WMKD: PLANNING-ORIENTED WORLD MODEL VIA FUTURE-GUIDED KNOWLEDGE DISTILLATION FOR END-TO-END AUTONOMOUS DRIVING

Anonymous authors
Paper under double-blind review

ABSTRACT

End-to-end autonomous driving requires anticipating how the surrounding scene will evolve to make safe and effective planning decisions. World models provide this context by learning to predict future scene evolution, yet a fundamental question remains: *"What should a world model predict for planning?"* Existing approaches largely formulate world modeling as reconstructing future observations or scene representations, often without explicitly modeling how these targets should be constructed to support the current planning decision. To this end, we present World Model Knowledge Distillation (WMKD), a teacher–student framework in which a future-guided teacher constructs a planning-oriented prediction target and a student learns to predict it from sensor observations. The teacher encodes ground-truth future semantic evolution under joint supervision from future semantic reconstruction and ego planning, and quantizes the resulting scene representation into discrete codes. The student's Discrete World Model (DWM) predicts these codes to anticipate planning-relevant future knowledge without privileged inputs at inference. To incorporate the predicted representation into planning, the Residual Fusion Module (RFM) fuses it with the observed BEV and their residual, while Planning Response Distillation (PRD) transfers the teacher's planning responses. The Trajectory Repropagation Planner (TRP) further re-encodes refined trajectory geometry to guide successive refinement. Extensive experiments across NAVSIM and Bench2Drive validate WMKD, achieving 92.0 PDMS and 90.6 EPDMS on NAVSIM, as well as a Driving Score (DS) of 87.56 and a Success Rate (SR) of 72.27% on Bench2Drive.

1 INTRODUCTION

End-to-end (E2E) autonomous driving replaces separately optimized perception, prediction, and planning modules (Zhou & Tuzel, 2018; Liu et al., 2023; Li et al., 2024c; Zhou et al., 2023; Moon et al., 2024; Yin et al., 2021; Yan et al., 2018; Ngiam et al., 2021; Jang et al., 2024) with unified policies trained directly from sensor observations (Zeng et al., 2019; Hu et al., 2023b; Jiang et al., 2023; Weng et al., 2024; Sun et al., 2025; Tang et al., 2025; Li et al., 2024d; Xiao et al., 2023). This formulation reduces error propagation across modules and allows scene representations to be optimized directly for driving. Expert trajectory imitation remains a central training signal for E2E driving, yet trajectory supervision alone provides limited information about how surrounding agents and scene structure evolve over the planning horizon. The current ego decision may depend on this future evolution, which is particularly important in interactive traffic scenarios (Zhou et al., 2025b; Hwang et al., 2024; Jiang et al., 2024; Mao et al., 2023; Wang et al., 2024a; Fu et al., 2025; Ding et al., 2024). To capture such future scene evolution, recent methods employ world models that predict future latent features (Li & Cui, 2024; Li et al., 2024a; Zheng et al., 2025; Yang et al., 2026), semantic maps (Li et al., 2025c; Zhang et al., 2025), occupancy (Yang et al., 2025; Mei et al., 2025; Tong et al., 2023), or sensor observations (Xia et al., 2026; Xiong et al., 2026; Hu et al., 2023a; Wang et al., 2024b; Li et al., 2025a) to support trajectory planning.

However, existing world models primarily focus on merely reconstructing future scene representations, while how to define their prediction targets for planning remains largely overlooked, as shown in Fig. 1(a). Latent-prediction methods commonly use self-supervised latent targets obtained by encoding

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

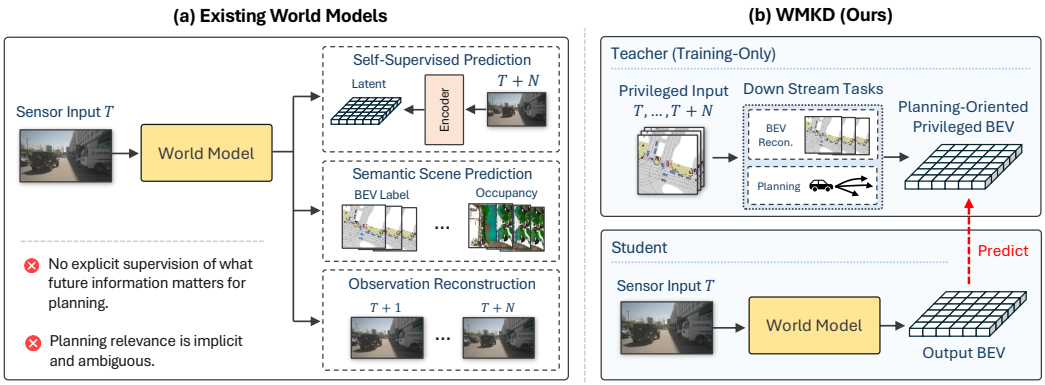

Figure 1: Comparison of world-model learning paradigms for end-to-end driving. (a) Existing world models predict self-supervised latent features, future observations, or semantic scene states, without explicitly specifying which future information is relevant to planning. (b) We instead construct a planning-oriented BEV target from privileged future observations under downstream task supervision and train a student world model to predict this target from the current observation.

future observations with the same encoder that processes current observations and is jointly updated during training (Li & Cui, 2024; Li et al., 2024a; Zheng et al., 2025; Yang et al., 2026). Features encoded by this shared encoder are optimized to support planning at their corresponding timesteps, rather than to emphasize the information relevant to the current ego decision. Indeed, such future-feature supervision has been shown to degrade planning performance (Zhang et al., 2026). Other methods predict semantic BEV maps (Li et al., 2025c; Zhang et al., 2025), occupancy (Yang et al., 2025; Mei et al., 2025; Tong et al., 2023), or sensor observations (Xia et al., 2026; Xiong et al., 2026; Hu et al., 2023a; Wang et al., 2024b; Li et al., 2025a; Zhou et al., 2025a). These targets explicitly describe future scenes, but are primarily optimized for reconstruction rather than for preserving information relevant to the current planning decision. A recent approach directly models transitions between current and future observations (Shang et al., 2026). Yet the downstream planner must still infer which information in these transition features matters for the current decision. These limitations in existing works raise a fundamental question: *"What should a world model predict for planning?"* This question motivates us to reconsider the target of the world model from the perspective of planning, rather than defining it directly from future latent, scenes, or transitions without considering their relevance to the planning decision.

To this end, we present World Model Knowledge Distillation (WMKD), a teacher–student framework in which a future-guided teacher constructs a planning-oriented target and a student learns to predict and exploit it for trajectory planning. As shown in Fig. 1(b), the teacher encodes ground-truth future semantic evolution under joint semantic reconstruction and ego planning supervision, and quantizes the resulting representation into discrete codes using a learnable codebook (Zhang et al., 2024). The student comprises four components: Discrete World Model (DWM), Residual Fusion Module (RFM), Planning Response Distillation (PRD), and Trajectory Repropagation Planner (TRP). DWM predicts the teacher codes from the observed BEV, enabling anticipation of planning-relevant future knowledge without privileged inputs. RFM fuses the reconstructed future representation with the observed BEV and their residual, while PRD aligns the student's planning responses with those of the teacher. We further propose TRP, which iteratively refines trajectory candidates and re-encodes their updated geometry for subsequent stages. Extensive experiments on NAVSIM and interactive closed-loop Bench2Drive validate WMKD, achieving 92.0 PDMS and 90.6 EPDMS on NAVSIM, as well as 87.56 DS and 72.27% SR on Bench2Drive.

Our key contributions are summarized as follows:

- We address the largely overlooked problem of designing world model prediction targets for planning, using a future-guided teacher to construct a planning-oriented target from future scene information relevant to the current planning decision.
- We propose World Model Knowledge Distillation (WMKD) to transfer this planning-oriented target to the student through the Discrete World Model (DWM), Planning Response Distillation (PRD), Residual Fusion Module (RFM), and Trajectory Repropagation Planner (TRP).

- Extensive experiments across NAVSIM benchmarks and the interactive closed-loop Bench2Drive benchmark validate the proposed framework, achieving 92.0 PDMS and 90.6 EPDMS on NAVSIM, as well as 87.56 DS and 72.27% SR on Bench2Drive.

2 Related Work

End-to-End (E2E) Autonomous Driving. Early autonomous driving systems relied on modular pipelines that separate perception, prediction, and planning (Zhou & Tuzel, 2018; Liu et al., 2023; Li et al., 2024c; Zhou et al., 2023; Moon et al., 2024; Yin et al., 2021; Yan et al., 2018; Ngiam et al., 2021; Jang et al., 2024). End-to-end (E2E) methods instead learn driving policies directly from sensor inputs (Hu et al., 2022; Chitta et al., 2022; Hu et al., 2023b; Jiang et al., 2023; Weng et al., 2024; Zeng et al., 2019; Xiao et al., 2023). Recent approaches employ anchor-based planning to provide trajectory priors and structure the search over candidate trajectories (Jiang et al., 2026a; Sun et al., 2025). In particular, dense trajectory scoring methods evaluate large candidate sets to select suitable driving trajectories (Li et al., 2024b; Sun et al., 2026; Li et al., 2025e). Other approaches explore iterative proposal refinement and trajectory selection (Guo et al., 2026; Yao et al., 2026; Tang et al., 2025), diffusion- and flow-based trajectory generation (Zheng et al., 2024; Liao et al., 2025; Xing et al., 2025; Wang et al., 2026), residual trajectory modeling (Zheng et al., 2026), and sparse-query architectures (Tang et al., 2025; Song et al., 2025; Zhao et al., 2026). Despite these advances, learning to imitate expert trajectories does not explicitly supervise how future changes in surrounding agents and scene structure should inform the current driving decision. WMKD addresses this limitation by incorporating predicted, planning-relevant future scene evolution into E2E trajectory planning.

World Models and Knowledge Distillation for Planning. World models provide predictive context for ego planning by learning future latent features (Li et al., 2024a; Li & Cui, 2024; Zheng et al., 2025; Yang et al., 2026), semantic BEV maps (Li et al., 2025c; Zhang et al., 2025), occupancy (Yang et al., 2025; Mei et al., 2025; Tong et al., 2023; Li et al., 2026a), or sensor observations (Xia et al., 2026; Xiong et al., 2026; Zhao et al., 2025; Zeng et al., 2026; Li et al., 2026b). Other approaches model temporal residuals to capture scene changes (Zhang et al., 2026) or incorporate language-based reasoning (Zhou et al., 2025a; Shang et al., 2026). In parallel, prior works on knowledge distillation transfer privileged scene representations (Chen et al., 2020; Zhang et al., 2023; Jia et al., 2023), planning-relevant features and responses (Feng et al., 2024; Yu et al., 2025), or policy knowledge informed by learned dynamics (Chen et al., 2021; Jiang et al., 2026b). Although privileged teachers can learn effective planning policies, regressing their exact feature values alone does not explicitly supervise how the future scene evolves from the student's current observation. World models address this need through explicit future prediction, but targets derived from future observations or scene dynamics are not necessarily optimized for their relevance to the current ego decision. WMKD bridges these approaches by learning a future prediction target from ground-truth scene evolution under the current planning objective and supervising discrete code prediction, allowing the student to anticipate scene changes that inform trajectory planning.

3 Method

As illustrated in Fig. 2, WMKD transfers planning-relevant future knowledge from a future-guided teacher to a student operating on sensor observations alone. The teacher constructs a planning-oriented target from ground-truth future evolution under semantic reconstruction and ego planning supervision, and quantizes it into discrete codes (Sec. 3.1). The student comprises four components: Discrete World Model (DWM), Residual Fusion Module (RFM), Planning Response Distillation (PRD), and Trajectory Repropagation Planner (TRP). DWM predicts the teacher codes from the observed scene, while RFM fuses the reconstructed future representation with the observed BEV and their residual to capture scene transitions. PRD further transfers the teacher's planning responses by aligning trajectory predictions and candidate scores, and TRP progressively refines trajectories by re-encoding their updated geometry for subsequent planning stages (Secs. 3.2 and 3.3).

3.1 Future-Guided Teacher Construction

Privileged Future Encoding. During training, the privileged teacher observes ground-truth BEV segmentation maps over the planning horizon, $\mathbf{S}_{T:T+N} \in \mathbb{R}^{(N+1)\times C_{\text{seg}}\times H\times W}$, where T is the

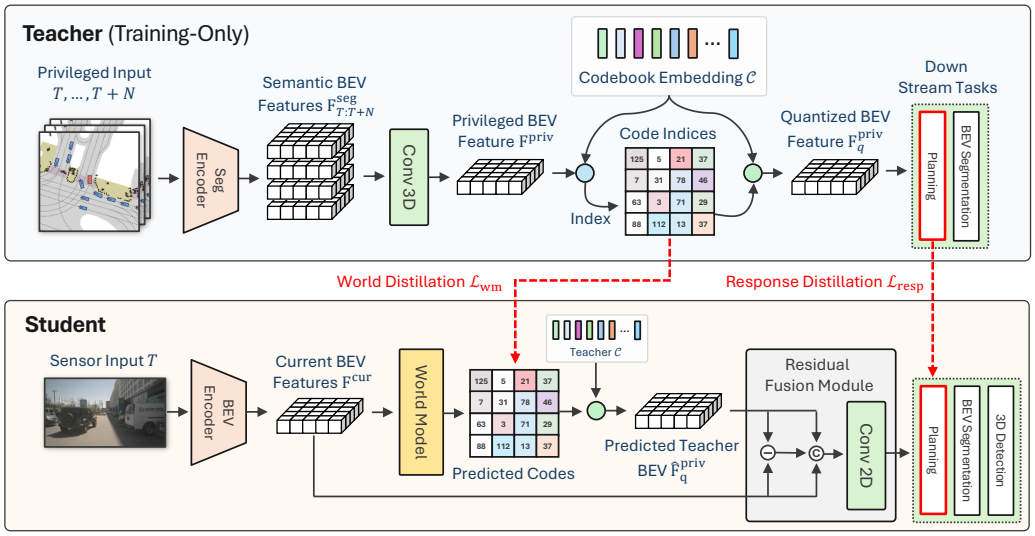

Figure 2: Overview of WMKD. (i) A training-only privileged teacher encodes future scene evolution into a discrete BEV representation. (ii) The current-only student predicts the corresponding teacher codes through world distillation and recovers the privileged future representation using the shared codebook. (iii) Residual Fusion Module (RFM) combines the predicted future representation with the current BEV, while Planning Response Distillation (PRD) transfers the teacher's planning behavior.

current timestep, N is the number of future frames, and C_{seg} is the number of semantic classes. A shared segmentation encoder $\mathcal{E}_{\text{seg}}$ extracts spatial features from each timestep as $\mathbf{F}^{\text{seg}}_{T:T+N} = \mathcal{E}_{\text{seg}}(\mathbf{S}_{T:T+N}) \in \mathbb{R}^{(N+1)\times C\times H\times W}$. A temporal encoder $\mathcal{E}_{\text{temp}}$ then aggregates their temporal evolution into $\mathbf{F}^{\text{priv}} = \mathcal{E}_{\text{temp}}(\mathbf{F}^{\text{seg}}_{T:T+N}) \in \mathbb{R}^{C\times H\times W}$. Thus, $\mathbf{F}^{\text{priv}}$ summarizes scene evolution over the entire planning horizon, while retaining the future information that the privileged teacher finds most relevant to the current planning decision.

Discretization. However, directly using this continuous privileged representation as the world model target requires the current-only student to reproduce exact feature values which are constructed from future observations. Such features contain fine-grained variations available to the privileged teacher that are difficult to recover at the same precision from the current observation alone. We therefore reduce the granularity of the prediction target by quantizing $\mathbf{F}^{\text{priv}}$ into a finite vocabulary of reusable future patterns. We use a learnable codebook $\mathcal{C} = \{\mathbf{e}_k\}_{k=1}^{K}$ to obtain a discrete indices map $\mathbf{Y}^{\text{priv}} \in \{1, \ldots, K\}^{H\times W}$ following VQ-Map (Zhang et al., 2024), yielding quantized representation $\mathbf{F}^{\text{priv}}_q \in \mathbb{R}^{C\times H\times W}$. This discretization converts exact continuous feature matching into prediction over reusable future patterns, providing a more tractable target for the current-only world model.

Planning-Oriented Supervision. Quantization reduces the complexity of the prediction target, but does not by itself determine what information each code should preserve. We therefore jointly optimize $\mathbf{F}^{\text{priv}}_q$ for ego planning and semantic reconstruction:

$$\mathcal{L}_{\text{teacher}} = \lambda_{\text{plan}}\mathcal{L}_{\text{plan}} + \lambda_{\text{map}}\mathcal{L}_{\text{map}} + \lambda_{\text{vq}}\mathcal{L}_{\text{vq}}, \tag{1}$$

where $\mathcal{L}_{\text{plan}}$, $\mathcal{L}_{\text{map}}$, and $\mathcal{L}_{\text{vq}}$ denote the planning, semantic-reconstruction, and vector-quantization objectives, respectively. Semantic reconstruction encourages the representation to encode future scene evolution, while planning supervision encourages it to capture future information useful for planning at the current timestep. The resulting code map therefore provides a compact, planning-oriented world model target for the student network.

3.2 Knowledge Distillation for Discrete World Model

Discrete World Model (DWM). After training the privileged teacher, we freeze both the teacher and its codebook and use the privileged code map $\mathbf{Y}^{\text{priv}} \in \{1, \ldots, K\}^{H\times W}$ as the prediction target of the student world model. Following prior works (Chitta et al., 2022; Zimmerlin et al., 2024), current multi-view images $\mathbf{I}_T$ and LiDAR observation $\mathbf{L}_T$ are encoded and fused into a current BEV feature

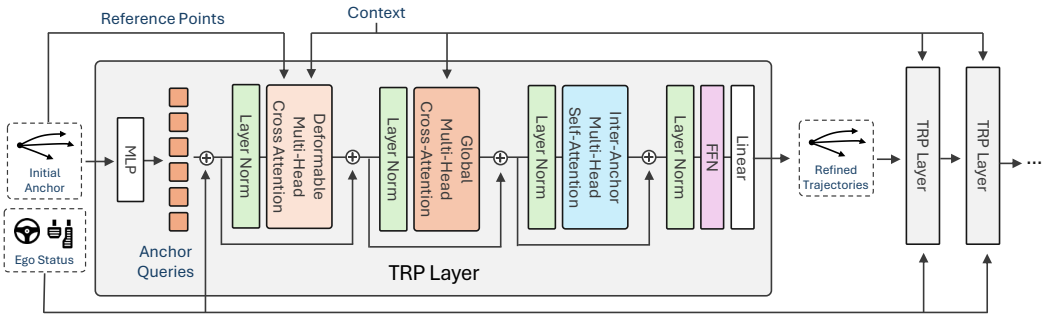

Figure 3: Overview of the Trajectory Repropagation Planner (TRP). (i) Initial trajectory anchors and ego status are encoded into planning queries. (ii) Each TRP layer uses trajectory-conditioned deformable attention, global cross-attention, and inter-anchor self-attention to refine trajectories. (iii) Refined trajectories are re-encoded as queries for the next layer, enabling progressive refinement.

$\mathbf{F}^{\mathrm{cur}} \in \mathbb{R}^{C\times H\times W}$. The student world model $\mathcal{W}_\theta$, composed of deformable self-attention layers (Li et al., 2024c), predicts privileged-code logits:

$$\mathbf{P}^{\mathrm{priv}} = \mathcal{W}_\theta(\mathbf{F}^{\mathrm{cur}}) \in \mathbb{R}^{K\times H\times W}, \qquad \mathcal{L}_{\mathrm{wm}} = \mathrm{CE}\big(\mathbf{P}^{\mathrm{priv}}, \mathbf{Y}^{\mathrm{priv}}\big), \tag{2}$$

where $\mathcal{L}_{\mathrm{wm}}$ denotes the world distillation loss, and $\mathrm{CE}(\cdot)$ is the cross-entropy loss over code indices. Through this objective, the student predicts the discrete future patterns encoded by the planning-oriented teacher from the current observation.

Residual Fusion Module (RFM). Predicted future representations (Zhang et al., 2025; Xia et al., 2026) and their residuals relative to the current scene (Zhang et al., 2026) provide useful context for downstream planning. We therefore propose the Residual Fusion Module (RFM), which combines the current BEV, predicted future representation, and their residual. First, the predicted privileged BEV is reconstructed from the frozen teacher codebook $\mathcal{C}$ using $\boldsymbol{\pi}^{\mathrm{priv}} = \mathrm{softmax}_k(\mathbf{P}^{\mathrm{priv}})$ and $\hat{\mathbf{F}}_q^{\mathrm{priv}} = \sum_{k=1}^{K} \boldsymbol{\pi}_k^{\mathrm{priv}} \mathbf{e}_k$, where $\boldsymbol{\pi}^{\mathrm{priv}}$ denotes the predicted code distribution. RFM then applies layer normalization (LN) to obtain $\bar{\mathbf{F}}^{\mathrm{cur}} = \mathrm{LN}(\mathbf{F}^{\mathrm{cur}})$ and $\bar{\mathbf{F}}^{\mathrm{priv}} = \mathrm{LN}(\hat{\mathbf{F}}_q^{\mathrm{priv}})$, constructs their residual, and fuses the three representations:

$$\mathbf{F}^{\Delta} = \bar{\mathbf{F}}^{\mathrm{priv}} - \bar{\mathbf{F}}^{\mathrm{cur}}, \qquad \mathbf{F}^{\mathrm{fuse}} = \mathcal{E}_{\mathrm{rfm}}([\bar{\mathbf{F}}^{\mathrm{cur}}; \bar{\mathbf{F}}^{\mathrm{priv}}; \mathbf{F}^{\Delta}]). \tag{3}$$

where $\mathcal{E}_{\mathrm{rfm}}$ is a lightweight 2D convolutional network. The fused representation provides the planner with the current scene, predicted future scene, and their explicit transition.

Planning Response Distillation (PRD). Representation-level world distillation transfers future knowledge but does not directly supervise the student's planning outputs. Inspired by prior works (Hong et al., 2022; Yu et al., 2025; Feng et al., 2024), we adopt response distillation to align the student with the privileged teacher at trajectory and candidate-evaluation levels. The response distillation loss is defined as:

$$\mathcal{L}_{\mathrm{resp}} = \lambda_{\mathrm{traj}}\mathcal{L}_{\mathrm{traj}}^{\mathrm{KD}} + \lambda_{\mathrm{imi}}\mathcal{L}_{\mathrm{imi}}^{\mathrm{KD}} + \lambda_{\mathrm{score}}\mathcal{L}_{\mathrm{score}}^{\mathrm{KD}}, \tag{4}$$

where λ_{traj}, λ_{imi}, and λ_{score} are weighting coefficients. $\mathcal{L}_{\mathrm{traj}}^{\mathrm{KD}}$ aligns teacher and student trajectories, $\mathcal{L}_{\mathrm{imi}}^{\mathrm{KD}}$ transfers the teacher's trajectory preference, and $\mathcal{L}_{\mathrm{score}}^{\mathrm{KD}}$ aligns their PDMS-based sub-scores. Together, these losses provide direct supervision on the student's planning outputs. Detailed formulations of the loss terms are provided in Appendix A.2.

3.3 Trajectory Repropagation Planner (TRP)

To effectively exploit the planning-oriented future representation transferred from the privileged teacher, we adopt an anchor-based cascaded planner that progressively refines trajectories across intermediate layers (Liao et al., 2025; Sun et al., 2025; Liu et al., 2025; Zheng et al., 2026). However, conventional cascaded planners propagate latent queries through successive decoder layers even as the trajectory references are continuously updated. This creates a mismatch between the planning query and trajectory geometry: the retrieved scene features correspond to the latest trajectory, whereas

the propagated query may still encode information associated with previous hypotheses. Such misalignment can limit the utilization of both planning-oriented future knowledge and current-scene context. To address this, we propose Trajectory Repropagation Planner (TRP), which reconstructs each planning query from the latest refined trajectory to maintain alignment throughout refinement.

As shown in Fig. 3, we design TRP with N_{trp} iterative refinement layers, each comprising three complementary attention modules:

$$\begin{aligned} \text{D-MHCA}(\mathbf{Q}, \mathbf{A}) &= \text{DeformAttn}(\mathbf{Q}, \mathbf{F}^{\text{fuse}}; \mathcal{R}(\mathbf{A})), \\ \text{G-MHCA}(\mathbf{Q}) = \text{MHA}(\mathbf{Q}, \mathbf{F}^{\text{cur}}, \mathbf{F}^{\text{cur}}), &\quad \text{IA-MHSA}(\mathbf{Q}) = \text{MHA}(\mathbf{Q}, \mathbf{Q}, \mathbf{Q}). \end{aligned} \tag{5}$$

Here, $\mathcal{R}(\cdot)$ converts trajectories into BEV reference points, and $\text{DeformAttn}(\cdot)$ samples features around these points using learned offsets and attention weights (Zhu et al., 2020). D-MHCA retrieves candidate-specific features from $\mathbf{F}^{\text{fuse}}$ along the latest trajectory hypotheses. Next, G-MHCA retrieves global context from the current BEV, with BEV features flattened into spatial tokens for cross-attention. Finally, IA-MHSA exchanges information among candidates. $\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ denotes multi-head attention (Vaswani et al., 2017) with query, key, and value inputs. Initial queries are constructed from trajectory anchors $\mathbf{A}^{(0)} \in \mathbb{R}^{N_{\text{anchor}} \times T \times 2}$ and ego-status embeddings $\mathbf{E}_{\text{ego}} \in \mathbb{R}^{1 \times C}$, which are shared across anchors. At each layer, the attention-refined queries $\mathbf{Q}^{(\ell)}$ predict trajectory offsets, and the updated trajectories are re-encoded to initialize the next layer:

$$\mathbf{A}^{(\ell)} = \mathbf{A}^{(\ell-1)} + \text{MLP}_{\text{reg}}(\mathbf{Q}^{(\ell)}), \qquad \mathbf{Q}_0^{(\ell+1)} = \text{MLP}_{\text{traj}}(\mathbf{A}^{(\ell)}) + \mathbf{E}_{\text{ego}}. \tag{6}$$

Queries from all refinement layers are concatenated to capture complementary evidence across refinement stages and predict the imitation and PDMS-based sub-scores for each anchor:

$$\left(p_i^{\text{imi}}, \{s_{i,m}\}_{m \in \mathcal{M}}\right) = \left(\text{MLP}_{\text{imi}}, \text{MLP}_{\text{score}}\right) \left([\mathbf{Q}_i^{(1)}; \cdots; \mathbf{Q}_i^{(N_{\text{trp}})}]\right). \tag{7}$$

The final-layer trajectories $\mathbf{A}^{(N_{\text{trp}})}$ are used as trajectory candidates, and the highest-scoring candidate is selected at inference (Li et al., 2024b; 2025c).

4 Experiments

4.1 Experimental Setup

4.1.1 Datasets and Metrics

NAVSIM and Metrics. NAVSIM (Dauner et al., 2024) evaluates E2E driving through non-reactive simulation on nuPlan logs (Caesar et al., 2021), with 1,192 `navtrain` and 136 `navtest` logs. We report the Predictive Driver Model Score (PDMS), which evaluates no-at-fault collision (NC), drivable-area compliance (DAC), ego progress (EP), time to collision (TTC), and comfort (Comf.). NAVSIM v2 (Cao et al., 2025) extends this evaluation with the Extended Predictive Driver Model Score (EPDMS), replacing comfort with lane keeping (LK), history-based comfort (HC), and extended comfort (EC), and adding driving-direction compliance (DDC) and traffic-light compliance (TLC).

Bench2Drive and Metrics. Bench2Drive (Jia et al., 2024) evaluates E2E driving through interactive closed-loop simulation in CARLA, covering safety-critical scenarios across 220 routes. We report Driving Score (DS) and Success Rate (SR) to assess overall closed-loop performance, together with fine-grained driving-ability scores for Merging, Overtaking, Emergency Braking, Give Way, Traffic Sign Compliance, and their abilities mean. Further details of the datasets and evaluation metrics are provided in Appendix B.1.1.

4.1.2 Implementation Details

We use ResNet-34 (He et al., 2016) and V2-99 (Lee et al., 2019) image backbones with input resolutions of 1024×256 and 2048×512, respectively, and fuse camera and rasterized LiDAR BEV features following TransFuser (Chitta et al., 2022). The teacher and student are trained from scratch on `navtrain` using AdamW with a learning rate of 6×10^{-4} on 8 NVIDIA RTX A6000 GPUs. The teacher is trained for 60 epochs with a total batch size of 512, taking 3 hours, while the student is trained for 100 epochs with a total batch size of 256, taking 10 hours. The privileged future

Table 1: Comparison of state-of-the-art methods on the NAVSIM v1 `navtest` split. C and L denote camera and LiDAR inputs, respectively. Best and second best results are highlighted in bold and underlined, respectively.

Method	Input	Backbone	PDMS ↑	NC ↑	DAC ↑	TTC ↑	Comf. ↑	EP ↑
Human	–	–	94.8	100.0	100.0	100.0	99.9	87.5
Traditional End-to-End Methods								
UniAD (2023b)	C	ResNet-34	83.4	97.8	91.9	92.9	**100.0**	78.8
Transfuser (2022)	C&L	ResNet-34	84.0	97.7	92.8	92.8	**100.0**	79.2
Hydra-MDP++ (2025b)	C	ResNet-34	86.6	97.6	96.0	93.1	**100.0**	80.4
DiffusionDrive (2025)	C&L	ResNet-34	88.1	98.2	96.2	94.7	**100.0**	82.2
GaussianFusion (2025)	C&L	ResNet-34	88.8	98.3	97.2	94.6	–	83.0
VADv2 (2026a)	C	ResNet-34	89.3	98.3	97.4	95.7	**100.0**	82.3
ResAD (2026)	C&L	ResNet-34	88.8	98.0	97.5	94.1	**100.0**	83.3
UniTeD (2026)	C	ResNet-34	90.2	**98.9**	97.2	96.6	**100.0**	84.1
MeanFuser (2026)	C	ResNet-34	89.0	98.6	97.0	95.0	**100.0**	82.8
World Models								
LAW (2024a)	C	ResNet-34	84.6	96.4	95.4	88.7	99.9	81.7
World4Drive (2025)	C	ResNet-34	85.1	97.4	94.3	92.8	**100.0**	79.9
WoTE (2025c)	C&L	ResNet-34	88.3	98.5	96.8	94.9	99.9	81.9
SeerDrive (2025)	C&L	ResNet-34	88.9	98.4	97.0	94.9	99.9	83.2
WorldRFT (2026)	C	ResNet-34	87.0	97.5	96.0	94.0	**100.0**	80.9
ResWorld (2026)	C&L	ResNet-34	89.0	**98.9**	96.5	95.6	**100.0**	83.1
WMKD (Ours)	C&L	ResNet-34	**92.0**	98.5	**98.5**	**97.9**	99.8	**85.7**
Hydra-MDP++ (2025b)	C	V2-99	91.0	98.6	98.6	95.1	**100.0**	85.7
GoalFlow (2025)	C&L	V2-99	90.3	98.4	98.3	94.6	**100.0**	85.0
GaussianFusion (2025)	C&L	V2-99	92.0	98.7	98.1	95.7	–	**88.2**
SeerDrive (2025)	C&L	V2-99	90.7	98.8	98.6	95.8	**100.0**	84.2
ResAD (2026)	C&L	V2-99	90.6	**98.9**	97.8	94.9	**100.0**	87.0
WMKD (Ours)	C&L	V2-99	**92.4**	98.5	**98.8**	**98.0**	99.7	86.0

Table 2: Comparison of state-of-the-art methods on the NAVSIM v2 `navtest` split. C and L denote camera and LiDAR inputs, respectively. Best and second best results are highlighted in bold and underlined, respectively, within each backbone group.

Method	Input	Backbone	EPDMS ↑	NC ↑	DAC ↑	DDC ↑	TLC ↑	EP ↑	TTC ↑	LK ↑	HC ↑	EC ↑
Human	–	–	94.5	100.0	100.0	99.8	100.0	87.4	100.0	100.0	98.1	90.1
Traditional End-to-End Methods												
TransFuser (2022)	C&L	ResNet-34	76.6	96.9	89.9	97.9	99.7	87.1	95.4	92.7	98.3	87.2
Hydra-MDP++ (2025b)	C	ResNet-34	81.4	97.2	97.5	99.4	99.6	83.1	96.5	94.4	98.2	70.9
DiffusionDrive (2025)	C&L	ResNet-34	84.8	98.2	96.2	99.5	99.8	87.4	97.4	97.0	98.3	87.8
GaussianFusion (2025)	C&L	ResNet-34	85.0	98.3	97.3	99.0	97.4	87.5	97.4	97.4	98.3	–
ResAD (2026)	C&L	ResNet-34	85.5	97.8	97.2	99.5	99.8	88.2	96.9	97.0	98.4	**88.2**
MeanFuser (2026)	C	ResNet-34	89.5	98.3	97.2	**99.6**	99.8	87.6	97.4	97.3	98.3	**88.2**
DiffRefiner (2026)	C	ResNet-34	86.2	98.5	97.4	**99.6**	99.8	87.6	97.7	97.7	98.3	86.2
UniTeD (2026)	C	ResNet-34	90.1	**99.1**	97.2	**99.6**	**99.9**	86.9	**98.5**	**98.4**	**99.9**	87.3
World Models												
WoTE (2025c)	C&L	ResNet-34	84.1	98.5	96.8	98.8	99.8	86.0	97.9	95.4	98.3	83.0
WMKD (Ours)	C&L	ResNet-34	**90.6**	98.5	**98.5**	99.4	99.7	**88.8**	97.9	96.6	98.3	85.4
Hydra-MDP++ (2025b)	C	V2-99	85.1	98.4	98.0	99.4	99.8	87.5	97.7	95.3	**98.3**	77.4
DiffRefiner (2026)	C	V2-99	87.4	**98.6**	98.5	**99.7**	**99.9**	87.6	**98.1**	**98.1**	**98.3**	**87.9**
WMKD (Ours)	C&L	V2-99	**91.1**	98.5	**98.8**	99.6	99.8	**88.8**	98.0	97.0	**98.3**	85.4

BEV spans a 4-second horizon with 8 future frames. The planner uses 256 trajectory anchors and $N_{\text{trp}} = 4$ TRP layers to predict 8 waypoints over a 4-second horizon. Further implementation details for NAVSIM and Bench2Drive are provided in Appendix B.1.2.

4.2 Quantitative Results

Evaluation on NAVSIM and NAVSIM v2. Tab. 1 presents the results of WMKD on the NAVSIM `navtest` split. With a ResNet-34 backbone, WMKD achieves 92.0 PDMS, including 97.9 TTC and 85.7 EP. These results indicate that planning-oriented future knowledge supports both collision-aware planning and driving progress. With the stronger V2-99 backbone, WMKD achieves 92.4 PDMS, showing that the framework remains effective across perception backbones. Tab. 2 further evaluates WMKD under the extended driving criteria of NAVSIM v2. It achieves 90.6 EPDMS with ResNet-34 and 91.1 EPDMS with V2-99. The V2-99 model also achieves 98.8 DAC and 88.8 EP, reflecting strong drivable-area compliance and driving progress. Together, these results demonstrate

Table 3: Comparison of methods on Bench2Drive under closed-loop metrics and driving abilities. Best and second best results are highlighted in bold and underlined, respectively.

Method	Overall		Multi-Ability (%) ↑					
	DS ↑	SR (%) ↑	Merging	Overtaking	Emergency Brake	Give Way	Traffic Sign	Mean
Traditional End-to-End Methods								
UniAD (2023b)	45.81	16.36	14.10	17.78	21.67	10.00	14.21	15.55
VAD (2023)	42.35	15.00	8.11	24.44	18.64	20.00	19.15	18.07
TF++ (2024)	84.21	67.27	58.75	57.77	83.33	<u>40.00</u>	82.11	64.39
SparseDrive (2025)	48.73	18.52	11.00	16.67	25.00	**50.00**	16.00	23.76
DriveTransformer (2025)	63.46	35.01	17.57	35.00	48.36	<u>40.00</u>	52.10	38.60
Hydra-NeXt (2025d)	73.86	50.00	40.00	64.44	61.67	**50.00**	50.00	53.22
MomAD (2025)	47.91	18.11	13.21	21.02	18.01	20.00	21.07	18.66
HiP-AD (2025)	86.77	69.09	50.00	**84.44**	83.33	<u>40.00</u>	72.10	65.98
VADv2 (2026a)	76.15	50.46	–	–	–	–	–	–
DiffRefiner (2026)	87.10	<u>71.40</u>	<u>63.80</u>	60.00	<u>85.00</u>	**50.00**	**86.30**	<u>69.00</u>
UniTeD (2026)	<u>87.30</u>	70.00	–	–	–	–	–	–
World Models								
WoTE (2025c)	61.71	31.36	–	–	–	–	–	–
SeerDrive (2025)	58.32	30.17	–	–	–	–	–	–
WMKD (Ours)	**87.56**	**72.27**	**66.86**	<u>65.18</u>	**88.33**	**50.00**	<u>85.28</u>	**71.13**

Table 4: Ablation study on the influence of each component on the NAVSIM v1 `navtest` split.

Model	Description	PDMS ↑	NC ↑	DAC ↑	EP ↑	TTC ↑
$\mathcal{M}_0$	Base Model	84.7	97.6	92.1	78.3	96.7
$\mathcal{M}_1$	$\mathcal{M}_0$ + DWM	89.5	98.4	96.7	82.4	97.6
$\mathcal{M}_2$	$\mathcal{M}_1$ + TRP	90.7	98.0	97.7	84.2	97.5
$\mathcal{M}_3$	$\mathcal{M}_2$ + PRD	91.1	98.4	97.8	84.5	97.8
$\mathcal{M}_4$	$\mathcal{M}_2$ + RFM	91.1	98.4	97.8	84.6	97.8
$\mathcal{M}_5$	$\mathcal{M}_4$ + PRD	91.3	98.3	98.1	84.8	97.9

Table 5: Ablation of future knowledge transfer on NAVSIM v1 `navtest` split.

Setting	PDMS ↑	NC ↑	DAC ↑	TTC ↑	EP ↑
Baseline	84.7	97.6	92.1	96.7	78.3
Direct Distillation					
Teacher BEV	88.5	98.1	95.9	93.3	81.5
w/ World Model					
Future BEV Seg.	89.2	98.5	96.2	97.8	82.1
Teacher BEV	89.1	98.5	96.1	97.7	82.1
Teacher Codes	89.5	98.4	96.7	97.6	82.4

the effectiveness of WMKD in transferring planning-oriented future knowledge across perception backbones and evaluation settings.

Evaluation on Bench2Drive. Tab. 3 evaluates WMKD in interactive closed-loop driving scenarios on Bench2Drive. WMKD achieves 87.56 DS, 72.27% SR, and 71.13% mean driving ability, demonstrating strong overall performance across diverse interactive scenarios. The method further achieves competitive results on challenging driving abilities such as Merging and Emergency Braking, reflecting its effectiveness in handling dynamic interactions. These results complement the NAVSIM evaluations and demonstrate that planning-oriented future knowledge remains effective not only in non-reactive simulation but also under closed-loop interaction with the dynamic traffic environment.

4.3 Ablation Studies

Effect of Discrete World Model and Planning Components. Tab. 4 evaluates the contributions of the proposed Discrete World Model (DWM) and planning components (i.e., TRP, RFM, and PRD). Starting from the baseline $\mathcal{M}_0$ with 84.7 PDMS, $\mathcal{M}_1$ introduces DWM and substantially improves performance to 89.5 PDMS, demonstrating the benefit of predicting planning-oriented future knowledge from the observed scene. Adding TRP in $\mathcal{M}_2$ further improves PDMS to 90.7, showing that iterative trajectory refinement effectively exploits the predicted future representation. Building on $\mathcal{M}_2$, PRD in $\mathcal{M}_3$ and RFM in $\mathcal{M}_4$ independently improve PDMS to 91.1 by transferring the teacher's planning responses and explicitly modeling the current-to-future transition, respectively. Finally, $\mathcal{M}_5$ combines RFM and PRD and achieves the best performance of 91.3 PDMS, demonstrating their complementary benefits within the unified framework.

Effect of Future Knowledge Transfer. Tab. 5 examines how privileged future knowledge should be transferred to the current-only student. (i) Direct feature distillation from the continuous teacher BEV improves PDMS from 84.7 to 88.5, confirming that privileged future information can provide effective planning supervision even through feature-level distillation. (ii) Using the same teacher BEV as an explicit world model prediction target further improves performance to 89.1 PDMS, showing that predicting future knowledge is more effective than feature-level alignment alone. (iii) Predicting future BEV segmentation achieves a comparable 89.2 PDMS, demonstrating that explicit

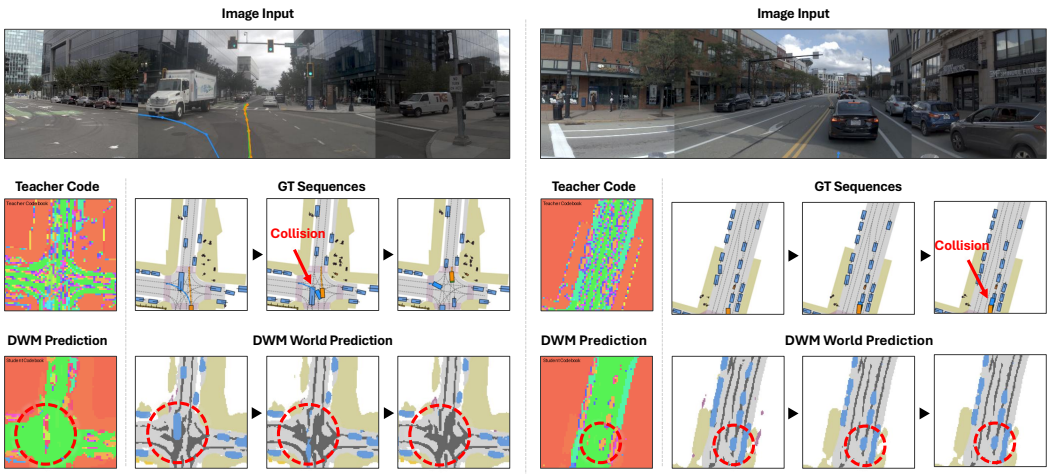

Figure 4: Visual comparison of WMKD. Trajectories are shown in blue for DiffusionDrive, orange for WMKD, and green for ground-truth. Red dashed circles highlight regions of interest in the world model predictions.

future-scene prediction also provides strong planning cues. (iv) Our discrete world model achieves the best performance of 89.5 PDMS, demonstrating the benefit of explicit world modeling and a compact discrete prediction target for transferring privileged future knowledge to the student. This result highlights the importance of the prediction target representation in effective world modeling. Additional ablation studies are provided in Appendix B.3.

4.4 Qualitative Results

Fig. 4 visualizes the privileged teacher codes, student predictions, and planning results in two interaction scenarios. Although the student predicts coarser code maps than the teacher, the recovered future sequences preserve important scene changes from the ground-truth future. (i) In the left example, WMKD anticipates a moving vehicle crossing the intersection and generates a trajectory that better follows the ground truth, whereas DiffusionDrive (Liao et al., 2025) collides with the vehicle. (ii) In the right example, DiffusionDrive fails to anticipate the stopped vehicle ahead and generates a colliding trajectory, while WMKD predicts the vehicle's future state and produces a more appropriate trajectory. These examples illustrate how discrete world modeling transfers planning-relevant future information from the privileged teacher, enabling better trajectory planning in dynamic traffic scenarios. Additional qualitative results are provided in Appendix B.4.

5 Conclusion

We present World Model Knowledge Distillation (WMKD) to address the overlooked question of what a world model should predict for planning in end-to-end autonomous driving. While existing works primarily focus on reconstructing future observations or scene representations, WMKD employs a future-guided teacher that encodes ground-truth scene evolution into a planning-oriented BEV representation under joint supervision from future semantic reconstruction and ego planning. The student predicts this representation from current observations through discrete code prediction. The Residual Fusion Module (RFM) combines the predicted representation with the current BEV and their residual, while Planning Response Distillation (PRD) transfers the teacher's trajectory predictions and candidate evaluations. The Trajectory Repropagation Planner (TRP) re-encodes refined trajectory geometry to guide successive refinement. WMKD achieves 92.0 PDMS and 90.6 EPDMS on NAVSIM, as well as a Driving Score (DS) of 87.56 and a Success Rate (SR) of 72.27% on Bench2Drive. These results highlight the importance of defining world model prediction targets according to their relevance to ego planning. We hope this work will inspire future research on planning-oriented world models for autonomous driving.

AI USE STATEMENT

In this work, we used generative AI tools for language editing, including grammar correction and sentence refinement. We have not used generative AI tools for generating research ideas, designing the methodology or experiments, analyzing experimental results, or generating scientific claims, and the rest of the required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools for improving the clarity and readability of the manuscript. We have reviewed all AI-assisted work. All AI-assisted language edits were manually reviewed by the authors to ensure that they preserved the intended technical meaning and scientific claims. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2021.

Wei Cao, Marcel Hallgarten, Tianyu Li, Daniel Dauner, Xunjiang Gu, Caojun Wang, Yakov Miron, Marco Aiello, Hongyang Li, Igor Gilitschenski, et al. Pseudo-simulation for autonomous driving. *arXiv preprint arXiv:2506.04218*, 2025.

Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In *Conference on robot learning*, pp. 66–75. PMLR, 2020.

Dian Chen, Vladlen Koltun, and Philipp Krähenbühl. Learning to drive from a world on rails. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15570–15579. IEEE, 2021.

Kashyap Chitta, Aditya Prakash, Bernhard Jaeger, Zehao Yu, Katrin Renz, and Andreas Geiger. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. *IEEE transactions on pattern analysis and machine intelligence*, 45(11):12878–12895, 2022.

Daniel Dauner, Marcel Hallgarten, Tianyu Li, Xinshuo Weng, Zhiyu Huang, Zetong Yang, Hongyang Li, Igor Gilitschenski, Boris Ivanovic, Marco Pavone, et al. Navsim: Data-driven non-reactive autonomous vehicle simulation and benchmarking. *Advances in Neural Information Processing Systems*, 37:28706–28719, 2024.

Kairui Ding, Boyuan Chen, Yuchen Su, Huan-ang Gao, Bu Jin, Chonghao Sima, Wuqiang Zhang, Xiaohui Li, Paul Barsch, Hongyang Li, et al. Hint-ad: Holistically aligned interpretability in end-to-end autonomous driving. *arXiv preprint arXiv:2409.06702*, 2024.

Kaituo Feng, Changsheng Li, Dongchun Ren, Ye Yuan, and Guoren Wang. On the road to portability: Compressing end-to-end motion planner for autonomous driving. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15099–15108. IEEE, 2024.

Haoyu Fu, Diankun Zhang, Zongchuang Zhao, Jianfeng Cui, Dingkang Liang, Chong Zhang, Dingyuan Zhang, Hongwei Xie, Bing Wang, and Xiang Bai. Orion: A holistic end-to-end autonomous driving framework by vision-language instructed action generation. *arXiv preprint arXiv:2503.19755*, 2025.

Ke Guo, Haochen Liu, Xiaojun Wu, Jia Pan, and Chen Lv. ipad: Iterative proposal-centric end-to-end autonomous driving. *IEEE Robotics and Automation Letters*, 2026.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Yu Hong, Hang Dai, and Yong Ding. Cross-modality knowledge distillation network for monocular 3d object detection. In *European Conference on Computer Vision*, pp. 87–104. Springer, 2022.

Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, and Gianluca Corrado. Gaia-1: A generative world model for autonomous driving, 2023a. URL https://arxiv.org/abs/2309.17080.

Shengchao Hu, Li Chen, Penghao Wu, Hongyang Li, Junchi Yan, and Dacheng Tao. St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning. In *European Conference on Computer Vision*, pp. 533–549. Springer, 2022.

Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17853–17862, 2023b.

Jyh-Jing Hwang, Runsheng Xu, Hubert Lin, Wei-Chih Hung, Jingwei Ji, Kristy Choi, Di Huang, Tong He, Paul Covington, Benjamin Sapp, et al. Emma: End-to-end multimodal model for autonomous driving. *arXiv preprint arXiv:2410.23262*, 2024.

Jiyun Jang, Mincheol Chang, Jongwon Park, and Jinkyu Kim. Finetuning pre-trained model with limited data for lidar-based 3d object detection by bridging domain gaps. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7145–7151. IEEE, 2024.

Xiaosong Jia, Yulu Gao, Li Chen, Junchi Yan, Patrick Langechuan Liu, and Hongyang Li. Driveadapter: Breaking the coupling barrier of perception and planning in end-to-end autonomous driving. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7919–7929. IEEE, 2023.

Xiaosong Jia, Zhenjie Yang, Qifeng Li, Zhiyuan Zhang, and Junchi Yan. Bench2drive: Towards multi-ability benchmarking of closed-loop end-to-end autonomous driving. *Advances in Neural Information Processing Systems*, 37:819–844, 2024.

Xiaosong Jia, Junqi You, Zhiyuan Zhang, and Junchi Yan. Drivetransformer: Unified transformer for scalable end-to-end autonomous driving. *arXiv preprint arXiv:2503.07656*, 2025.

Bo Jiang, Shaoyu Chen, Qing Xu, Bencheng Liao, Jiajie Chen, Helong Zhou, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Vad: Vectorized scene representation for efficient autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8340–8350, 2023.

Bo Jiang, Shaoyu Chen, Bencheng Liao, Xingyu Zhang, Wei Yin, Qian Zhang, Chang Huang, Wenyu Liu, and Xinggang Wang. Senna: Bridging large vision-language models and end-to-end autonomous driving. *arXiv preprint arXiv:2410.22313*, 2024.

Bo Jiang, Shaoyu Chen, Hao Gao, Bencheng Liao, Qian Zhang, Wenyu Liu, and Xinggang Wang. Vadv2: End-to-end vectorized autonomous driving via probabilistic planning. In *International Conference on Learning Representations*, volume 2026, pp. 68886–68900, 2026a.

Guangfeng Jiang, Yueru Luo, Jun Liu, Yi Huang, Yiyao Zhu, Zhan Qu, Dave Zhenyu Chen, Bingbing Liu, and Xu Yan. Wpt: World-to-policy transfer via online world model distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17842–17852, 2026b.

Youngwan Lee, Joong-won Hwang, Sangrok Lee, Yuseok Bae, and Jongyoul Park. An energy and gpu-computation efficient backbone network for real-time object detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 752–760. IEEE, 2019.

Jingyu Li, Bozhou Zhang, Xin Jin, Jiankang Deng, Xiatian Zhu, and Li Zhang. Imagidrive: A unified imagination-and-planning framework for autonomous driving. *arXiv preprint arXiv:2508.11428*, 2025a.

Jingyu Li, Junjie Wu, Dongnan Hu, Xiangkai Huang, Bin Sun, Zhihui Hao, Xianpeng Lang, Xiatian Zhu, and Li Zhang. Sgdrive: Scene-to-goal hierarchical world cognition for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4032–4042, June 2026a.

Kailin Li, Zhenxin Li, Shiyi Lan, Yuan Xie, Zhizhong Zhang, Jiayi Liu, Zuxuan Wu, Zhiding Yu, and Jose M Alvarez. Hydra-mdp++: Advancing end-to-end driving via expert-guided hydra-distillation. *arXiv preprint arXiv:2503.12820*, 2025b.

Peidong Li and Dixiao Cui. Navigation-guided sparse scene representation for end-to-end autonomous driving. *arXiv preprint arXiv:2409.18341*, 2024.

Yingyan Li, Lue Fan, Jiawei He, Yuqi Wang, Yuntao Chen, Zhaoxiang Zhang, and Tieniu Tan. Enhancing end-to-end autonomous driving with latent world model. *arXiv preprint arXiv:2406.08481*, 2024a.

Yingyan Li, Yuqi Wang, Yang Liu, Jiawei He, Lue Fan, and Zhaoxiang Zhang. End-to-end driving with online trajectory evaluation via bev world model. *arXiv preprint arXiv:2504.01941*, 2025c.

Yingyan Li, Shuyao Shang, Weisong Liu, Bing Zhan, Haochen Wang, Yuqi Wang, Yuntao Chen, Xiaoman Wang, Yasong An, Chufeng Tang, et al. Drivevla-w0: World models amplify data scaling law in autonomous driving. In *International Conference on Learning Representations*, volume 2026, pp. 7890–7911, 2026b.

Zhenxin Li, Kailin Li, Shihao Wang, Shiyi Lan, Zhiding Yu, Yishen Ji, Zhiqi Li, Ziyue Zhu, Jan Kautz, Zuxuan Wu, et al. Hydra-mdp: End-to-end multimodal planning with multi-target hydra-distillation. *arXiv preprint arXiv:2406.06978*, 2024b.

Zhenxin Li, Shihao Wang, Shiyi Lan, Zhiding Yu, Zuxuan Wu, and Jose M Alvarez. Hydra-next: Robust closed-loop driving with open-loop training. *arXiv preprint arXiv:2503.12030*, 2025d.

Zhenxin Li, Wenhao Yao, Zi Wang, Xinglong Sun, Joshua Chen, Nadine Chang, Maying Shen, Zuxuan Wu, Shiyi Lan, and Jose M. Alvarez. Generalized trajectory scoring for end-to-end multimodal planning, 2025e. URL https://arxiv.org/abs/2506.06664.

Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Qiao Yu, and Jifeng Dai. Bevformer: learning bird's-eye-view representation from lidar-camera via spatiotemporal transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024c.

Zhiqi Li, Zhiding Yu, Shiyi Lan, Jiahan Li, Jan Kautz, Tong Lu, and Jose M Alvarez. Is ego status all you need for open-loop end-to-end autonomous driving? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14864–14873, 2024d.

Bencheng Liao, Shaoyu Chen, Haoran Yin, Bo Jiang, Cheng Wang, Sixu Yan, Xinbang Zhang, Xiangyu Li, Ying Zhang, Qian Zhang, et al. Diffusiondrive: Truncated diffusion model for end-to-end autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12037–12047, 2025.

Shuai Liu, Quanmin Liang, Zefeng Li, Boyang Li, and Kai Huang. Gaussianfusion: Gaussian-based multi-sensor fusion for end-to-end autonomous driving. *arXiv preprint arXiv:2506.00034*, 2025.

Zhijian Liu, Haotian Tang, Alexander Amini, Xinyu Yang, Huizi Mao, Daniela L Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In *2023 IEEE international conference on robotics and automation (ICRA)*, pp. 2774–2781. IEEE, 2023.

Jiageng Mao, Yuxi Qian, Junjie Ye, Hang Zhao, and Yue Wang. Gpt-driver: Learning to drive with gpt. *arXiv preprint arXiv:2310.01415*, 2023.

Jianbiao Mei, Yu Yang, Xuemeng Yang, Licheng Wen, Jiajun Lv, Botian Shi, and Yong Liu. Vision-centric 4d occupancy forecasting and planning via implicit residual world models. *arXiv preprint arXiv:2510.16729*, 2025.

Seokha Moon, Hyun Woo, Hongbeen Park, Haeji Jung, Reza Mahjourian, Hyung-gun Chi, Hyerin Lim, Sangpil Kim, and Jinkyu Kim. Visiontrap: Vision-augmented trajectory prediction guided by textual descriptions. In *European Conference on Computer Vision*, pp. 361–379. Springer, 2024.

Jiquan Ngiam, Benjamin Caine, Vijay Vasudevan, Zhengdong Zhang, Hao-Tien Lewis Chiang, Jeffrey Ling, Rebecca Roelofs, Alex Bewley, Chenxi Liu, Ashish Venugopal, et al. Scene transformer: A unified architecture for predicting multiple agent trajectories. *arXiv preprint arXiv:2106.08417*, 2021.

Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10425–10433, 2020. doi: 10.1109/CVPR42600.2020.01044.

Shuyao Shang, Bing Zhan, Yunfei Yan, Yuqi Wang, Yingyan Li, Yasong An, Xiaoman Wang, Jierui Liu, Lu Hou, Lue Fan, et al. Dynvla: Learning world dynamics for action reasoning in autonomous driving. *arXiv preprint arXiv:2603.11041*, 2026.

Ziying Song, Caiyan Jia, Lin Liu, Hongyu Pan, Yongchang Zhang, Junming Wang, Xingyu Zhang, Shaoqing Xu, Lei Yang, and Yadan Luo. Don't shake the wheel: Momentum-aware planning in end-to-end autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 22432–22441, 2025.

Wenchao Sun, Xuewu Lin, Yining Shi, Chuang Zhang, Haoran Wu, and Sifa Zheng. Sparsedrive: End-to-end autonomous driving via sparse scene representation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8795–8801. IEEE, 2025.

Wenchao Sun, Xuewu Lin, Keyu Chen, Zixiang Pei, Xiang Li, Yining Shi, and Sifa Zheng. Sparsedrivev2: Scoring is all you need for end-to-end autonomous driving. *arXiv preprint arXiv:2603.29163*, 2026.

Yingqi Tang, Zhuoran Xu, Zhaotie Meng, and Erkang Cheng. Hip-ad: Hierarchical and multi-granularity planning with deformable attention for autonomous driving in a single decoder. *arXiv preprint arXiv:2503.08612*, 2025.

Wenwen Tong, Chonghao Sima, Tai Wang, Li Chen, Silei Wu, Hanming Deng, Yi Gu, Lewei Lu, Ping Luo, Dahua Lin, et al. Scene as occupancy. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8372–8381. IEEE, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Junli Wang, Yinan Zheng, Xueyi Liu, Zebin Xing, Pengfei Li, Kun Ma, Hangjun Ye, Guang Chen, Guang Li, Long Chen, et al. Meanfuser: Fast one-step multi-modal trajectory generation and adaptive reconstruction via meanflow for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17884–17893, 2026.

Shihao Wang, Zhiding Yu, Xiaohui Jiang, Shiyi Lan, Min Shi, Nadine Chang, Jan Kautz, Ying Li, and Jose M Alvarez. Omnidrive: A holistic llm-agent framework for autonomous driving with 3d perception, reasoning and planning. *CoRR*, 2024a.

Yuqi Wang, Jiawei He, Lue Fan, Hongxin Li, Yuntao Chen, and Zhaoxiang Zhang. Driving into the future: Multiview visual forecasting and planning with world model for autonomous driving. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14749–14759. IEEE, 2024b.

Xinshuo Weng, Boris Ivanovic, Yan Wang, Yue Wang, and Marco Pavone. Para-drive: Parallelized architecture for real-time autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15449–15458, 2024.

Tianze Xia, Yongkang Li, Lijun Zhou, Jingfeng Yao, Kaixin Xiong, Haiyang Sun, Bing Wang, Kun Ma, Guang Chen, Hangjun Ye, et al. Drivelaw: Unifying planning and video generation in a latent driving world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 39701–39712, 2026.

Yi Xiao, Felipe Codevilla, Diego Porres, and Antonio M López. Scaling vision-based end-to-end autonomous driving with multi-view attention learning. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1586–1593. IEEE, 2023.

Zebin Xing, Xingyu Zhang, Yang Hu, Bo Jiang, Tong He, Qian Zhang, Xiaoxiao Long, and Wei Yin. Goalflow: Goal-driven flow matching for multimodal trajectories generation in end-to-end autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 1602–1611, 2025.

Zhexiao Xiong, Xin Ye, Burhan Yaman, Sheng Cheng, Yiren Lu, Jingru Luo, Nathan Jacobs, and Liu Ren. Unidrive-wm: Unified understanding, planning and generation world model for autonomous driving. *arXiv preprint arXiv:2601.04453*, 2026.

Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10), 2018. ISSN 1424-8220. doi: 10.3390/s18103337. URL https://www.mdpi.com/1424-8220/18/10/3337.

Pengxuan Yang, Ben Lu, Zhongpu Xia, Chao Han, Yinfeng Gao, Teng Zhang, Kun Zhan, XianPeng Lang, Yupeng Zheng, and Qichao Zhang. Worldrft: Latent world model planning with reinforcement fine-tuning for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11649–11657, 2026.

Yu Yang, Jianbiao Mei, Yukai Ma, Siliang Du, Wenqing Chen, Yijie Qian, Yuxiang Feng, and Yong Liu. Driving in the occupancy world: Vision-centric 4d occupancy forecasting and planning via world models for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 9327–9335, 2025.

Wenhao Yao, Zhenxin Li, Shiyi Lan, Zi Wang, Xinglong Sun, Jose M Alvarez, and Zuxuan Wu. Drivesuprim: Towards precise trajectory selection for end-to-end planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11910–11918, 2026.

Liuhan Yin, Runkun Ju, Guodong Guo, and Erkang Cheng. Diffrefiner: Coarse to fine trajectory planning via diffusion refinement with semantic interaction for end to end autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12009–12017, 2026.

Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11784–11793, 2021.

Rui Yu, Xianghang Zhang, Runkai Zhao, Huaicheng Yan, and Meng Wang. Distilldrive: End-to-end multi-mode autonomous driving distillation by isomorphic hetero-source planning model. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26188–26197. IEEE, 2025.

Shuang Zeng, Xinyuan Chang, Mengwei Xie, Xinran Liu, Yifan Bai, Zheng Pan, Mu Xu, and Xing Wei. Futuresightdrive: Thinking visually with spatio-temporal cot for autonomous driving. *Advances in Neural Information Processing Systems*, 38:67299–67318, 2026.

Wenyuan Zeng, Wenjie Luo, Simon Suo, Abbas Sadat, Bin Yang, Sergio Casas, and Raquel Urtasun. End-to-end interpretable neural motion planner. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8660–8669, 2019.

Bozhou Zhang, Nan Song, Jingyu Li, Xiatian Zhu, Jiankang Deng, and Li Zhang. Future-aware end-to-end driving: Bidirectional modeling of trajectory planning and scene evolution. *arXiv preprint arXiv:2510.11092*, 2025.

Jimuyang Zhang, Zanming Huang, and Eshed Ohn-Bar. Coaching a teachable student. In *CVPR*, 2023.

Jinqing Zhang, Zehua Fu, Zelin Xu, Wenying Dai, Qingjie Liu, and Yunhong Wang. Resworld: Temporal residual world model for end-to-end autonomous driving. *arXiv preprint arXiv:2602.10884*, 2026.

Yiwei Zhang, Jin Gao, Fudong Ge, Guan Luo, Bing Li, Zhaoxiang Zhang, Haibin Ling, and Weiming Hu. Vq-map: Bird's-eye-view map layout estimation in tokenized discrete space via vector quantization. *Advances in Neural Information Processing Systems*, 37:70453–70475, 2024.

Bo Zhao, Xinting Zhao, Naifan Li, Erkang Cheng, and Haibin Ling. United: Unified temporal diffusion for joint perception and planning in autonomous driving. *arXiv preprint arXiv:2606.25736*, 2026.

Zhida Zhao, Talas Fu, Yifan Wang, Lijun Wang, and Huchuan Lu. From forecasting to planning: Policy world model for collaborative state-action prediction. *arXiv preprint arXiv:2510.19654*, 2025.

Wenzhao Zheng, Ruiqi Song, Xianda Guo, Chenming Zhang, and Long Chen. Genad: Generative end-to-end autonomous driving. In *European Conference on Computer Vision*, pp. 87–104. Springer, 2024.

Yupeng Zheng, Pengxuan Yang, Zebin Xing, Qichao Zhang, Yuhang Zheng, Yinfeng Gao, Pengfei Li, Teng Zhang, Zhongpu Xia, Peng Jia, et al. World4drive: End-to-end autonomous driving via intention-aware physical latent world model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 28632–28642, 2025.

Zhiyu Zheng, Shaoyu Chen, Haoran Yin, Xinbang Zhang, Jialv Zou, Xinggang Wang, Qian Zhang, and Lefei Zhang. Resad: Normalized residual trajectory modeling for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3729–3739, 2026.

Xin Zhou, Dingkang Liang, Sifan Tu, Xiwu Chen, Yikang Ding, Dingyuan Zhang, Feiyang Tan, Hengshuang Zhao, and Xiang Bai. Hermes: A unified self-driving world model for simultaneous 3d scene understanding and generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27817–27827, 2025a.

Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4490–4499, 2018.

Zewei Zhou, Tianhui Cai, Seth Z Zhao, Yun Zhang, Zhiyu Huang, Bolei Zhou, and Jiaqi Ma. Autovla: A vision-language-action model for end-to-end autonomous driving with adaptive reasoning and reinforcement fine-tuning. *arXiv preprint arXiv:2506.13757*, 2025b.

Zikang Zhou, Jianping Wang, Yung-Hui Li, and Yu-Kai Huang. Query-centric trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17863–17873, 2023.

Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020.

Julian Zimmerlin, Jens Beißwenger, Bernhard Jaeger, Andreas Geiger, and Kashyap Chitta. Hidden biases of end-to-end driving datasets. *arXiv preprint arXiv:2412.09602*, 2024.

APPENDIX

A ADDITIONAL METHOD DETAILS

A.1 CODEBOOK CONSTRUCTION

Following VQ-Map (Zhang et al., 2024), we construct a discrete representation of the privileged BEV feature through vector quantization. Let $F^{\text{priv}} \in \mathbb{R}^{C\times H\times W}$ denote the continuous representation produced by the privileged teacher. We introduce a learnable codebook $\mathcal{C} = \{\mathbf{e}_k\}_{k=1}^{K}$, where $\mathbf{e}_k \in \mathbb{R}^C$ is the k-th code embedding. At each BEV location (h, w), the code index is determined by nearest-neighbor assignment:

$$Y^{\text{priv}}_{h,w} = \arg\min_{k\in\{1,\dots,K\}} \left\| F^{\text{priv}}_{h,w} - \mathbf{e}_k \right\|_2^2, \qquad F^{\text{priv}}_{q,h,w} = \mathbf{e}_{Y^{\text{priv}}_{h,w}}, \tag{8}$$

yielding the discrete code map $Y^{\text{priv}} \in \{1,\dots,K\}^{H\times W}$ and the quantized representation $F^{\text{priv}}_q \in \mathbb{R}^{C\times H\times W}$. The codebook is optimized using the vector-quantization objective (Zhang et al., 2024):

$$\mathcal{L}_{\text{vq}} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\left[\left\|\text{sg}(F^{\text{priv}}_{h,w}) - F^{\text{priv}}_{q,h,w}\right\|_2^2 + \beta\left\|F^{\text{priv}}_{h,w} - \text{sg}(F^{\text{priv}}_{q,h,w})\right\|_2^2\right]. \tag{9}$$

where $\text{sg}(\cdot)$ denotes stop-gradient and $\beta = 4$ controls the commitment loss. The first term updates the code embeddings toward their assigned privileged features, while the second encourages the encoder outputs to remain close to their selected embeddings. The resulting quantized representation is jointly optimized for ego planning and future semantic reconstruction through the teacher objective $\mathcal{L}_{\text{teacher}}$. After teacher training, the codebook is frozen and Y^{priv} serves as the prediction target for the student's Discrete World Model (DWM).

A.2 DETAILS FOR OBJECTIVES

Planning Loss. Following prior works (Li et al., 2024b; 2025c), we supervise trajectory regression, imitation scores, and PDMS-based sub-scores. The same planning loss is applied independently to both the teacher and the student. Let $\mathbf{A}^{(0)} \in \mathbb{R}^{N_{\text{anchor}}\times T\times 2}$ denote the initial trajectory anchors and $\boldsymbol{\tau}^{\text{GT}} \in \mathbb{R}^{T\times 2}$ the expert trajectory. We use $\mathcal{T}$, p_i^{imi}, and $s_{i,m}$ to denote the trajectory predictions, imitation probabilities, and predicted sub-scores of the corresponding model. For a reference trajectory $\boldsymbol{\tau}$, we define the anchor distance and its corresponding imitation target as:

$$d_i(\boldsymbol{\tau}) = \frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{A}^{(0)}_{i,t} - \boldsymbol{\tau}_t\right\|_2^2, \qquad q_i(\boldsymbol{\tau}) = \frac{\exp[-d_i(\boldsymbol{\tau})]}{\sum_{j=1}^{N_{\text{anchor}}}\exp[-d_j(\boldsymbol{\tau})]}. \tag{10}$$

Following the winner-take-all strategy (Liao et al., 2025), only the refined trajectory corresponding to the anchor closest to the expert trajectory receives regression supervision:

$$i^*_{\text{GT}} = \arg\min_i d_i(\boldsymbol{\tau}^{\text{GT}}), \qquad \mathcal{L}_{\text{traj}} = \frac{1}{T}\sum_{t=1}^{T}\left\|\mathcal{T}_{i^*_{\text{GT}},t} - \boldsymbol{\tau}^{\text{GT}}_t\right\|_1. \tag{11}$$

Let p_i^{imi} denote the predicted imitation probability for anchor i. The imitation loss is the cross-entropy between the predicted distribution and the distance-based target (Li et al., 2024b; 2025c):

$$\mathcal{L}_{\text{imi}} = -\sum_{i=1}^{N_{\text{anchor}}} q_i(\boldsymbol{\tau}^{\text{GT}})\log p_i^{\text{imi}}. \tag{12}$$

Following prior works (Li et al., 2024b; 2025c), we compute PDMS-based sub-scores $\bar{s}_{i,m}$ for each initial trajectory anchor offline, where $m \in \mathcal{M}$ indexes the sub-metrics. The predicted sub-scores $s_{i,m}$ are supervised by binary cross-entropy:

$$\mathcal{L}_{\text{score}} = -\frac{1}{N_{\text{anchor}}|\mathcal{M}|}\sum_{i=1}^{N_{\text{anchor}}}\sum_{m\in\mathcal{M}}\left[\bar{s}_{i,m}\log s_{i,m} + (1-\bar{s}_{i,m})\log(1-s_{i,m})\right]. \tag{13}$$

The planning loss shared by the teacher and student is as follows:

$$\mathcal{L}_{\text{plan}} = \lambda_{\text{traj}}\mathcal{L}_{\text{traj}} + \lambda_{\text{imi}}\mathcal{L}_{\text{imi}} + \lambda_{\text{score}}\mathcal{L}_{\text{score}}. \tag{14}$$

Planning Response Distillation (PRD). Planning Response Distillation (PRD) uses the teacher's selected trajectory as the pseudo-GT for student trajectory regression and imitation supervision, while the teacher's predicted PDMS-based sub-scores supervise student scoring. Let $i_{\text{sel}}^{\text{tea}}$ denote the candidate selected by the teacher. The pseudo-GT trajectory is $\boldsymbol{\tau}^{\text{tea}} = \mathcal{T}_{i_{\text{sel}}^{\text{tea}}}^{\text{tea}}$.

As in the original planning loss, the student regresses only the trajectory corresponding to the anchor closest to the pseudo-GT:

$$i_{\text{KD}}^{*} = \arg\min_{i} d_i(\boldsymbol{\tau}^{\text{tea}}), \qquad \mathcal{L}_{\text{traj}}^{\text{KD}} = \frac{1}{T}\sum_{t=1}^{T}\left\|\mathcal{T}_{i_{\text{KD}}^{*},t}^{\text{stu}} - \boldsymbol{\tau}_t^{\text{tea}}\right\|_1. \tag{15}$$

The imitation target is likewise recomputed using the teacher's selected trajectory:

$$\mathcal{L}_{\text{imi}}^{\text{KD}} = -\sum_{i=1}^{N_{\text{anchor}}} q_i(\boldsymbol{\tau}^{\text{tea}}) \log p_i^{\text{stu,imi}}. \tag{16}$$

For candidate scoring, the teacher's predicted sub-scores replace the offline targets:

$$\mathcal{L}_{\text{score}}^{\text{KD}} = -\frac{1}{N_{\text{anchor}}|\mathcal{M}|}\sum_{i=1}^{N_{\text{anchor}}}\sum_{m\in\mathcal{M}}\left[s_{i,m}^{\text{tea}}\log s_{i,m}^{\text{stu}} + (1-s_{i,m}^{\text{tea}})\log(1-s_{i,m}^{\text{stu}})\right]. \tag{17}$$

Total Objective. The overall student objective combines planning, perception, world model distillation, and planning response distillation losses:

$$\mathcal{L}_{\text{student}} = \lambda_{\text{plan}}\mathcal{L}_{\text{plan}} + \lambda_{\text{map}}\mathcal{L}_{\text{map}} + \lambda_{\text{3D}}\mathcal{L}_{\text{3D}} + \lambda_{\text{wm}}\mathcal{L}_{\text{wm}} + \lambda_{\text{resp}}\mathcal{L}_{\text{resp}}. \tag{18}$$

Following TransFuser (Chitta et al., 2022), $\mathcal{L}_{\text{map}}$ is an auxiliary perception objective shared by the teacher and student, while $\mathcal{L}_{\text{3D}}$ is applied only to the student.

B Additional Experiments

B.1 Experimental Setup

B.1.1 Datasets and Metrics

NAVSIM and Metrics. We evaluate WMKD on NAVSIM (Dauner et al., 2024) and NAVSIM v2 (Cao et al., 2025), which assess driving performance through non-reactive simulation on nuPlan logs (Caesar et al., 2021). NAVSIM uses the Predictive Driver Model Score (PDMS), which combines no-at-fault collision (NC), drivable-area compliance (DAC), ego progress (EP), time to collision (TTC), and comfort (Comf.):

$$\text{PDMS} = \text{NC}\cdot\text{DAC}\cdot\frac{5\text{EP} + 5\text{TTC} + 2\text{Comf.}}{12}. \tag{19}$$

NC and DAC serve as multiplicative safety penalties, while EP, TTC, and Comf. form a weighted measure of driving progress, collision risk, and motion quality.

NAVSIM v2 extends this evaluation with the Extended Predictive Driver Model Score (EPDMS). It introduces driving-direction compliance (DDC) and traffic-light compliance (TLC) as additional penalties, and replaces the original comfort metric with lane keeping (LK), history-based comfort (HC), and extended comfort (EC):

$$\text{EPDMS} = \text{NC}\cdot\text{DAC}\cdot\text{DDC}\cdot\text{TLC}\cdot\frac{5\text{EP} + 5\text{TTC} + 2\text{LK} + 2\text{HC} + 2\text{EC}}{16}. \tag{20}$$

EPDMS additionally accounts for violations in the recorded human trajectory when determining applicable penalties. Both metrics are computed per frame and averaged over the evaluation sequence. Further details are provided in Tab. A1.

Bench2Drive and Metrics. We evaluate WMKD on Bench2Drive (Jia et al., 2024), a CARLA-based benchmark that assesses driving performance in interactive and safety-critical scenarios. Following our experimental setup, we use the dataset collected by TF++ (Zimmerlin et al., 2024) for training and evaluate closed-loop performance on 220 predefined routes. Bench2Drive allows the ego vehicle's actions to affect subsequent observations and interactions with surrounding traffic.

Table A1: Metrics used in PDMS and EPDMS for the NAVSIM benchmark.

Metric	EPDMS only	Weight	Range
No at-fault Collisions (NC)		multiplier	$\{0, \frac{1}{2}, 1\}$
Drivable Area Compliance (DAC)		multiplier	$\{0, 1\}$
Ego Progress (EP)		5	$[0, 1]$
Time to Collision (TTC)		5	$\{0, 1\}$
Comfort (C)		2	$\{0, 1\}$
Driving Direction Compliance (DDC)	✓	multiplier	$\{0, \frac{1}{2}, 1\}$
Traffic Light Compliance (TLC)	✓	multiplier	$\{0, 1\}$
Lane Keeping (LK)	✓	2	$\{0, 1\}$
History Comfort (HC)	✓	2	$\{0, 1\}$
Extended Comfort (EC)	✓	2	$\{0, 1\}$

We report Driving Score (DS), which combines route completion with penalties for driving infractions, and Success Rate (SR), which measures the proportion of successfully completed routes. We additionally report Efficiency (Eff.) and Comfort (Com.) to assess driving efficiency and motion quality. To examine performance beyond aggregate scores, we evaluate five fine-grained driving abilities: Merging, Overtaking, Emergency Braking, Give Way, and Traffic Sign compliance. The mean driving-ability score summarizes performance across these scenarios, providing a complementary measure of the model's behavior in diverse interactive driving situations.

B.1.2 Implementation Details

Implementation for NAVSIM. We follow TransFuser (Chitta et al., 2022) to encode and fuse camera and LiDAR observations into a BEV representation. For the ResNet-34 (He et al., 2016) setting, three forward-facing camera views are cropped, resized, and concatenated into a 1024×256 image. For V2-99 (Lee et al., 2019), we use an input resolution of 2048×512. LiDAR observations are rasterized into a BEV representation and fused with camera features by the perception encoder. The teacher receives ground-truth BEV maps from the current frame to the next 8 future frames, spanning a 4-second horizon.

The privileged representation is quantized using a codebook with 128 entries of dimension 256. The student world model consists of 2 deformable self-attention layers with 8 attention heads and predicts the teacher code index at each BEV location. For planning, initial trajectory anchors are obtained by K-means clustering over expert trajectories (Jiang et al., 2026a; Li et al., 2024b; 2025b). We use $N_{\text{anchor}} = 256$ trajectory anchors and $N_{\text{trp}} = 4$ TRP layers. The planner predicts 8 future waypoints over a 4-second planning horizon. Following prior works (Li et al., 2025c; 2024b), we pre-compute PDMS-based sub-scores for each trajectory anchor and use them as supervision for scoring. For teacher training, we set $\lambda_{\text{plan}} = 12$, $\lambda_{\text{map}} = 14$, and $\lambda_{\text{vq}} = 1$. For student training, we use $\lambda_{\text{plan}} = 12$, $\lambda_{\text{map}} = 14$, $\lambda_{\text{3D}} = 10$, $\lambda_{\text{wm}} = 1$, and $\lambda_{\text{resp}} = 1$. Within the planning objectives, we set $\lambda_{\text{traj}} = 8$, $\lambda_{\text{imi}} = 1$, and $\lambda_{\text{score}} = 5$.

The teacher and student are trained from scratch on the NAVSIM `navtrain` split using AdamW with an initial learning rate of 6×10^{-4} and a weight decay of 10^{-4}. The teacher is trained for 60 epochs with a total batch size of 512, taking 3 hours, while the student is trained for 100 epochs with a total batch size of 256, taking 10 hours. Both models are trained on 8 NVIDIA RTX A6000 GPUs. For fast ablation studies, student training is shortened to 20 epochs. Inference speed is measured with a batch size of 1 on a single NVIDIA RTX A6000 GPU.

Implementation for Bench2Drive. For Bench2Drive evaluation, we follow the implementation of TF++ (Zimmerlin et al., 2024) which takes a front-view RGB image at a resolution of 1024×384 together with a single-frame LiDAR input, using RegNetY-032 (Radosavovic et al., 2020) backbone for both modalities. The teacher and student models are trained for 30 and 10 epochs, respectively. Unless otherwise specified, all other architectural and methodological settings follow those used for NAVSIM implementations.

B.2 Additional Quantitative Results

Model Size and Computational Efficiency. Tab. A2 compares planning performance and inference efficiency on a single NVIDIA RTX A6000 GPU. With one TRP layer, WMKD achieves 90.5 PDMS at 52 FPS. Increasing N_{trp} to 4 improves PDMS to 91.4 at 38 FPS, outperforming Diffusion-

Drive (Liao et al., 2025) (88.1 PDMS, 39 FPS) and ResAD (Zheng et al., 2026) (88.8 PDMS, 37 FPS) at comparable speeds. Using 8 layers increases the parameter count from 77M to 85M and reduces throughput to 27 FPS without further gains. Thus, 4 TRP layers provide the highest PDMS with a favorable performance–efficiency trade-off.

B.3 Additional Ablation Studies

Effect of Teacher Tasks. Tab. A3 studies how teacher supervision affects the quality of knowledge transferred to the student. Planning-only supervision in (a) achieves 90.8 PDMS, but provides limited supervision on the structure of future scene evolution. Conversely, segmentation-only supervision in (b) reaches 91.0 PDMS by preserving future scene structure, but does not explicitly emphasize its relevance to the current planning decision. Combining both objectives in (c) provides complementary supervision and achieves the best performance of 91.3 PDMS, jointly capturing structured future evolution and planning-relevant information. Further adding 3D detection in (d) decreases PDMS to 90.7, indicating that additional object-level supervision is not necessarily aligned with the desired prediction target for DWM.

Effect of Privileged Timesteps. Tab. A4 examines the importance of temporal scene evolution in privileged knowledge transfer. Using only the 0s frame, the teacher and student achieve 92.4 and 90.5 PDMS, respectively. Notably, the 0s student performs similarly to the TRP-only model (90.3 PDMS), indicating that privileged distillation without future evolution provides only marginal gains. Adding observations at 0s, 2s, and 4s improves their performance to 93.1 and 90.9 PDMS, while dense 0.5-second sampling further increases it to 93.9 and 91.3 PDMS. These consistent gains show that the benefit of privileged distillation primarily comes from modeling scene evolution over the planning horizon, enabling the teacher to construct a more informative planning-oriented target.

Effect of TRP and WMKD. Tab. A5 evaluates the benefit of TRP and the complete WMKD framework. Starting from the baseline with 84.7 PDMS, TRP alone already provides a strong planner, substantially improving performance to 90.3 PDMS. Nevertheless, the full WMKD framework further increases PDMS to 91.3, showing that planner refinement alone is insufficient to achieve the best performance and that planning-oriented future knowledge provides additional complementary gains.

Table A2: Model parameter size, inference speed, and planning performance.

Method	Params (M)	PDMS ↑	FPS ↑
Traditional End-to-End Methods			
TransFuser (2022)	56	84.0	63
Hydra-MDP (2024b)	65	86.5	20
GoalFlow (2025)	62	85.7	10
DiffusionDrive (2025)	61	88.1	39
MeanFuser (2026)	59	89.0	59
ResAD (2026)	68	88.8	37
World Models			
WoTE (2025c)	64	88.3	43
SeerDrive (2025)	68	88.9	40
Ours ($N_{trp} = 1$)	70	90.5	52
Ours ($N_{trp} = 2$)	73	91.1	45
Ours ($N_{trp} = 4$)	77	91.4	38
Ours ($N_{trp} = 8$)	85	91.1	27

Table A3: Ablation of privileged teacher tasks on the NAVSIM v1 `navtest` split. All metrics report the performance of the distilled student.

ID	Teacher Tasks			PDMS ↑	NC ↑	DAC ↑	TTC ↑	EP ↑
	Plan	Seg.	3D					
(a)	✓	-	-	90.8	98.4	97.5	97.9	84.2
(b)	-	✓	-	91.0	98.5	97.5	97.9	84.6
(c)	✓	✓	-	91.3	98.3	98.1	97.9	84.8
(d)	✓	✓	✓	90.7	98.1	97.7	97.6	84.2

Table A4: Ablation of privileged timesteps for teacher model on the NAVSIM v1 `navtest` split.

Model	PDMS ↑	NC ↑	DAC ↑	TTC ↑	EP ↑
Privileged Timesteps: 0s					
Teacher	92.4	98.5	99.1	97.9	85.7
Student	90.5	98.3	97.3	97.7	84.0
Privileged Timesteps: 0–2–4s					
Teacher	93.1	99.5	98.6	99.0	86.7
Student	90.9	98.4	97.7	97.8	84.5
Privileged Timesteps: 0–0.5–· · ·–3.5–4s					
Teacher	93.9	99.7	99.1	99.5	87.5
Student	91.3	98.3	98.1	97.9	84.8

Table A5: Ablation study of TRP and WMKD on the NAVSIM v1 `navtest` split.

Model	PDMS ↑	NC ↑	DAC ↑	TTC ↑	EP ↑
Baseline	84.7	97.6	92.1	96.7	78.3
(+) TRP	90.3	98.1	97.3	97.5	83.8
WMKD (Ours)	91.3	98.3	98.1	97.9	84.8

Effect of Codebook Size. Table A6 examines the effect of codebook size on world model distillation. A codebook with 128 entries achieves the highest PDMS of 91.3, outperforming codebooks with 64 and 256 entries by 0.5 and 0.7 points, respectively. A smaller codebook of 64 entries may lack the capacity to represent diverse future scene dynamics, limiting the information transferred from

Table A6: Ablation study of the codebook entry size on the NAVSIM v1 `navtest` split.

ID	Codebook Size	PDMS ↑	NC ↑	DAC ↑	TTC ↑	EP ↑
(a)	64	90.8	98.3	97.7	97.6	84.3
(b)	128	91.3	98.3	98.1	97.9	84.8
(c)	256	90.6	98.2	97.6	97.6	84.1

Table A7: Ablation study of residual fusion on the NAVSIM v1 `navtest` split.

Model	PDMS ↑	NC ↑	DAC ↑	TTC ↑	EP ↑
WMKD (Ours)	91.3	98.3	98.1	97.9	84.8
(-) $\mathbf{F}^{\Delta}$ Fusion	91.1	98.2	97.9	97.7	84.6

Table A8: Ablation of TRP attention modules and trajectory repropagation on the NAVSIM v1 `navtest` split. Note that WMKD is applied.

ID	Attention			PDMS ↑	NC ↑	DAC ↑	TTC ↑	EP ↑
	D	G	IA					
Query Propagation								
(a)	✓	-	-	90.4	98.3	97.3	97.7	83.9
(b)	✓	-	✓	90.9	98.4	97.6	97.8	84.5
(c)	✓	✓	-	91.0	98.5	97.6	97.9	84.6
(d)	✓	✓	✓	90.6	98.3	97.3	97.8	84.1
Trajectory Repropagation								
(e)	✓	-	-	90.7	98.1	97.7	97.6	84.1
(f)	✓	-	✓	91.1	98.3	98.0	97.8	84.5
(g)	✓	✓	-	91.1	98.2	97.9	97.7	84.5
(h)	✓	✓	✓	91.3	98.3	98.1	97.9	84.8

the privileged teacher. Conversely, increasing the size to 256 introduces more discrete codes that the student must distinguish from current observations, making future code prediction more challenging. The 128-entry codebook offers a balance between representation capacity and prediction difficulty, yielding the highest DAC, TTC, and EP while matching the best NC.

Effect of Explicit Transition Fusion. Tab. A7 examines the contribution of the explicit current-to-future transition in RFM. Removing $\mathbf{F}^{\Delta}$ decreases PDMS from 91.3 to 91.1. This result indicates that explicitly modeling the difference between the current and predicted future representations provides additional information beyond directly fusing the two states.

Ablation of the TRP Architecture. Tab. A8 examines the attention modules and trajectory repropagation strategy of TRP, with the world model applied to all variants. Under conventional query propagation, (a) D-MHCA alone achieves 90.4 PDMS. Adding (b) IA-MHSA or (c) G-MHCA improves performance to 90.9 and 91.0 PDMS, respectively. However, combining all three modules in (d) reduces performance to 90.6 PDMS. We attribute this degradation to IA-MHSA mixing information across anchors, which can progressively weaken the correspondence between each propagated query and its own refined trajectory geometry. Trajectory repropagation mitigates this mismatch by reconstructing each query from its updated anchor-specific geometry at every layer. Accordingly, (f) and (g) reach 91.1 PDMS, while the full configuration (h) achieves the best performance of 91.3 PDMS, improving over its query-propagation counterpart (d) by 0.7 PDMS.

B.4 Additional Qualitative Results

Fig. A1 presents additional qualitative results across six driving scenes, comparing the student's predicted codes and future BEV sequences with the corresponding teacher codes and ground-truth sequences. The examples include straight road segments and intersections with different surrounding-vehicle configurations. Across successive future timesteps, the predicted BEV sequences capture the road layout and the evolving positions of surrounding vehicles. These visualizations illustrate how WMKD transfers privileged future information to the student and uses it to anticipate scene evolution for planning.

C Limitations

WMKD remains dependent on the performance of the privileged teacher, as the quality of the learned prediction target is influenced by the future knowledge captured by the teacher. Moreover, training the teacher requires ground-truth future semantic labels, which may be costly to obtain and limit the applicability of the framework to datasets with such annotations. Future work will investigate more effective privileged teacher learning and explore alternative sources of future supervision, including self-supervised learning from unlabeled driving sequences. These directions may reduce reliance on privileged annotations while preserving the planning relevance of the learned future representations.



Figure A1: Additional qualitative results of WMKD across six driving scenes. For each scene, we visualize the camera input, teacher code, predicted code, ground-truth (GT) future BEV segmentation sequence, and predicted future BEV sequence over three timesteps. The examples cover different road layouts and traffic configurations, illustrating how the predicted representations capture scene structure and the evolution of surrounding vehicles. Trajectories are shown in orange for WMKD and green for ground truth.