.

Figure 3 shows that the training and validation objectives and the validation minADE/minFDE metrics decrease rapidly during the early epochs and then stabilize, indicating stable optimization over the 30-epoch schedule.

V. Main Results

A. Comparison With Sota

1) Overall Performance: Table III summarizes closed-loop results on the *Normal*, *Hard*, and *Long-tail* benchmarks. BIPP achieves the highest overall score among the compared baselines in all three evaluated settings. Notably, its margin over the next-best result widens as the benchmark becomes more challenging: from 1.45/1.67 points on *Normal* (NR/R) to 2.08/1.86 points on *Hard* (NR/R), and further to 12.85 points on *Long-tail*. This trend suggests that BIPP is particularly effective in the evaluated safety-critical and strongly coupled situations, such as dense traffic and high-conflict intersection negotiation, where reciprocal adaptation between the AV and surrounding agents is essential. The results of GF w/o refine. exhibit progressive error accumulation during closed-loop execution.

TABLE III
The Overall Closed-Loop Performance Across Different Benchmarks

Planner	Normal		Hard		Long-tail
	NR	R	NR	R	
UrbanDriver	58.58	53.27	39.93	41.25	9.28
IDM	79.70	81.31	50.71	58.89	31.64
PDM-Closed	92.05	90.05	61.39	67.74	39.07
GameFormer	89.93	85.23	64.46	64.60	11.55
GF w/o refine.	16.44	19.14	12.91	16.63	1.09
DTPP	86.47	83.15	58.34	58.60	26.44
BIPP (Ours)	**93.50**	**91.72**	**66.54**	**69.60**	**51.92**

TABLE IV
Closed-Loop Planning Results on the Normal Benchmark Across Metrics

Planner	Collision	Drivable	TTC	Progress	Route ratio
UrbanDriver	73.00	93.50	66.50	96.50	90.13
IDM	92.25	93.50	87.00	96.00	86.54
PDM-Closed	98.00	99.50	93.50	100.00	87.36
GameFormer	96.75	97.50	91.50	99.50	**93.23**
GF w/o refine.	43.25	37.00	34.00	95.50	65.71
DTPP	95.50	99.50	92.50	98.50	80.02
BIPP (Ours)	**98.50**	**99.75**	**97.00**	**100.00**	91.89

2) Performance Across Metrics: Table IV shows the closed-loop metric breakdown on the Normal benchmark. BIPP achieves the best results on the safety-critical metrics, including *Collision*, *TTC*, and *Drivable*, and also obtains the highest *Progress*. For *Route ratio*, BIPP is slightly lower than GameFormer (91.89 vs. 93.23), but achieves substantially better *Collision*, *Drivable*, and *TTC* scores, indicating a more favorable safety–efficiency balance rather than simply maximizing route progress.

Figure 4 further compares BIPP with PDM-Closed and DTPP on the Hard benchmark. BIPP consistently outperforms both baselines across both safety and efficiency metrics, with the largest advantage observed in *TTC*. This pattern highlights

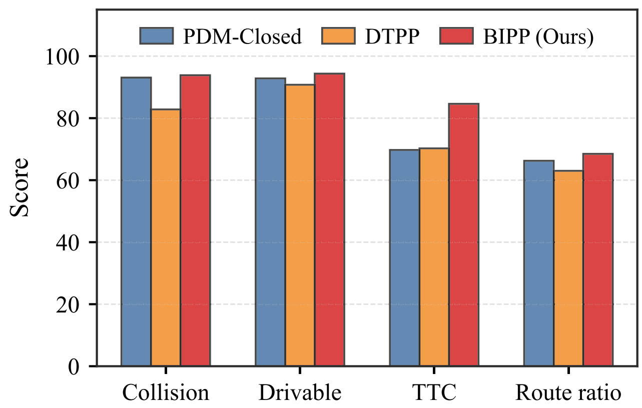


Fig. 4. Safety and efficiency metrics on the Hard benchmark.

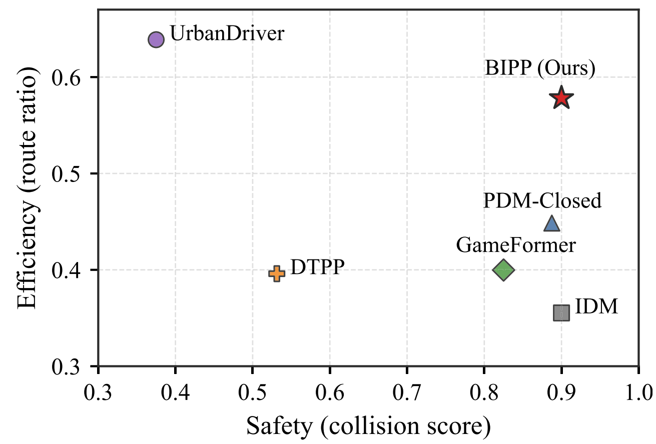


Fig. 6. The safety and efficiency performance on the Long-tail benchmark.

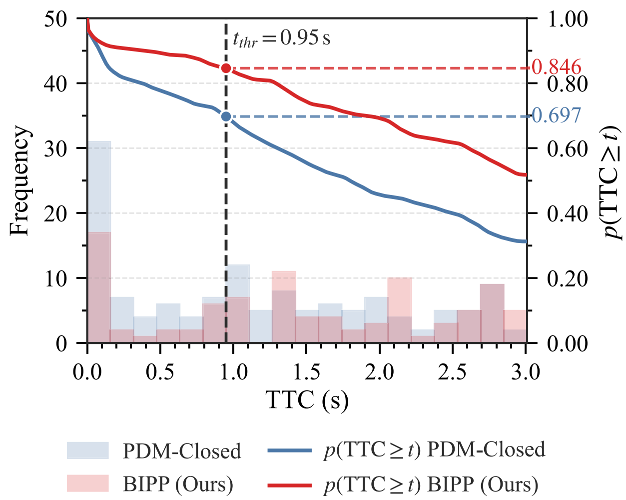


Fig. 5. TTC comparison on the Hard benchmark.

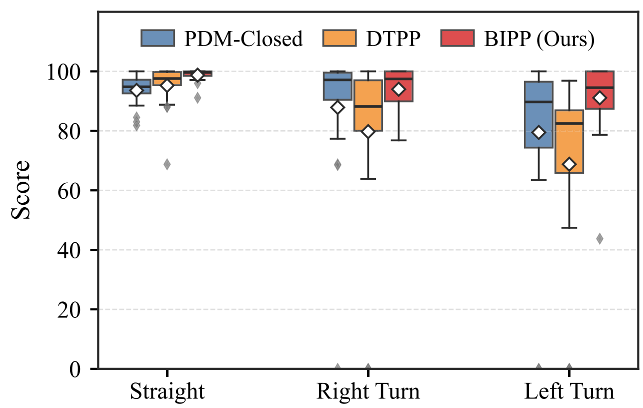


Fig. 7. Performance in intersection scenarios on the Normal benchmark.

the benefit of the proposed bidirectional interaction modeling and hybrid planning design, which enables the planner to maintain safer temporal margins while preserving effective forward progress in highly challenging scenes. Figure 5 provides a more detailed comparison of the *TTC* distributions between BIPP and PDM-Closed on the Hard benchmark. BIPP exhibits substantially fewer *TTC* violations, and its *TTC* survival curve remains consistently above that of PDM-Closed. These results suggest that BIPP maintains a larger safety margin in complex interactive scenarios, thereby reducing the likelihood of risky close-range interactions and collisions.

Figure 6 highlights the safety–efficiency trade-off on the Long-tail benchmark through the joint comparison of *Collision* and *Route ratio*. BIPP achieves the highest *Collision* score while retaining a competitive *Route ratio*, placing it in the most favorable region of this trade-off. Although UrbanDriver attains a marginally higher *Route ratio*, its substantially lower *Collision* score indicates that this gain comes at a considerable cost to safety. These results demonstrate that BIPP achieves a better balance between safe interaction and efficient route progress in long-tail scenarios.

3) Performance Across Scenario Types: Figure 7 compares BIPP, PDM-Closed, and DTPP in intersection scenarios on the Normal benchmark. Across starting-straight, left-turn, and right-turn scenarios, BIPP achieves higher average scores. The advantage is especially pronounced in left-turn and right-turn cases, where the baselines occasionally fail completely and receive zero scores, whereas BIPP exhibits no such failure cases. Moreover, the score distribution of BIPP is more concentrated and skewed toward the high-score region, indicating not only better average performance but also stronger stability and reliability across diverse intersection interactions.

Table V and Figure 8 further compare BIPP with the baselines across scenario types on the Long-tail benchmark. BIPP achieves the highest overall score and the best score in six of the eight categories, with particularly large relative gains in *Construction zone*, *Accident*, and *Overtaking*. Performance remains limited in *Accident*, *Jaywalking pedestrian*, and low-traffic-density lane-change scenarios, indicating that rare agent behaviors and flexible lateral interactions remain challenging.

B. Stage-Wise Refinement and Runtime Analysis

1) Stage-Wise Refinement Analysis: We analyze the intermediate outputs of the same trained three-stage BIPP model

TABLE V
CLOSED-LOOP PLANNING RESULTS ON THE INTERPLAN BENCHMARK ACROSS SCENARIO TYPES

Planner	Score	Constr.	Acc.	Jayw.	Nudge	Overt.	LTD	MTD	HTD
UrbanDriver	9.28	0.00	0.00	20.00	19.50	0.00	15.41	0.00	19.29
IDM	31.64	0.00	0.00	**65.65**	0.00	0.00	62.72	62.07	62.72
PDM-Closed	39.07	18.40	0.00	38.94	58.33	8.24	**62.93**	62.78	62.93
GameFormer	11.55	0.00	0.00	55.67	0.00	0.00	0.00	36.73	0.00
DTPP	26.44	9.02	18.05	31.19	0.00	0.00	26.69	65.17	61.39
BIPP	**51.92**	**60.94**	**33.17**	43.35	**59.57**	**44.86**	42.97	**65.45**	**65.06**

Abbreviations: Constr. = Construction zone; Acc. = Encountering an accident; Jayw. = Jaywalking pedestrian; Nudge = Nudging around a parked vehicle; Overt. = Overtaking a parked vehicle through the oncoming lane; LTD/MTD/HTD = Lane changes to low/medium/high traffic density.

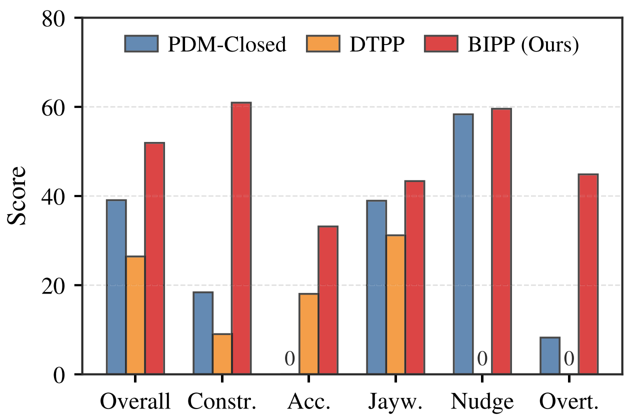


Fig. 8. Performance across different scenario types on the Long-tail benchmark.

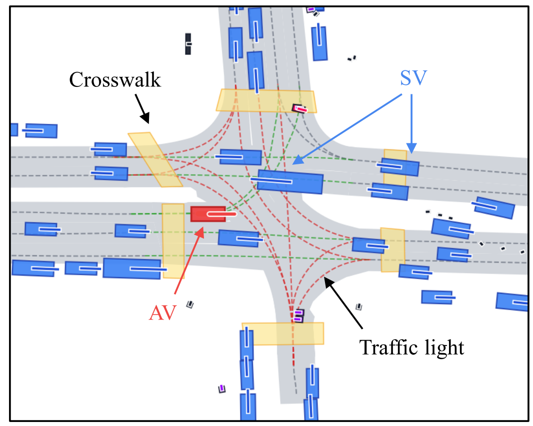


Fig. 9. Illustration of the unprotected left-turn scenario in dense traffic.

TABLE VI
STAGE-WISE REFINEMENT OF BIPP OVER THE COMMON FIRST-3-S EVALUATION WINDOW

Metric	Stage 1	Stage 2	Stage 3
Collision Rate (%) ↓	7.13	6.86	**2.45**
Closest Gap (m) ↑	3.483	3.652	**3.841**
Normalized Entropy ↓	0.648	0.567	**0.431**
Top-1 Confidence Margin ↑	0.174	0.207	**0.299**

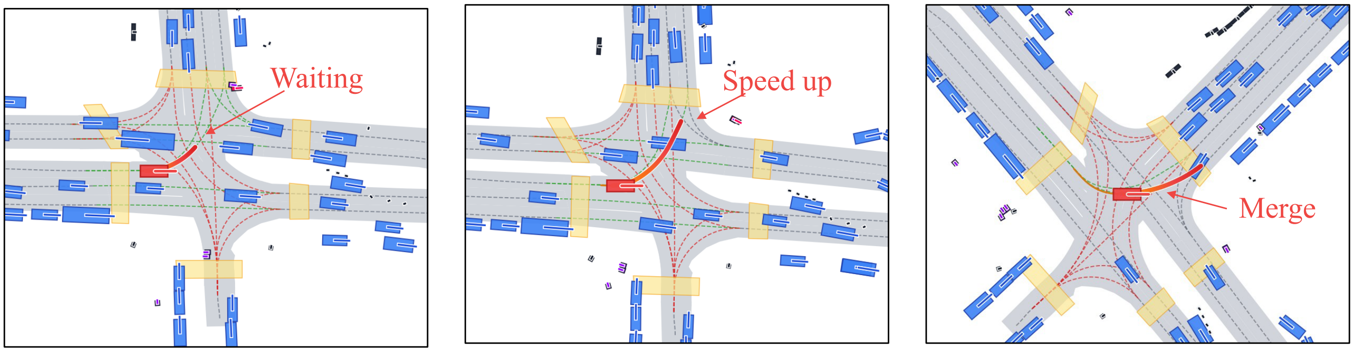
TABLE VII
RUNTIME BREAKDOWN OF BIPP INFERENCE

Module	Runtime (ms)	Ratio (%)
Trajectory sampling	38.9	41.0
Scene encoding	4.9	5.1
Stage 1 interaction	17.5	18.4
Stage 2 interaction	16.8	17.7
Stage 3 interaction	16.9	17.8
Total	**95.0**	**100.0**

using two geometric-safety metrics (Collision Rate and Closest Gap) and two decision-certainty metrics (Normalized Entropy and Top-1 Confidence Margin). Although the three stages internally cover 3, 5, and 8 s, respectively, all stage outputs are uniformly truncated to the first 3 s for this analysis so that the comparison uses an identical temporal window. Collision Rate is the proportion of selected joint proposals with any AV–SA oriented-box overlap, whereas Closest Gap is the average minimum signed gap between the AV and SA boxes over this common window. Normalized Entropy measures the ambiguity of the probability distribution over the 30 joint planning candidates, while Top-1 Confidence Margin is the probability difference between the highest- and second-highest-scored candidates. Therefore, lower Collision Rate and Normalized Entropy, together with higher Closest Gap and Top-1 Confidence Margin, indicate safer proposals and more decisive candidate selection.

As shown in Table VI, Collision Rate decreases from 7.13% to 6.86% and 2.45%, while Closest Gap increases from 3.483 to 3.652 and 3.841 m. Meanwhile, Normalized Entropy decreases from 0.648 to 0.567 and 0.431, and Top-1 Confidence Margin increases from 0.174 to 0.207 and 0.299. Thus, the selected joint proposals become progressively safer and candidate selection becomes more certain, providing empirical evidence of stage-wise refinement.

2) Runtime Analysis: To characterize the computational cost of BIPP, we profile the inference latency of the complete three-stage Python implementation on a workstation equipped with an Intel Core i9-14900K CPU and an NVIDIA RTX 5090D GPU. The runtime is decomposed into trajectory sampling, scene encoding, and the three parallel-iterative interaction stages, as shown in Table VII. BIPP requires



(a) BIPP, $t = 1.5$s (b) BIPP, $t = 5.8$s (c) BIPP, $t = 9.7$s

Fig. 10. Qualitative results of the unprotected left-turn maneuver.

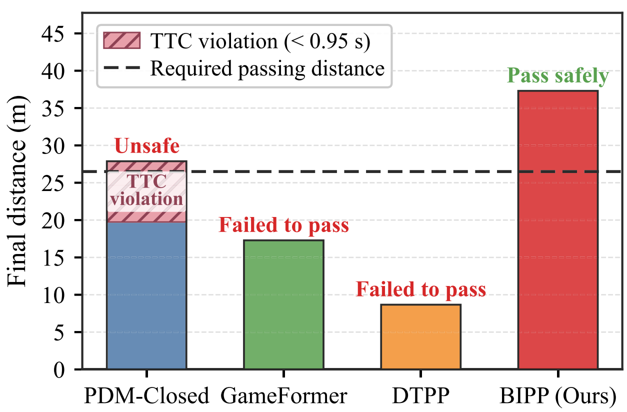


Fig. 11. Comparison of traveled distance and safety performance in the unprotected left-turn scenario.

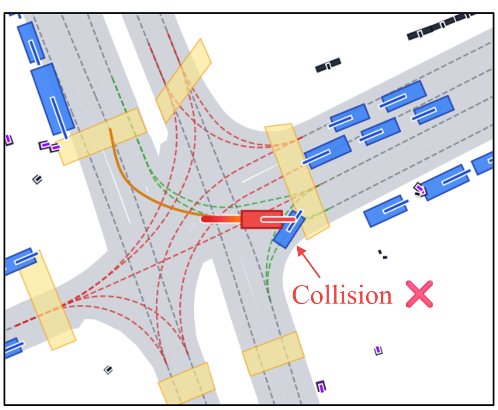


Fig. 12. Illustration of the collision accident of PDM-Closed.

95.0 ms per frame on average, corresponding to approximately 10.5 FPS on this workstation.

As shown in Table VII, the similar per-stage costs indicate that interaction latency grows approximately linearly with the number of stages. Since trajectory sampling is the largest individual runtime component, future work will prioritize improving its efficiency.

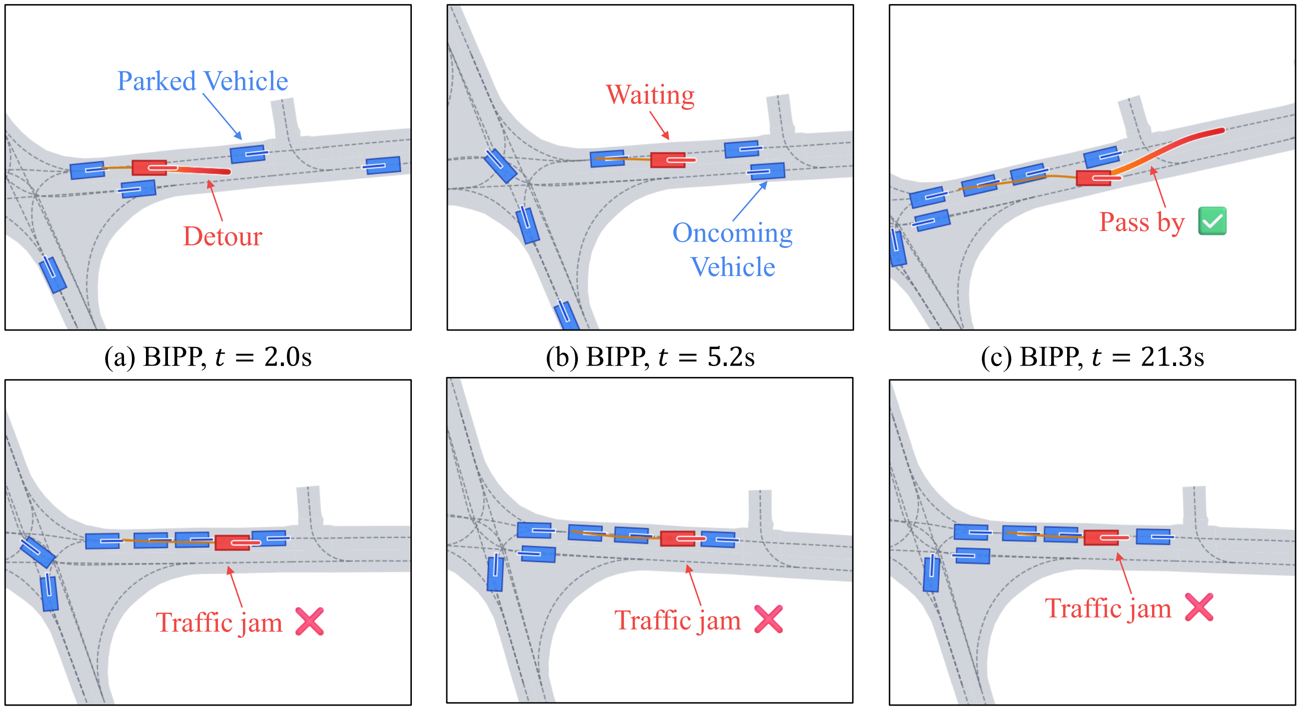
C. Case Study

1) Unprotected Left Turn in Dense Traffic: In a dense intersection scenario shown in Fig. 9, the ego vehicle (red) intends to execute an unprotected left turn while multiple vehicles in the opposing lane are moving straight through the intersection. BIPP first plans a low-speed yielding trajectory to let the oncoming vehicles pass (Fig. 10(a)). As the conflicting traffic gradually clears, the AV smoothly accelerates to enter the intersection. During the turning maneuver, it further encounters a right-turning vehicle, with which it interacts safely and efficiently, and eventually completes the merge into the target through lane.

Fig. 11 quantitatively compares BIPP with the baseline methods in the same scenario. BIPP completes the intersection passage with a final travel distance of 37.31 m and no TTC violation. PDM-Closed reaches 27.88 m but incurs a TTC violation and collides with the right-turning vehicle, thereby failing to complete the maneuver safely, as shown in Fig. 12. GameFormer and DTPP record no TTC violation but reach only 17.29 m and 8.67 m, respectively, and do not complete the intersection passage. Taken together, these results show that BIPP achieves the greatest progress among the compared methods without triggering a TTC violation in its rollout.

2) Overtaking a Parked Vehicle: In the long-tail scenario where a parked vehicle obstructs the middle of the road, Fig. 13(a)-(c) illustrates that BIPP successfully performs a safe and efficient detour maneuver. At $t = 2.0$ s, the AV plans a rightward lane-change trajectory to circumvent the parked vehicle. As the maneuver unfolds, the AV further recognizes an oncoming vehicle in the opposite lane and correspondingly slows down to yield at $t = 5.2$ s. After the oncoming vehicle passes, the AV accelerates again and completes the maneuver safely. This behavior demonstrates that BIPP can not only generate a feasible bypassing trajectory, but also adaptively adjust its motion according to evolving interactive traffic conditions.

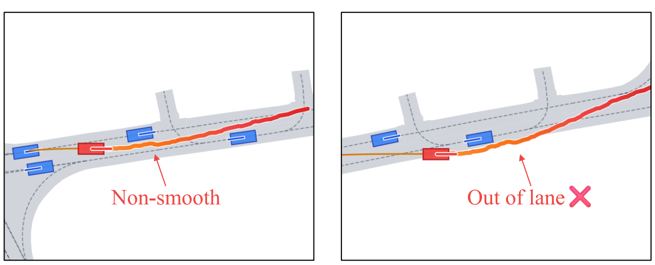
By comparison, PDM-Closed, DTPP and GameFormer all adopt a conservative stop-and-wait behavior that fails to resolve the blockage and eventually induces traffic jam, as shown in Fig. 13(d)-(e). Figure 14 further analyzes the planning process of the purely learning-based GameFormer w/o refine. Although it is able to generate a nominal bypassing



(a) BIPP, $t = 2.0$s (b) BIPP, $t = 5.2$s (c) BIPP, $t = 21.3$s

(d) PDM-Closed, $t = 21.3$s (e) DTPP, $t = 21.3$s (f) Gameformer, $t = 21.3$s

Fig. 13. Qualitative results of the overtaking scenario.



(a) GF w/o refine, $t = 2.6$s (b) GF w/o refine, $t = 6.1$s

Fig. 14. Qualitative results illustrating the closed-loop error accumulation of GameFormer in the overtaking scenario.

TABLE VIII
INFLUENCE OF ORDER OF PLANNING AND PREDICTION

Order	NR-Score	TTC	Route ratio	Comfort	interPlan
Pre-Plan	92.36	95.00	88.06	89.50	44.39
Plan-Pre	91.57	94.50	89.70	87.50	39.52
Parallel	**93.50**	**97.00**	**91.89**	**89.50**	**51.92**

trajectory, the low trajectory quality leads to persistent execution deviations and error accumulation, eventually resulting in an out-of-lane failure.

D. Ablation Study

1) Influence of Execution Order: Table VIII compares the three dependency structures under otherwise identical settings. The *Parallel* strategy achieves the highest NR-Score and interPlan score, consistently outperforming both *Pre-then-Plan* and *Plan-then-Pre*.

TABLE IX
MODULE-LEVEL ABLATION OF INTERACTION COMPONENTS

Variant	NR-Score	TTC	Route ratio	Comfort	interPlan
w/o future	89.66	97.00	83.69	93.50	34.61
w/o history	92.79	96.50	86.46	**95.00**	45.99
w/o map	87.22	**97.50**	79.30	93.50	39.62
BIPP (Ours)	**93.50**	97.00	**91.89**	89.50	**51.92**

Compared with *Plan-then-Pre*, *Pre-then-Plan* achieves a slightly higher *TTC* (95.00 vs. 94.50) but a lower *Route ratio* (88.06 vs. 89.70), showing that sequential architectures are sensitive to which branch is executed first. *Parallel* eliminates this within-stage order sensitivity by removing the directed dependency between prediction and planning: both branches are updated from the same previous-stage joint future, and neither receives the current-stage output of the other. Their updated representations are then combined and propagated to the next stage for reciprocal refinement. This order-insensitive structure improves consistency between the AV plan and predicted surrounding-agent responses, increasing both *TTC* to 97.00 and *Route ratio* to 91.89.

2) Influence of Interaction Components: Table IX reports controlled module-level ablations under otherwise identical settings. All three ablations reduce both NR-Score and interPlan, with more pronounced degradation on the interaction-focused interPlan benchmark. In particular,

TABLE X
INFLUENCE OF NUMBER OF ITERATION STAGES

Iteration	NR-Score	TTC	Route ratio	Comfort	interPlan
One stage	91.78	94.00	92.65	85.50	40.42
Two stages	92.68	95.00	90.87	83.00	47.70
Three stages	**93.50**	**97.00**	91.89	**89.50**	**51.92**
Four stages	93.06	95.00	**93.64**	80.05	50.99

TABLE XI
INFLUENCE OF NUMBER OF LONGITUDINAL QUERIES

Number	NR-Score	TTC	Route ratio	Comfort	interPlan
5	92.07	94.50	**92.31**	87.00	43.71
10	**93.50**	**97.00**	91.89	**89.50**	**51.92**
15	92.77	95.50	91.50	87.00	42.38

removing the future interaction stream decreases interPlan by 17.31 points, demonstrating the importance of cross-stage joint-future conditioning. The consistent degradation caused by removing the historical interaction stream or map cross-attention further shows that observed motion evidence and environmental context provide complementary grounding for interaction reasoning. Although some ablated variants obtain marginally higher *TTC* or *Comfort*, these gains are accompanied by substantial losses in route progress and overall closed-loop performance.

3) Influence of Multi-Stage Interaction: As shown in Table X, increasing the number of stages from one to three progressively improves overall closed-loop performance. The improvement is particularly evident on interPlan, indicating that repeated updates of the predicted and planned intentions help the model progressively refine its understanding of coupled AV–SA interactions, even though all variants cover the same total planning horizon. The four-stage variant remains competitive and achieves the highest *Route ratio*, but it does not provide a further overall improvement. Moreover, the runtime breakdown in Table VII shows that each interaction stage has a similar computational cost of approximately 17 ms, such that adding stages increases interaction latency approximately linearly. Therefore, three stages retain the benefit of iterative interaction refinement while providing a better balance between closed-loop performance and computational efficiency.

4) Influence of Number of Longitudinal Queries: Table XI analyzes the impact of longitudinal query granularity on multimodal decision coverage. The model achieves optimal overall performance with 10 longitudinal queries: its NR-Score, *TTC*, *Comfort*, and interPlan score are 93.50, 97.00, 89.50, and 51.92, respectively, compared with 92.07, 94.50, 87.00, and 43.71 for 5 queries and 92.77, 95.50, 87.00, and 42.38 for 15 queries. Fewer queries limit the diversity of speed profiles, whereas simply increasing the number of queries does not provide a monotonic benefit, suggesting that the additional coverage is offset by more difficult mode learning and selection. Ten queries therefore provide the best empirical balance between multimodal behavioral coverage and decision discrimination.

VI. CONCLUSION

In this paper, we proposed BIPP, a bidirectional interactive prediction-and-planning framework to address insufficient interaction modeling and closed-loop error accumulation in autonomous driving. The main conclusions are summarized as follows:

(1) Parallel-iterative prediction and planning: BIPP performs prediction and planning in parallel and refines them iteratively across stages, removing the artificial unidirectional dependency in conventional pipelines. Compared with *Pre-then-Plan*, *Plan-then-Pre*, and representative planners including PDM-Closed, DTPP, and GameFormer, BIPP achieves the best overall closed-loop performance. It also attains superior *TTC* and competitive *Route ratio*, indicating a better balance between safety and efficiency.

(2) Bidirectional interaction modeling: By jointly leveraging reliable historical states and updated future intentions, BIPP progressively refines the joint AV–SA future context. Removing the future interaction stream decreases interPlan from 51.92 to 34.61, while the stage-wise analysis shows consistent improvements in geometric safety and decision certainty. These results demonstrate the contribution of bidirectional interaction to reciprocal reasoning and adaptive decision selection.

(3) Hybrid planning strategy: By combining rule-based lateral trajectory generation with learning-based longitudinal decision-making, the hybrid planner constructs a compact and geometrically feasible candidate space while retaining behavioral adaptability. This planning formulation complements the interaction mechanism by improving trajectory feasibility and closed-loop execution performance.

Overall, extensive closed-loop evaluations on the nuPlan and interPlan benchmarks demonstrate that the complementary combination of interaction-aware refinement and hybrid trajectory generation enables BIPP to achieve the best overall performance among the compared baselines in the evaluated settings. The uneven category-level results also identify remaining challenges in rare agent behaviors and flexible lateral interactions.

The rule-based lateral planner remains limited in highly constrained spaces. Future work will investigate more flexible lateral trajectory generation and extend BIPP toward an end-to-end perception-to-planning framework that jointly models upstream perception uncertainty and downstream planning decisions. We will also assess its potential for real-vehicle deployment through embedded-platform optimization and on-road testing.

REFERENCES

[1] S. Sacone, "Platoon control in traffic networks: New challenges and opportunities," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 4, pp. 149–183, Apr. 2024.
[2] C. Pasquale, S. Sacone, S. Siri, and M. Papageorgiou, "Optimal control for reducing congestion and improving safety in freeway systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3613–3625, Nov. 2018.

[3] M. Zhu, Y. Wang, Z. Pu, J. Hu, X. Wang, and R. Ke, "Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102662.

[4] K. Muhammad, A. Ullah, J. Lloret, J. D. Ser, and V. H. C. de Albuquerque, "Deep learning for safe autonomous driving: Current challenges and future directions," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4316–4336, Jul. 2021.

[5] P. Cai, H. Wang, Y. Sun, and M. Liu, "DQ-GAT: Towards safe and efficient autonomous driving with deep Q-learning and graph attention networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21102–21112, Nov. 2022.

[6] H. Tong, L. Chu, Z. Chen, Y. Liu, Y. Zhang, and J. Hu, "Multi-objective autonomous eco-driving strategy: A pathway to future green mobility," *Green Energy Intell. Transp.*, vol. 4, no. 4, Aug. 2025, Art. no. 100279.

[7] W. Luo, C. Park, A. Cornman, B. Sapp, and D. Anguelov, "JFP: Joint future prediction with interactive multi-agent modeling for autonomous driving," in *Proc. Conf. Robot Learn.*, 2023, pp. 1457–1467.

[8] Z. Huang, H. Liu, and C. Lv, "GameFormer: Game-theoretic modeling and learning of transformer-based interactive prediction and planning for autonomous driving," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 3880–3890.

[9] A. Irshayyid, J. Chen, and G. Xiong, "A review on reinforcement learning-based highway autonomous vehicle control," *Green Energy Intell. Transp.*, vol. 3, no. 4, Aug. 2024, Art. no. 100156.

[10] Y. Hu et al., "Planning-oriented autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 17853–17862.

[11] B. Jiang et al., "VAD: Vectorized scene representation for efficient autonomous driving," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 8306–8316.

[12] A. Sadat, S. Casas, M. Ren, X. Wu, P. Dhawan, and R. Urtasun, "Perceive, predict, and plan: Safe motion planning through interpretable semantic representations," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 414–430.

[13] S. Hagedorn, M. Hallgarten, M. Stoll, and A. Condurache, "Rethinking the integration of prediction and planning in deep learning-based automated driving systems: A review," 2024, *arXiv:2308.05731*.

[14] Z. Song et al., "Don't shake the wheel: Momentum-aware planning in end-to-end autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2025, pp. 22432–22441.

[15] B. Zhang, N. Song, X. Jin, and L. Zhang, "Bridging past and future: End-to-end autonomous driving with historical prediction and planning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 6854–6863.

[16] M. Knoche, D. de Geus, and B. Leibe, "DONUT: A decoder-only model for trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2025, pp. 28903–28912.

[17] W. Sun, X. Lin, Y. Shi, C. Zhang, H. Wu, and S. Zheng, "SparseDrive: End-to-end autonomous driving via sparse scene representation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2025, pp. 8795–8801.

[18] Z. Lan, Y. Jiang, Y. Mu, C. Chen, and S. E. Li, "Sep.: Towards efficient scene representation learning for motion prediction," 2023, *arXiv:2309.15289*.

[19] D. Dauner, M. Hallgarten, A. Geiger, and K. Chitta, "Parting with misconceptions about learning-based vehicle motion planning," in *Proc. Conf. Robot Learn.*, 2023, pp. 1268–1281.

[20] Z. Zhang et al., "Closed-loop supervised fine-tuning of tokenized traffic models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 5422–5432.

[21] Z. Li, S. Wang, S. Lan, Z. Yu, Z. Wu, and J. M. Alvarez, "Hydra-NeXt: Robust closed-loop driving with open-loop training," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2025, pp. 27305–27314.

[22] X. Weng, B. Ivanovic, Y. Wang, Y. Wang, and M. Pavone, "PARA-drive: Parallelized architecture for real-time autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 15449–15458.

[23] A. Cui, S. Casas, A. Sadat, R. Liao, and R. Urtasun, "LookOut: Diverse multi-future prediction and planning for self-driving," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16107–16116.

[24] N. Rhinehart, R. Mcallister, K. Kitani, and S. Levine, "PRECOG: Prediction conditioned on goals in visual multi-agent settings," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2821–2830.

[25] H. Song, W. Ding, Y. Chen, S. Shen, M. Y. Wang, and Q. Chen, "PiP: Planning-informed trajectory prediction for autonomous driving," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 598–614.

[26] Z. Huang, H. Liu, J. Wu, and C. Lv, "Conditional predictive behavior planning with inverse reinforcement learning for human-like autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7244–7258, Jul. 2023.

[27] E. Ohn-Bar, A. Prakash, A. Behl, K. Chitta, and A. Geiger, "Learning situational driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 11296–11305.

[28] E. Amirloo, A. Rasouli, P. Lakner, M. Rohani, and J. Luo, "LatentFormer: Multi-agent transformer-based interaction modeling and trajectory prediction," 2022, *arXiv:2203.01880*.

[29] W. Zheng, R. Song, X. Guo, C. Zhang, and L. Chen, "GenAD: Generative end-to-end autonomous driving," in *Proc. Eur. Conf. Comput. Vis. Cham*, Switzerland: Springer, 2024, pp. 87–104.

[30] X. Chen, J. Yan, W. Liao, T. He, and P. Peng, "Int2Planner: An intention-based multi-modal motion planner for integrated prediction and planning," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, no. 14, pp. 14558–14566.

[31] J. Fu, Y. Shen, Z. Jian, S. Chen, J. Xin, and N. Zheng, "InteractionNet: Joint planning and prediction for autonomous driving with transformers," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2023, pp. 9332–9339.

[32] S. Pini, C. S. Perone, A. Ahuja, A. S. R. Ferreira, M. Niendorf, and S. Zagoruyko, "Safe real-world autonomous driving by learning to predict and plan with a mixture of experts," 2022, *arXiv:2211.02131*.

[33] X. Jia et al., "Think twice before driving: Towards scalable decoders for end-to-end autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, Jun. 2023, pp. 21983–21994.

[34] Y. Chen, P. Karkus, B. Ivanovic, X. Weng, and M. Pavone, "Tree-structured policy planning with learned behavior models," 2023, *arXiv:2301.11902*.

[35] R. Chekroun, T. Gilles, M. Toromanoff, S. Hornauer, and F. Moutarde, "MBAPPE: MCTS-built-around prediction for planning explicitly," in *Proc. IEEE Intell. Vehicles Symp.* (IV), Jun. 2024, pp. 2062–2069.

[36] B. Zhang, N. Song, J. Li, X. Zhu, J. Deng, and L. Zhang, "Future-aware end-to-end driving: Bidirectional modeling of trajectory planning and scene evolution," 2025, *arXiv:2510.11092*.

[37] Z. Chen, M. Ye, S. Xu, T. Cao, and Q. Chen, "PPAD: Iterative interactions of prediction and planning for end-to-end autonomous driving," 2023, *arXiv:2311.08100*.

[38] Z. Huang, P. Karkus, B. Ivanovic, Y. Chen, M. Pavone, and C. Lv, "DTPP: Differentiable joint conditional prediction and cost evaluation for tree policy planning in autonomous driving," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 6806–6812.

[39] B. Paden, M. Cáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Trans. Intell. Vehicles*, vol. 1, no. 1, pp. 33–55, Mar. 2016.

[40] W. Zhan, J. Chen, C.-Y. Chan, C. Liu, and M. Tomizuka, "Spatially-partitioned environmental representation and planning architecture for on-road autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 632–639.

[41] J. Chen, W. Zhan, and M. Tomizuka, "Autonomous driving motion planning with constrained iterative LQR," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 2, pp. 244–254, Jun. 2019.

[42] O. Speidel, J. Ruof, and K. Dietmayer, "Graph-based motion planning for automated vehicles using multi-model branching and admissible heuristics," in *Proc. IEEE Int. Conf. Auto. Syst. (ICAS)*, Aug. 2021, pp. 1–5.

[43] S. Teng et al., "Motion planning for autonomous driving: The state of the art and future perspectives," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 6, pp. 3692–3711, Jun. 2023.

[44] H. Shao et al., "LMDrive: Closed-loop end-to-end driving with large language models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 15120–15130.

[45] Y. Wang, J. He, L. Fan, H. Li, Y. Chen, and Z. Zhang, "Driving into the future: Multiview visual forecasting and planning with world model for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 14749–14759.

[46] X. Jia, J. You, Z. Zhang, and J. Yan, "DriveTransformer: Unified transformer for scalable end-to-end autonomous driving," 2025, *arXiv:2503.07656*.

[47] X. Jia, Z. Yang, Q. Li, Z. Zhang, and J. Yan, "Bench2Drive: Towards multi-ability benchmarking of closed-loop end-to-end autonomous driving," in *Proc. Adv. Neural Inf. Process. Syst.*, 2024, pp. 819–844.

[48] L. Zhang, Z. Peng, Q. Li, and B. Zhou, "CAT: Closed-loop adversarial training for safe end-to-end driving," in *Proc. Conf. Robot Learn.*, 2023, pp. 2357–2372.

[49] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 627–635.

[50] M. Bansal, A. Krizhevsky, and A. Ogale, "ChauffeurNet: Learning to drive by imitating the best and synthesizing the worst," 2018, *arXiv:1812.03079*.

[51] J. Zhou et al., "Exploring imitation learning for autonomous driving with feedback synthesizer and differentiable rasterization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Prague, Czech Republic, Sep. 2021, pp. 1450–1457.

[52] J. Cheng, Y. Chen, X. Mei, B. Yang, B. Li, and M. Liu, "Rethinking imitation-based planners for autonomous driving," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 14123–14130.

[53] J. Cheng, R. Xin, S. Wang, and M. Liu, "MPNP: Multi-policy neural planner for urban driving," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 10549–10554.

[54] J. Cheng, Y. Chen, and Q. Chen, "PLUTO: Pushing the limit of imitation learning-based planning for autonomous driving," 2024, *arXiv:2404.14327*.

[55] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.

[56] N. Karnchanachari et al., "Towards learning-based planning: The nuPlan benchmark for real-world autonomous driving," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 629–636.

[57] M. Hallgarten, J. Zapata, M. Stoll, K. Renz, and A. Zell, "Can vehicle motion planning generalize to realistic long-tail scenarios?" in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2024, pp. 5388–5395.

[58] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, pp. 1805–1824, Aug. 2000.

[59] O. Scheel, L. Bergamini, M. Wolczyk, B. Osiński, and P. Ondruska, "Urban driver: Learning to drive from real-world demonstrations using policy gradients," in *Proc. Conf. Robot Learn.*, 2022, pp. 718–728.

Xiang Zhao received the B.E. degree from Tsinghua University, Beijing, China, in 2023, where he is currently pursuing the Ph.D. degree in mechanical engineering with the School of Vehicle and Mobility. His research interests include connected and automated vehicles and trajectory planning and prediction.

Keqiang Li received the B.E. degree from Tsinghua University of China, Beijing, China, in 1985, and the M.S. and Ph.D. degrees in mechanical engineering from Chongqing University, Chongqing, China, in 1988 and 1995, respectively. He is currently a Professor with the School of Vehicle and Mobility, Tsinghua University. He has authored more than 200 articles and more than 80 patents in China and Japan. His main research interests include automotive control systems, driver assistance systems, and networked dynamics and control.

Jianqiang Wang received the B.Tech., M.S., and Ph.D. degrees from Jilin University, Changchun, China, in 1994, 1997, and 2002, respectively. He is currently a Professor with the School of Vehicle and Mobility, Tsinghua University, Beijing, China. He has authored more than 150 articles and a co-inventor of 99 patent applications. He was involved in over ten sponsored projects. His research interests include intelligent vehicles, driving assistance systems, and driver behavior.

Yunhao Hu received the B.E. degree from Harbin Institute of Technology, Harbin, China, in 2018, and the M.Sc. degree from KTH Royal Institute of Technology, Sweden, in 2020. He is currently pursuing the Ph.D. degree in mechanical engineering with the School of Vehicle and Mobility, Tsinghua University. His research interests include connected and automated vehicles and verification and validation.

Hanxiao Ren received the B.E. degree from Tsinghua University, Beijing, China, in 2023, where he is currently pursuing the Ph.D. degree in mechanical engineering with the School of Vehicle and Mobility. His research interests include connected and automated vehicles and cooperative perception and prediction.

Yugong Luo received the B.E. and M.S. degrees from Chongqing University, Chongqing, China, in 1996 and 1999, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2003. He is currently a Professor with the School of Vehicle and Mobility, Tsinghua University. His research interests include intelligent connected electric vehicle dynamics and control.