WORLDOCC: A REAL-TIME OCCUPANCY SIMULATOR TOWARDS CLOSED-LOOP EVALUATION WITH DIVERSE AND LEARNABLE TRAFFIC BEHAVIORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Closed-loop simulation for autonomous driving has garnered widespread attention in recent years, yet, to our knowledge, no occupancy simulator is at once real-time, high-fidelity, long-horizon, and able to exercise diverse, learnable traffic. This shortfall stems largely from the prevailing practice of building occupancy simulators on a learned world model, which entangles scene evolution with the decision making of traffic participants (i.e., actors). We therefore rethink *whether an occupancy simulator must rely on a learned world model*, and propose WorldOcc, an occupancy simulator that is fast, faithful, and sustained over long rollouts, without a learned world model. At its core is a *policy-scene disentangled architecture* that models the scene transition and the actor policy independently, so each serves its own end, the transition for fidelity and the policy for diversity and flexibility. For high fidelity over a long horizon, we design a *lossless rigid-operator transition*, Ego Warp-Recenter and Actor Erase-Print, that advances the occupancy state with bounded errors. For diverse traffic behaviors, we tailor a *plug-and-play actor policy network* that learns and deploys behaviors, from normal to adversarial, effectively inside the simulator. WorldOcc simulates at about 1270 FPS on an RTX 4090 GPU while holding IoU $\geq 70\%$ over 16 s. Beyond simulation, we introduce WorldOcc-Bench, to our knowledge the first closed-loop benchmark for occupancy planners with diverse traffic, benchmarking six recent planner families. WorldOcc not only reveals the vulnerability of occupancy planners under closed-loop conditions with diverse traffic, but also shows potential as an effective post-training environment.

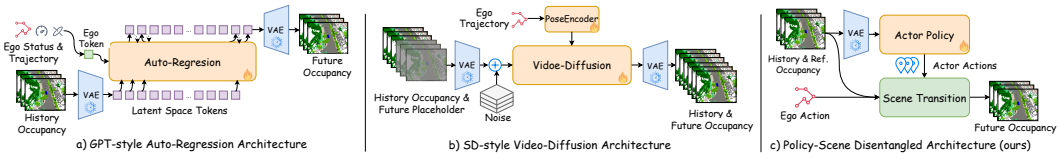


Figure 1: Occupancy simulators based on the learned world models, in (a) GPT-style (Radford et al., 2018) autoregressive or (b) Stable Diffusion-style (Rombach et al., 2022) video-diffusion form. Both entangle scene evolution and actor decision. In contrast, (c) WorldOcc is a *policy-scene disentangled architecture*, with a lossless rigid-operator transition for the high-fidelity, long-horizon scene evolution and a plug-and-play actor policy network for the diverse, learnable traffic behaviors.

1 INTRODUCTION

Closed-loop simulation for autonomous driving has drawn growing attention in recent years. Open-loop metrics such as L2 and collision are efficient but correlate poorly with closed-loop metrics, while closed-loop evaluation remains difficult to scale. PDMS (Dauner et al., 2024) was therefore proposed as a pseudo-closed-loop metric and is now widely adopted (Zhang et al., 2025). However, a recent study (Wang et al., 2026b) finds that PDMS does not always rank planners consistently with

Code and webpage are available at: `https://anonymous.4open.science/r/page-14F1`

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.

closed-loop evaluation, so real-time closed-loop evaluation remains essential. To obtain closed-loop feedback, the community invests heavily in simulators. Generative simulators, such as DrivingSphere (Yan et al., 2025) and DriveArena (Yang et al., 2025), synthesize the future scene, and reconstruction simulators, such as UniSim (Yang et al., 2023) and HUGSIM (Zhou et al., 2026), re-render a recorded scene from a driven pose. Meanwhile, semantic occupancy is a compelling representation for autonomous driving, as it captures fine-grained 3D structure, can be obtained economically from sparse LiDAR, and adapts to both vision and LiDAR modalities (Tian et al., 2023; Zheng et al., 2024). Yet, to the best of our knowledge, no occupancy-modality closed-loop simulator is at once real-time, high-fidelity, long-horizon, and able to exercise diverse, learnable traffic.

Most prior work builds occupancy simulators with a learned world model that forecasts the next occupancy under an ego condition, mainly as a GPT-style autoregressive model such as I^2-World (Liao et al., 2025b) and OccWorld (Zheng et al., 2024), or a Stable Diffusion-style video-diffusion model such as DOME (Gu et al., 2024) and OccSora (Wang et al., 2026a) (Figure 1). The learned world model entangles scene evolution with the decision making of traffic participants, leaving three limitations. First, the generated future is *low-fidelity*, with the fidelity metrics IoU/mIoU measuring no higher than 50% so far. Second, it *decays temporally*, as the reported horizons span only about 3 s (Liao et al., 2025b), too brief for closed-loop simulations. Third, *traffic behaviors lack diversity and flexibility*, because an actor's behavior is fixed to the recording and a user cannot benchmark against diverse traffic or re-drive a participant into a safety-critical interaction. We therefore ask a more fundamental question: *Must an occupancy simulator rely on a learned world model?*

To answer this, we propose WorldOcc, a closed-loop occupancy simulator that is fast, faithful, and sustained over long rollouts, without a learned world model. At its core is a *policy-scene disentangled architecture* that models the scene transition and the actor policy independently, so the transition can be optimal for fidelity while the policy is for diversity and flexibility. For high fidelity over a long horizon, we design a *lossless rigid-operator transition* comprising Ego Warp-Recenter and Actor Erase-Print, which advances the occupancy state with bounded errors. For diverse traffic behaviors, we tailor a *plug-and-play actor policy network* that learns and deploys behaviors, from normal to adversarial, effectively inside our simulator. Harnessing our simulator, we take a step towards closed-loop evaluation with diverse traffic, proposing WorldOcc-Bench on Occ3D-nuScenes (Tian et al., 2023), to our knowledge the first closed-loop benchmark for occupancy planners. Together, WorldOcc not only reveals the vulnerability of occupancy planners under closed-loop conditions with diverse traffic, but also shows potential as an effective post-training environment.

Our contributions are as follows.

- We introduce WorldOcc, to the best of our knowledge the first occupancy-modality closed-loop simulator that is at once real-time, high-fidelity, and long-horizon, with diverse, learnable traffic behaviors, based on a *policy-scene disentangled architecture* that separates scene evolution from actor behavior, without relying on a learned world model.
- We design a *lossless rigid-operator transition* comprising Ego Warp-Recenter and Actor Erase-Print, that achieves nearly lossless transition of occupancy states and holds fidelity over a long horizon. We further devise a *plug-and-play actor policy network* to learn and deploy diverse traffic behaviors effectively inside our simulator.
- We present WorldOcc-Bench, to our knowledge the first closed-loop benchmark for occupancy planners with diverse traffic, benchmarking six recent planner families. Together, WorldOcc not only reveals the vulnerability of occupancy planners under closed-loop conditions with diverse traffic, but also shows potential as an effective post-training environment.
- Extensive experiments show that WorldOcc simulates in real time with high fidelity (IoU $\geq 70\%$) over 16 s on the challenging Occ3D datasets, and that our actor policy network provides realistic and diverse traffic behaviors for our benchmark.

2 Related Work

2.1 Closed-Loop Simulators and Benchmarks for Autonomous Driving

Closed-loop Simulators. Reconstruction simulators re-render recorded scenes into realistic observations but expose no learnable actor policy (UniSim (Yang et al., 2023), HUGSIM (Zhou et al.,

2026)). Game engine simulators such as CARLA (Dosovitskiy et al., 2017) provide reactive actors but synthetic observations, and generative simulators such as DrivingSphere (Yan et al., 2025) and DriveArena (Yang et al., 2025) synthesize realistic scenes with flexible controls but not in real-time. WorldOcc bridges the gap as a real-time closed-loop occupancy simulator with diverse, learnable traffic behaviors. *Benchmarks of Planners.* CARLA-based suites (Bench2Drive (Jia et al., 2024), DriveE2E (Yu et al., 2025)) are closed-loop but keep the synthetic image modality. NAVSIM (Dauner et al., 2024) replays recordings but, as a short-horizon non-reactive pseudo-simulation, cannot exercise the ego's responses to traffic. NuPlan-R (Peng et al., 2026) builds learned reactive traffic but only imitates recorded behaviors. In contrast, WorldOcc's plug-and-play actor policy can learn diverse traffic behaviors inside the simulator. Table 1 compares WorldOcc with prior simulators and with closed-loop planning benchmarks along five distinct axes.

Table 1: **Comparison with closed-loop simulators and planning benchmarks.** WorldOcc is a real-time occupancy simulator that supports diverse, learnable traffic behaviors.

Method	Modality	Real-world	Long horizon	Real-time	Learnable Traffic Behavior
CARLA (Dosovitskiy et al., 2017)	Image & LiDAR	×	✓	✓	✓
UniSim (Yang et al., 2023)	Image & LiDAR	✓	✓	×	×
NAVSIM (Dauner et al., 2024)	Image	✓	×	✓	×
DriveArena (Yang et al., 2025)	Image	✓	✓	×	×
DrivingSphere (Yan et al., 2025)	Image & Occ.	✓	✓	×	×
HUGSIM (Zhou et al., 2026)	Image	✓	✓	✓	×
nuPlan-R (Peng et al., 2026)	Map & State	✓	✓	✓	✓
Bench2Drive (Jia et al., 2024)	Image & LiDAR	×	✓	✓	×
DriveE2E (Yu et al., 2025)	Image	×	✓	✓	×
WorldOcc (ours)	Occ.	✓	✓	✓	✓

2.2 Occupancy World Model for Planning, Simulation, and Scene Generation

Occupancy world models predict the evolution of the semantic scene, and mainly fall into three lines based on their downstream applications. The planning line couples the forecast with driving, giving the ego a predictive future to plan against (OccWorld (Zheng et al., 2024), GEM (Chen et al., 2026), OccTENS (Jin et al., 2026)); some additionally involve scene understanding and question answering (OccLLaMA (Wei et al., 2024), Occ-LLM (Xu et al., 2025)). The simulation line synthesizes future occupancy under an ego condition (DOME (Gu et al., 2024), I^2-World (Liao et al., 2025b), OccSora (Wang et al., 2026a)), and the scene generation line generates occupancy data from layouts (AnyScene (Zhang et al., 2026), UniScene (Li et al., 2025)), from a one-frame static map (OccSim (Liu et al., 2026)), or as large-scale 4D scenes (DynamicCity (Bian et al., 2025)). WorldOcc is an occupancy simulator but *not a learned world model*, providing a closed-loop benchmark for planners while the underlying occupancy data can be supplied by scene generators, bridging the two lines.

3 Method

3.1 Policy-Scene Disentangled Architecture

Prior world model simulators couple the static scene and the dynamic actors into a single learned model, making both sub-optimal. To address this, we design a *policy-scene disentangled architecture* (Figure 2b). The occupancy scene is advanced by the *lossless rigid operator transition* of Section 3.2, while the controlled traffic participants are governed by a plug-and-play actor policy network of Section 3.3. First, we formulate the closed-loop simulation in occupancy as a partially observable Markov decision process (POMDP) (Kaelbling et al., 1998), $\langle S, A, T, R, \Omega, O, \gamma \rangle$, with state $s_t \in S$, action $a_t \in A$, transition T, reward function R with per-agent reward $r_t^{(i)} \in \mathbb{R}$, observation $o_t \in \Omega$, observation function O, and discount γ. At time t the closed-loop state and an agent's observation are

$$s_t = (X_t, E_t, G_t), \qquad o_t = O(s_t) = (B_t, \eta_t), \tag{1}$$

where $X_t \in [0, C)^{H \times W \times Z}$ is the ego-centric semantic occupancy scene of $H \times W \times Z$ voxels of cell size ℓ, with C classes, B_t the agent's observation of X_t, and η_t the agent's own state ($\eta_t = E_t$ for the ego, $\eta_t = G_t$ for an actor). E_t and $G_t = \{G_t^{(i)}\}_{i=1}^{n}$ are the ego and collective actor states,

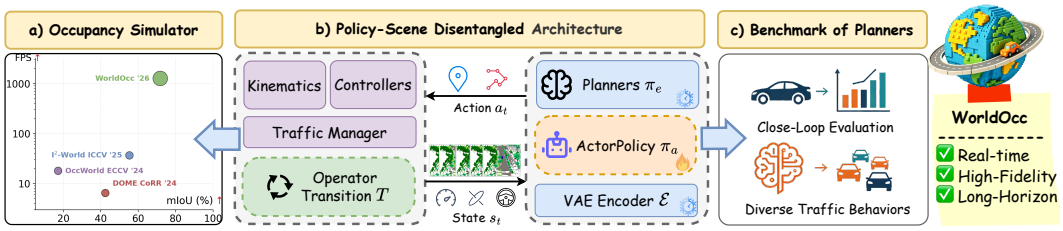


Figure 2: **The WorldOcc simulator.** (a) Occupancy fidelity (mIoU) vs. throughput (FPS) on Occ3D-nuScenes, measured on an RTX 4090, with bubble area encoding the supported simulation horizon. (b) Our policy-scene disentangled architecture (Section 3.1). (c) We introduce a benchmark of planners with closed-loop evaluation and diverse traffic behaviors based on WorldOcc.

respectively (Appendix B). The ego and the actors act through their policies π_e and π_a,

$$a_t^e = \pi_e(o_t) = (w_{t+1}, \ldots, w_{t+6}) \in \mathbb{R}^{2\times 6}, \qquad a_t = \pi_a(o_t) = \{a_t^{(i)} \in \mathbb{R}^2\}_{i=1}^n, \tag{2}$$

where π_e is an *external* planner under evaluation and π_a a *plug-in* actor policy that the user can set (Section 3.3). The ego action $a_t^e = (w_{t+1}, \ldots, w_{t+6}) \in \mathbb{R}^{2\times 6}$ is a six-waypoint trajectory with each waypoint $w_{t+1} \in \mathbb{R}^2$, following the planning protocol of (Zheng et al., 2024), and the collective actor action $a_t = \{a_t^{(i)}\}_{i=1}^n$ collects each actor's aim point $a_t^{(i)} \in \mathbb{R}^2$. Unlike prior work (Gu et al., 2024; Liao et al., 2025b) that executes the actions (e.g., planned trajectory or high-level command) directly in the simulator, we add kinematics and controllers for more physically plausible traffic behaviors (Appendix B). The kinematics and controllers execute the actions a_t^e and a_t into driven poses E_{t+1} and G_{t+1}, so the transition depends only on these poses and the current state s_t,

$$s_{t+1} = T(s_t, E_{t+1}, G_{t+1}). \tag{3}$$

We characterize our occupancy simulator by the pair (T, π_a), where we model T and π_a independently, so that T is optimal for fidelity while the policy π_a is for diversity and flexibility.

3.2 Lossless Rigid-Operator Transition

Prior world model-based simulators learn the transition function T_θ from data, but fail to maintain high fidelity over a long horizon. Motivated by NAVSIM (Dauner et al., 2024), which evaluates planners by replaying real-world, recorded scenes, we take the recorded states s'_t and s'_{t+1} as *reference states*. We notice the large *spatial-temporal overlap of scene* (Table 5) and the *bounded trajectory deviation* of planners (Table 16), allowing s'_{t+1} to be a high-fidelity surrogate of the closed-loop current state s_t, so we modify T to operate on s'_{t+1} instead,

$$s_{t+1} = T\big(s'_{t+1}, E_{t+1}, G_{t+1}\big), \tag{4}$$

where s_{t+1} is the closed-loop next state. The reference state $s'_{t+1} = (X'_{t+1}, E'_{t+1}, G'_{t+1})$ provides the fidelity required by a reliable and trustworthy benchmark, while supplying the recorded traffic behavior with no re-driving overhead. The spatial-temporal overlap of scene is a property of occupancy data, while the bounded trajectory deviation is a valid assumption of trained planners, as detailed in Appendix A. We further decouple T into Ego Warp-Recenter and Actor Erase-Print to execute each kind of action.

Ego Warp-Recenter. We align the reference state s'_{t+1} to the driven ego pose. Let E'_t and E'_{t+1} be the reference ego poses at times t and $t{+}1$, with positions p'_t, p'_{t+1} and headings θ'_t, θ'_{t+1}, and let E_{t+1} be the driven ego pose at $t{+}1$, with position p_{t+1} and heading θ_{t+1}. We re-express the driven pose in the frame of the reference state,

$$\bar{p}_{t+1} = \mathrm{Rot}(\theta'_t - \theta'_{t+1})\, p_{t+1} + \mathrm{Rot}(-\theta'_{t+1})(p'_t - p'_{t+1}), \qquad \bar{\varphi} = \theta_{t+1} - (\theta'_{t+1} - \theta'_t), \tag{5}$$

where $\mathrm{Rot}(\alpha)$ is a rotation matrix for angle α, so $P = (\bar{p}_{t+1}, \bar{\varphi})$ is the driven pose re-expressed in the frame of the reference state. Each reference-frame position c' maps to the re-centred cell centre c and its class is placed at the nearest re-centred cell, $\lfloor \frac{1}{\ell} c + \frac{1}{2} \rfloor$. The Recenter operator applies this mapping to every reference cell, producing the closed-loop next occupancy state X_{t+1},

$$c = \mathrm{Rot}(-\bar{\varphi})(c' - \bar{p}_{t+1}), \qquad X_{t+1} = \mathrm{Recenter}(X'_{t+1};\, P), \tag{6}$$

Positions that map outside the re-centred grid are discarded (Algorithm 1). Unlike latent-space warping, such as (Liu et al., 2026), our operator can be viewed as an occupancy-space warping.

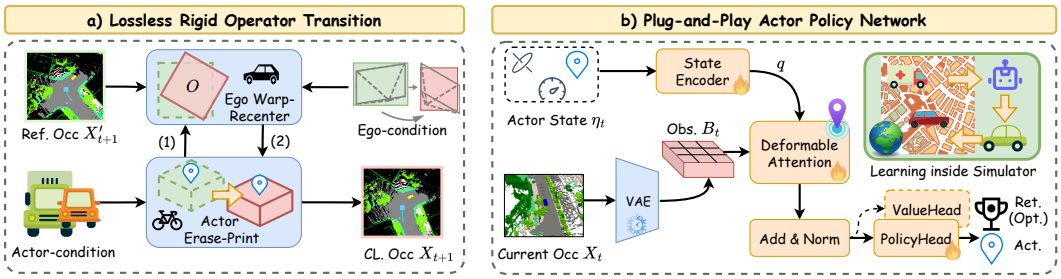


Figure 3: (a) The *lossless rigid-operator transition* comprising Ego Warp-Recenter and Actor Erase-Print using the next recorded state as a reference state, for reliable and trustworthy benchmarking. (b) The *plug-and-play actor policy network* refines the actor query q with a localized scene context from the observation B_t, and learns diverse traffic behaviors effectively inside our simulator.

Actor Erase-Print. We erase a real scene actor's reference body and re-drive it under a novel behavior. For a controlled actor i, we set the voxels inside its recorded bounding box at pose $G'^{(i)}_{t+1}$ to the free class, giving the erased reference-state occupancy $\bar{X}'_{t+1}$. We re-express the actor's driven pose $G^{(i)}_{t+1} = (p^{(i)}_{t+1}, \theta^{(i)}_{t+1})$ in the driven ego coordinate, analogous to Eq. (5),

$$\bar{G}^{(i)}_{t+1} = \big(\mathrm{Rot}(-\bar{\varphi})(p^{(i)}_{t+1} - \bar{p}_{t+1}),\ \theta^{(i)}_{t+1} - \bar{\varphi}\big), \tag{7}$$

We then advance the erased reference-state occupancy by the Ego Warp-Recenter operator and print the actor's captured voxel body there,

$$X_{t+1} = \mathrm{Recenter}\left(\bar{X}'_{t+1};\ P\right) \cup \Phi\big(\bar{G}^{(i)}_{t+1}\big), \tag{8}$$

where X_{t+1} is the next closed-loop occupancy in the driven ego frame, $\mathrm{Recenter}(\cdot)$ is the operator defined by Eq. (6), and $\Phi(\cdot)$ places the actor's voxel body at the given pose (Algorithm 2).

3.3 Plug-and-Play Actor Policy Network

Unlike prior approaches that predict a scene-policy coupled future, our actor policy is plug-and-play, trained and deployed inside our simulator with no finetuning of a heavy-weight world model.

Actor Policy Network. We first encode the occupancy state into a BEV observation of feature dimension d with a frozen tokenizer (Figure 3b) $\mathcal{E}$, and map each actor's status $G^{(i)}_t$ to an actor query q with an MLP,

$$B_t = \mathcal{E}(X_t) \in \mathbb{R}^{H\times W\times d}, \quad q = \mathrm{MLP}(G^{(i)}_t) \in \mathbb{R}^d, \tag{9}$$

We then refine q with deformable attention (Zhu et al., 2021) over the scene B_t around the actor's current position $p^{(i)}_t$, so each decision depends on the actor's localized perception of the environment,

$$f_t = \mathrm{LayerNorm}\left(q + \mathrm{DeformAttention}(q, B_t; p^{(i)}_t)\right) \in \mathbb{R}^d, \tag{10}$$

The feature f_t feeds a policy head, and optionally a value head. The action space is the normalized next waypoint $a^{(i)}_t \in [-1,1]^2$, unnormalized by the per-class displacement limit. The network weights are shared across actors, so at inference it can drive multiple controlled actors in a batch. To augment our simulator with natural and diverse interactions from traffic participants, we instantiate the actor policy network on normal and adversarial behaviors, detailed as follows.

Normal Traffic Behavior. Given that the underlying recorded data (i.e., nuScenes (Caesar et al., 2020)) is predominantly benign, we obtain a normal behavior by training the actor policy network to imitate recorded traffic (Hussein et al., 2017). We instantiate a deterministic policy head,

$$a^{(i)}_t = \mu_\theta(f_t), \tag{11}$$

supervised by the recorded next pose $a^{\star(i)}_t$ in the current ego frame (Appendix C). We minimize the ℓ_1 behavior-cloning objective

$$\mathcal{L}_{\mathrm{L1}}(\theta) = \mathbb{E}_t\left[\|\tanh(\mu_\theta(f_t)) - a^{\star(i)}_t\|_1\right], \tag{12}$$

Adversarial Traffic Behavior. Unlike a normal behavior, a safety-critical adversarial behavior is rare in the benign recordings, so imitating it would require collecting extra demonstration dataset. We instead learn the adversarial behavior directly from interactions inside our simulator with deep reinforcement learning. We pair a stochastic policy head with a shared-feature value head,

$$\pi(\cdot \mid o_t) = \mathcal{N}\big(\mu_\theta(f_t),\, \sigma_\theta\big), \qquad V_t = v_\phi(f_t), \tag{13}$$

where $\mu_\theta(f_t)$ and σ_θ are the parameterized Gaussian mean and standard deviation, and $V_t = v_\phi(f_t)$ is the value head's estimate of the expected discounted return $R_t^{(i)} = \sum_{k=0}^{\infty} \gamma^k r_{t+k}^{(i)}$. We design an adversarial reward, consisting of a large reward on colliding with the ego, a penalty for leaving the sensing range, a distance-progress shaping toward the ego, and a blocking penalty,

$$r_t^{(i)} = \begin{cases} r_{\text{coll}}, & \text{the driven actor bbox intersects the ego (terminal),} \\ r_{\text{out}}, & \text{the actor leaves the sensing range (terminal),} \\ \lambda\,(d_t^{(i)} - d_{t+1}^{(i)}) - \kappa\,\beta_t^{(i)}, & \text{otherwise,} \end{cases} \tag{14}$$

where $d_t^{(i)} = \|p_t^{(i)}\|$ is the actor-to-ego distance, $r_{\text{coll}} > 0$ and $r_{\text{out}} < 0$ are the terminal collision reward and out-of-range penalty, the progress term $\lambda(d_t^{(i)} - d_{t+1}^{(i)})$ rewards closing the distance with the ego, $\beta_t^{(i)} \in [0, 1]$ is the safety-braked fraction of the frame, and λ, κ weight progress and blocking. We optimize with Proximal Policy Optimization (PPO) (Schulman et al., 2017).

3.4 Closed-Loop Benchmark with Diverse Traffic

Closed-loop evaluation with diverse traffic is essential for autonomous driving (Jia et al., 2024; Yu et al., 2025; Li et al., 2026). We design *WorldOcc-Bench* on the Occ3D-nuScenes (Tian et al., 2023), a closed-loop benchmark in the occupancy modality with diverse traffic. It evaluates a planner under three traffic settings, namely (1) *recorded*, where actors follow their recorded motion, (2) *normal*, where the actors are re-driven by the normal actor policy, and (3) *adversarial*, where actors are re-driven into safety-critical interactions with the adversarial actor policy.

Traffic Manager. We tailor a *range-aware spatial-temporal traffic manager* that selects which actors to re-drive. Motivated by nuPlan-R (Peng et al., 2026), we sample actors proximal to the ego with a per-class reachability range, and randomize the frame where the policy starts to control them, so re-driven actors span different types, positions, and timings. We design two schedules: *spatial* installs one batch of k actors at a single start frame for concurrent pressure, while *temporal* staggers k actors over independently drawn start frames for a sequence of separate encounters. Sweeping k along a per-schedule ladder exposes the actor load at which each planner degrades.

Multi-Scenario Scoring. Because WorldOcc runs in real time, we can evaluate a planner on thousands of scenarios within a reasonable time. We therefore introduce a *multi-scenario* closed-loop score, *WorldOcc-Score*, summarizing each planner's performance under each setting and schedule. Within a scenario, PDMS (Dauner et al., 2024) averages a per-step component product over the N driven frames of that scenario,

$$\text{PDMS} = \frac{1}{N} \sum_t \Big(\prod_{m \in \{\text{NC,DAC}\}} m_t \Big) \cdot \frac{\sum_{m \in \{\text{TTC,C}\}} w_m\, m_t}{\sum_{m \in \{\text{TTC,C}\}} w_m}, \tag{15}$$

where no collision (NC), drivable-area compliance (DAC), time-to-collision (TTC), and comfort (C) score safety and ride quality, with w_m the weight of component m in the weighted average over TTC and C, and route completion (RC) is the fraction of the route driven. Following (Zhou et al., 2026), we gate the PDMS to prevent short driven routes from dominating the score, and aggregate over all scenarios $\mathcal{Q}_\sigma$ and schedules $\mathcal{S}$ within a setting,

$$\text{WorldOcc-Score}_S = \frac{1}{|\mathcal{S}|} \sum_{\sigma \in \mathcal{S}} \frac{1}{|\mathcal{Q}_\sigma|} \sum_{j \in \mathcal{Q}_\sigma} \text{PDMS}_j \cdot \text{RC}_j, \tag{16}$$

where $S \in \mathcal{T} = \{\text{recorded}, \text{normal}, \text{adversarial}\}$ is the traffic setting, $\mathcal{S}$ its schedules, and $\mathcal{Q}_\sigma$ the scenario cells under schedule σ. The recorded traffic setting re-drives no actor, so $\mathcal{S} = \{\text{recorded}\}$, whereas the normal and adversarial settings use both schedules, $\mathcal{S} = \{\text{spatial}, \text{temporal}\}$. We report the score under each traffic setting, and their mean as an overall score.

Table 2: Occupancy fidelity on **Occ3D-nuScenes** in IoU/mIoU ↑(%). †Reported from DynamicCity's conditional model on the Occ3D-nuScenes dataset.

Method	Year	1 s	2 s	3 s	8 s	16 s	Avg	FPS
DOME (Gu et al., 2024)	2024	43.99/35.11	35.36/25.89	29.74/20.29	–	–	36.36/27.10	6.5
OccWorld (Zheng et al., 2024)	2024	34.63/25.78	25.07/15.14	20.18/10.51	–	–	26.63/17.14	18.0
I^2-World (Liao et al., 2025b)	2025	54.29/47.62	49.43/38.58	45.69/32.98	–	–	49.80/39.73	37.0
DynamicCity† (Bian et al., 2025)	2025	34.12/26.18	25.82/16.94	–	–	–	30.0/21.6	–
WorldOcc (ours)	2026	73.4/72.4	75.0/72.7	77.0/77.0	76.3/72.7	73.9/70.5	**75.3/74.0**	**1270**

4 Experiment Results

4.1 Experimental Setup

Datasets. We evaluate our simulator on the Occ3D-nuScenes and Occ3D-Waymo datasets (Tian et al., 2023). **Simulator evaluation.** We compare occupancy fidelity in IoU/mIoU and throughput in FPS following (Zheng et al., 2024), over the 1–3 s horizon the baselines cover, grouping them into two families, autoregressive I^2-World (Liao et al., 2025b) and OccWorld (Zheng et al., 2024), and video-diffusion DOME (Gu et al., 2024) and DynamicCity (Bian et al., 2025). **Planner benchmarking.** We benchmark occupancy-based and recent end-to-end planners adapted to occupancy via the frozen I^2-tokenizer (Liao et al., 2025b), namely GT-Following, OccWorld (Zheng et al., 2024), Ego-MLP, Go-Straight, the BEV-Planner family (Li et al., 2024), DiffusionDrive (Liao et al., 2025a), ResWorld (Zhang et al., 2025), SSR (Li & Cui, 2025), and WorldEngine (BC/RLFT) (Li et al., 2026). We also report reproduced open-loop scores following (Li et al., 2024).

4.2 Implementation Details

Our rigid-operator transition runs on both a numba-JIT[1] CPU and a CUDA GPU. The ego plan is solved by an iLQR (Karnchanachari et al., 2024) once per occupancy frame and executed by the bicycle model at a fixed 20 Hz; actors are driven by an IDM (Treiber et al., 2000)-inspired waypoint-following controller over a unicycle model. Appendix B gives simulator details, Appendix C the actor-policy training details, and Appendix D the benchmark details.

4.3 A Real-Time, Long-Horizon, and High-Fidelity Simulator

Long-horizon fidelity. We first measure long-horizon fidelity under recorded ego actions in Table 2. WorldOcc holds IoU between 73% and 80%, flat across the whole 16 s horizon, whereas the learned world models decay over the 1–3 s they cover and stay far lower, with I^2-World dropping from 54% to 46% and DOME from 44% to 30%. We present our results on Occ3D-Waymo (Table 3). There we compare only with I^2-World, as DOME and OccWorld report no Occ3D-Waymo results or checkpoints and DynamicCity's conditional model is Occ3D-nuScenes only, so none of the three matches our protocol. WorldOcc holds 92.6–93.5% IoU across the whole 16 s horizon at the native 10 Hz, against I^2-World's zero-shot inference of 57.6–64.6% over the 1–3 s it covers. The gap is similar under the 2 Hz protocol (93.3–93.6% vs. 46.9–58.3%). **Avg** is the whole 16 s horizon for WorldOcc and the 1–3 s mean for I^2-World. We also compare the teacher-forcing setting, the perception fidelity following (Lee et al., 2024), and the counterfactual-maneuver fidelity in Appendix B, showing the high-fidelity property of WorldOcc comprehensively.

Throughput. On an RTX 4090, WorldOcc advances the occupancy state at about 1270 FPS on GPU, over thirty times I^2-World's reported 37.0 FPS, over seventy times OccWorld's 18.0 FPS, and about two hundred times DOME's 6.5 FPS. Even on a CPU, WorldOcc still achieves 436 FPS.

Learned Actor Policies. We evaluate the learned actor policies under the GT-Following ego against two scripted behaviors, namely (1) *normal scripted*, which re-drives each actor along its recording, and (2) *aggressive scripted*, which chases the ego, over five seeds (Fig. 4). The *adversarial* policy's ego collision rate grows with the number of controlled actors k, reaching 0.46 spatially and 0.74 temporally at k=16, above both scripted baselines, while the *normal* policy's rate of 0.125 ± 0.067

[1] https://numba.pydata.org/

Table 3: Occupancy fidelity on **Occ3D-Waymo** in IoU/mIoU ↑(%). †Reported from I^2-World's zero-shot evaluation. DOME and OccWorld report no Occ3D-Waymo results or checkpoints, and DynamicCity's conditional model is Occ3D-nuScenes only; all three are thus omitted.

Method	**Freq**	1 s	2 s	3 s	8 s	16 s	**Avg**
I^2-World† (Liao et al., 2025b)	2 Hz	58.29/42.84	51.84/35.58	46.94/30.72	–	–	52.36/36.38
WorldOcc (ours)	2 Hz	93.40/95.09	93.29/94.78	93.56/94.68	94.73/95.52	95.44/95.29	**95.44/95.29**
I^2-World† (Liao et al., 2025b)	10 Hz	64.58/49.35	60.71/43.03	57.63/38.99	–	–	60.97/43.73
WorldOcc (ours)	10 Hz	93.47/93.60	92.98/92.69	92.59/91.92	93.02/92.27	92.84/91.65	**92.84/91.65**

matches the scripted re-drive's 0.111 ± 0.058, confirming that the actor policy network can learn distinctive behaviors. We further report the naturalness metrics, tracking accuracy of the normal policy, and training compute in Appendix C, showing efficiency and effectiveness of our policies.

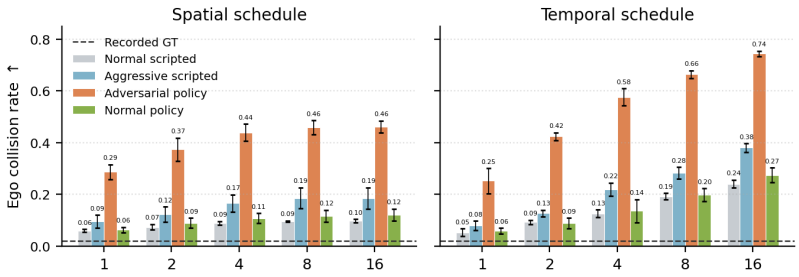


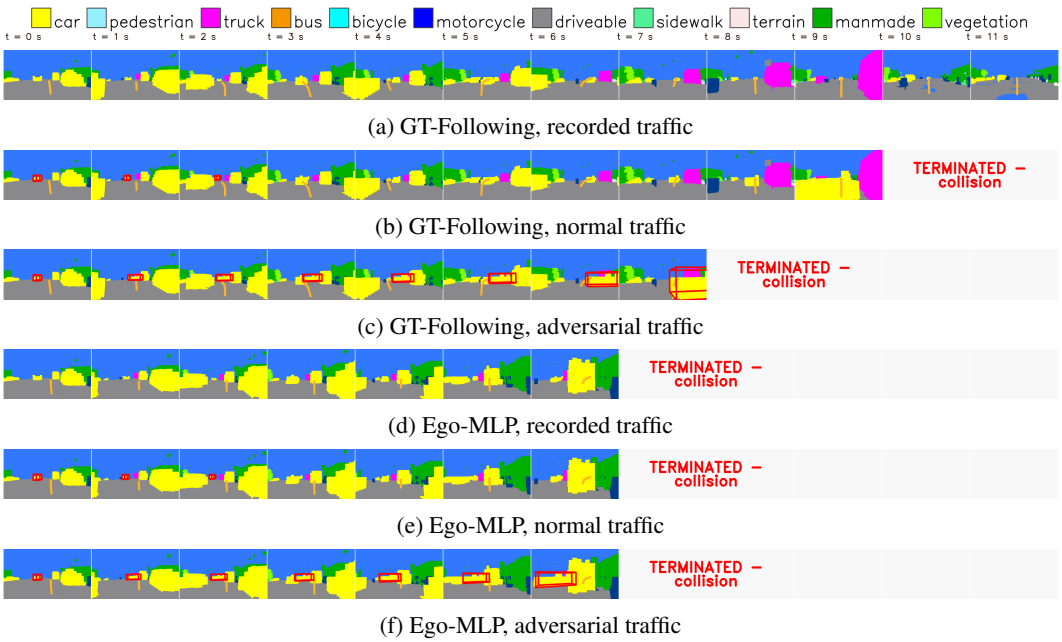
Figure 4: Ego collision rate vs. the number of controlled actors k, under the spatial and temporal schedules. The dashed line is the recorded GT, which is invariant to the number of controlled actors.

4.4 The WorldOcc-Bench Leaderboard

Open-Loop vs. Closed-Loop Scores. Table 4 reports WorldOcc-Score across planners. Every planner falls well short of GT-Following, and the recorded-setting scores are nearly uncorrelated with open-loop ADE (Pearson $r = +0.14$) and collision rate ($r = +0.12$, Fig. 9), confirming that open-loop metrics misrank planners (Li et al., 2024; Dauner et al., 2024). ResWorld reproduces both reported open-loop variants (L2 0.35/0.32 m) yet collapses in closed loop (0.034/0.049), and Ego-MLP is near-best by open-loop ADE (0.36 m) yet scores 0.183. The recorded traffic setting alone is already a challenge from self-conditioning covariate shift, since each planner is fed its own *driven* occupancy, leading to a poor RC of 0.07–0.33. Overall, the benchmark presents a substantial challenge to current end-to-end and occupancy-based planners.

Vulnerability under Diverse Traffic. We further expose planners' vulnerability under diverse traffic behaviors. The *normal* setting uses the learned behavior-cloned policy, which imitates recorded motion, so most planners stay close to their recorded score, with the perception-free Ego-MLP moving only $0.244 \rightarrow 0.241$. The *adversarial* setting, driven by the reward-tuned policy, degrades nearly every planner. The occupancy-native OccWorld drops under adversarial ($0.193 \rightarrow 0.191 \rightarrow 0.184$), so its failure is driven-occupancy drift rather than a modality mismatch, while the perception-free Ego-MLP collapses only under adversarial ($0.244 \rightarrow 0.241 \rightarrow 0.065$). Even the GT-Following oracle falls from 0.864 to 0.804 to 0.753, so the traffic settings, not self-conditioning, drive the degradation. Figure 5 shows the planner-by-traffic matrix for one scene in the front-camera semantic view, where the GT-Following oracle completes only the recorded clip and collides under both normal and adversarial traffic, while Ego-MLP collides in all three settings, mirroring the scores of Table 4.

Post-Training Environment. Beyond evaluation, the simulator also serves as a post-training environment. We port WorldEngine's trajectory-vocabulary planner (Li et al., 2026) and follow their Reinforcement Learning Fine-Tuning (RLFT) recipe to fine-tune its scoring head. The planner scores a fixed 4096-candidate vocabulary against the occupancy latent, and one-step RLFT lifts the recorded score from 0.273 to 0.293, the best after the GT-Following oracle. We further study



(a) GT-Following, recorded traffic

(b) GT-Following, normal traffic

(c) GT-Following, adversarial traffic

(d) Ego-MLP, recorded traffic

(e) Ego-MLP, normal traffic

(f) Ego-MLP, adversarial traffic

Figure 5: Planner-by-traffic matrix: rollouts of one scene in the front-camera semantic view with the ego plan overlaid, showing the GT-Following oracle and Ego-MLP under recorded, normal, and adversarial traffic. The re-driven actor is outlined in red. Rows that terminate abnormally end with their termination reason, while normally-terminating rows run the full clip with no banner.

multi-step RLFT beyond WorldEngine's setting and find the WorldOcc-Score peaks at 0.309 with two steps (Appendix D); long-horizon RLFT on a driving policy is left to future work.

Table 4: WorldOcc-Bench results with reproduced open-loop results. †GT-Following is a human reference rather than a learned planner.

	PDMS subscores					WorldOcc-Score				Open-loop	
Baseline	**NC** ↑	**DAC** ↑	**TTC** ↑	**Comf.** ↑	**RC** ↑	**Recorded** ↑	**Normal** ↑	**Adversarial** ↑	**Overall** ↑	**L2 (m)** ↓	**Coll. (%)** ↓
GT-Following†	0.992	0.913	0.991	0.871	0.937	0.864	0.804	0.753	0.807	0.00	0.00
OccWorld (Zheng et al., 2024)	0.953	0.875	0.946	0.740	0.254	0.193	0.191	0.184	0.189	0.63	3.18
Ego-MLP (Li et al., 2024)	0.966	0.781	0.957	0.910	0.286	0.244	0.241	0.065	0.183	0.36	0.56
Go-Straight (Li et al., 2024)	0.963	0.793	0.962	1.000	0.254	0.227	0.224	0.221	0.224	0.55	1.05
BEV-Planner+ (Li et al., 2024)	0.963	0.788	0.957	0.865	0.304	0.251	0.246	0.241	0.246	0.34	0.50
BEV-Planner++ (Li et al., 2024)	0.966	0.797	0.961	0.930	0.296	0.252	0.249	0.242	0.248	0.37	0.78
DiffusionDrive (Liao et al., 2025a)	0.944	0.792	0.937	0.797	0.288	0.249	0.250	0.228	0.242	0.32	0.42
ResWorld (Zhang et al., 2025)	0.922	0.696	0.905	0.610	0.070	0.034	0.035	0.035	0.034	0.35	0.40
ResWorld♢ (Zhang et al., 2025)	0.905	0.715	0.886	0.506	0.093	0.057	0.045	0.045	0.049	0.32	0.38
SSR (Li & Cui, 2025)	0.959	0.761	0.952	0.790	0.255	0.214	0.212	0.196	0.207	0.34	0.57
WorldEngine (BC) (Li et al., 2026)	0.941	0.874	0.930	0.791	0.312	0.273	0.268	0.247	0.263	—	—
WorldEngine (RLFT) (Li et al., 2026)	**0.976**	**0.970**	**0.969**	0.960	**0.331**	**0.293**	**0.287**	**0.263**	**0.281**	—	—

5 Conclusion

In this paper, we introduced WorldOcc, a real-time, high-fidelity, long-horizon occupancy simulator that avoids a learned world model. Our policy-scene disentangled architecture separates scene evolution from actor behavior, so the transition serves fidelity and the policy serves diversity. Harnessing it, we propose WorldOcc-Bench, a closed-loop benchmark for occupancy planners with diverse traffic, exposing planner vulnerability and showing potential as an effective post-training environment.

Limitations and Future Work. WorldOcc is a simulator, not a learned world model, so it cannot synthesize scenes beyond the recording. Future work will combine it with occupancy scene generators (e.g., OccSim, AnyScene), extend RLFT of planners to longer horizons, and learn more realistic, long-tail traffic behaviors, such as cutting-in, emergency braking, taking over, and running a red light.

AI USE STATEMENT

We used generative AI tools only to draft and polish the text. Experiments, data, and results were produced by our own code, and we take responsibility for all content.

REPRODUCIBILITY STATEMENT

All experiments use public datasets. The operators, kinematics, controllers, benchmark protocol, and metrics are specified in the main text and appendices. Planners use released checkpoints, and code will be released upon publication.

ETHICS STATEMENT

This work uses the public Occ3D dataset and the published planners, simulators. No human subjects or personal data are involved, and the simulator is used only for offline evaluation and RL training.

REFERENCES

Hengwei Bian, Lingdong Kong, Haozhe Xie, Liang Pan, Yu Qiao, and Ziwei Liu. Dynamiccity: Large-scale 4d occupancy generation from dynamic scenes. In *International Conference on Learning Representations (ICLR)*, 2025.

Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pp. 11618–11628. Computer Vision Foundation / IEEE, 2020.

Cheng Chen, Hao Huang, and Saurabh Bagchi. Gem: Gaussian evolution model for occupancy forecasting and motion planning. *arXiv preprint arXiv:2605.17682*, 2026.

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

Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio López, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pp. 1–16. PMLR, 2017.

Songen Gu, Wei Yin, Bu Jin, Xiaoyang Guo, Junming Wang, Haodong Li, Qian Zhang, and Xiaoxiao Long. Dome: Taming diffusion model into high-fidelity controllable occupancy world model. *CoRR*, abs/2410.10429, 2024.

Ahmed Hussein, Mohamed Medhat Gaber, Eyad Elyan, and Chrisina Jayne. Imitation learning: A survey of learning methods. *ACM Computing Surveys*, 50(2):1–35, 2017.

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

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 *IEEE/CVF International Conference on Computer Vision, ICCV 2023*, pp. 8306–8316. IEEE, 2023.

Bu Jin, Songen Gu, Xiaotao Hu, Yupeng Zheng, Xiaoyang Guo, Qian Zhang, Xiaoxiao Long, and Wei Yin. Occtens: 3d occupancy world model via temporal next-scale prediction. *IEEE Robotics and Automation Letters*, 11(3):3566–3573, 2026.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2):99–134, 1998.

Napat Karnchanachari, Dimitris Geromichalos, Kok Seang Tan, Nanxiang Li, Christopher Eriksen, Shakiba Yaghoubi, Noushin Mehdipour, Gianmarco Bernasconi, Whye Kit Fong, Yiluan Guo, and Holger Caesar. Towards learning-based planning: The nuplan benchmark for real-world autonomous driving. In *IEEE International Conference on Robotics and Automation, ICRA 2024*, pp. 629–636. IEEE, 2024.

Jumin Lee, Sebin Lee, Changho Jo, Woobin Im, Juhyeong Seon, and Sung-Eui Yoon. Semcity: Semantic scene generation with triplane diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Bohan Li, Jiazhe Guo, Hongsi Liu, Yingshuang Zou, Yikang Ding, Xiwu Chen, Hu Zhu, Feiyang Tan, Chi Zhang, Tiancai Wang, Shuchang Zhou, Li Zhang, Xiaojuan Qi, Hao Zhao, Mu Yang, Wenjun Zeng, and Xin Jin. Uniscene: Unified occupancy-centric driving scene generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025*, pp. 11971–11981. Computer Vision Foundation / IEEE, 2025.

Peidong Li and Dixiao Cui. Navigation-guided sparse scene representation for end-to-end autonomous driving. In *International Conference on Learning Representations (ICLR)*, 2025.

Tianyu Li, Li Chen, Caojun Wang, Haochen Liu, Kashyap Chitta, Zhenjie Yang, Yuhang Lu, Naisheng Ye, Yihang Qiu, Yufei Wang, Luoxi Zou, Jiaxin Peng, Jin Pan, Zhaoyu Su, Andrei Bursuc, Shengbo Eben Li, Andreas Geiger, Peng Su, and Hongyang Li. World engine: Towards the era of post-training for autonomous driving, 2026.

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, CVPR 2024*, 2024.

Bencheng Liao, Shaoyu Chen, Haoran Yin, Bo Jiang, Cheng Wang, Sixu Yan, Xinbang Zhang, Xiangyu Li, Ying Zhang, Qian Zhang, and Xinggang Wang. Diffusiondrive: Truncated diffusion model for end-to-end autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025*, pp. 12037–12047. Computer Vision Foundation / IEEE, 2025a.

Zhimin Liao, Ping Wei, Ruijie Zhang, Shuaijia Chen, Haoxuan Wang, and Ziyang Ren. I2-world: Intra-inter tokenization for efficient dynamic 4d scene forecasting. In *IEEE/CVF International Conference on Computer Vision, ICCV 2025*, pp. 25810–25819. IEEE, 2025b.

Tianran Liu, Shengwen Zhao, Mozhgan Pourkeshavarz, Weican Li, and Nicholas Rhinehart. Occsim: Multi-kilometer simulation with long-horizon occupancy world models. *arXiv preprint arXiv:2603.28887*, 2026.

Mingxing Peng, Ruoyu Yao, Xusen Guo, and Jun Ma. nuplan-r: A closed-loop planning benchmark for autonomous driving via reactive multi-agent simulation. In *IEEE Intelligent Vehicles Symposium (IV)*, pp. 702–709, 2026.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.

Christian Schroeder de Witt, Tarun Gupta, Denys Makoviychuk, Viktor Makoviychuk, Philip H. S. Torr, Mingfei Sun, and Shimon Whiteson. Is independent learning all you need in the starcraft multi-agent challenge? *arXiv preprint arXiv:2011.09533*, 2020.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017.

Xiaoyu Tian, Tao Jiang, Longfei Yun, Yucheng Mao, Huitong Yang, Yue Wang, Yilun Wang, and Hang Zhao. Occ3d: A large-scale 3d occupancy prediction benchmark for autonomous driving. In *Advances in Neural Information Processing Systems 36*, 2023.

Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, 62(2):1805–1824, 2000.

Lening Wang, Wenzhao Zheng, Yilong Ren, Han Jiang, Zhiyong Cui, Haiyang Yu, and Jiwen Lu. Occsora: 4d occupancy generation models as world simulators for autonomous driving. *IEEE Transactions on Image Processing*, 35:4947–4960, 2026a.

Yiru Wang, Anqing Jiang, Shuo Wang, Yuwen Heng, Hai Yang, Yang Chen, and Hao Sun. Do open-loop metrics predict closed-loop driving? a cross-benchmark correlation study of NAVSIM and Bench2Drive. *arXiv preprint arXiv:2605.00066*, 2026b.

Julong Wei, Shanshuai Yuan, Pengfei Li, Qingda Hu, Zhongxue Gan, and Wenchao Ding. Occllama: An occupancy-language-action generative world model for autonomous driving. *arXiv preprint*, 2024.

Huihui Xu et al. Occ-llm: Enhancing autonomous driving with occupancy-based large language models. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

Tianyi Yan, Dongming Wu, Wencheng Han, Junpeng Jiang, Xia Zhou, Kun Zhan, Cheng-zhong Xu, and Jianbing Shen. Drivingsphere: Building a high-fidelity 4d world for closed-loop simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27531–27541, 2025.

Xuemeng Yang, Licheng Wen, Tiantian Wei, Yukai Ma, Jianbiao Mei, Xin Li, Wenjie Lei, Daocheng Fu, Pinlong Cai, Min Dou, Liang He, Yong Liu, Botian Shi, and Yu Qiao. Drivearena: A closed-loop generative simulation platform for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26933–26943, 2025.

Ze Yang, Yun Chen, Jingkang Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. Unisim: A neural closed-loop sensor simulator. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023*, pp. 1389–1399. IEEE, 2023.

Haibao Yu, Wenxian Yang, Ruiyang Hao, Chuanye Wang, Jiaru Zhong, Ping Luo, and Zaiqing Nie. Drivee2e: Closed-loop benchmark for end-to-end autonomous driving through real-to-simulation. *arXiv preprint arXiv:2509.23922*, 2025.

Fengcong Zhang, Yuanming Duan, Xinyu Zhang, Zhaotie Meng, Guoliang Xi, Zhanming Zhou, Yicheng Li, Duo Yin, Huifeng Zhai, and Yang Gao. Resworld: Temporal residual world model for end-to-end autonomous driving. In *Proceedings of the Thirteenth International Conference on Learning Representations, ICLR 2025*, 2025.

Haiming Zhang, Junfei Zhou, Feng Jiang, Jingzhong Li, Zhenglong Guo, Penglin Dai, Jifeng Dai, Yan Xie, and Benjin Zhu. Anyscene: Towards highly controllable driving scene generation at anywhere and beyond. *CoRR*, abs/2605.26113, 2026.

Wenzhao Zheng, Weiliang Chen, Yuanhui Huang, Borui Zhang, Yueqi Duan, and Jiwen Lu. Occworld: Learning a 3d occupancy world model for autonomous driving. In *Computer Vision - ECCV 2024*, volume 15071 of *Lecture Notes in Computer Science*, pp. 55–72. Springer, 2024.

Hongyu Zhou, Longzhong Lin, Jiabao Wang, Yichong Lu, Dongfeng Bai, Bingbing Liu, Yue Wang, Andreas Geiger, and Yiyi Liao. Hugsim: A real-time, photo-realistic and closed-loop simulator for autonomous driving. *IEEE Trans. Pattern Anal. Mach. Intell.*, 48(4):4673–4691, 2026.

Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable DETR: Deformable transformers for end-to-end object detection. In *9th International Conference on Learning Representations, ICLR 2021*. OpenReview.net, 2021.

APPENDIX

In this appendix, we supplement the following materials to support the findings and conclusions drawn in the main body of this paper.

CONTENTS

A Operator Transition Details **14**

A.1 Spatial-Temporal Overlap of Scene 14

A.2 Bounded Trajectory Deviation 14

A.3 Bounded Operator Errors 15

A.4 Properties of Actor Erase-Print 16

B Simulator Details **16**

B.1 Datasets 16

B.2 Substep Calibration 16

B.3 Teacher-Forcing Protocols 16

B.4 Perceptual Fidelity 17

B.5 Agent States 17

B.6 Bicycle Ego Kinematic 18

B.7 Unicycle Actor Kinematic 18

B.8 The Controllers 19

B.9 Traffic Manager 19

B.10 Operator Timing 19

B.11 Sensitivity of Occupancy IoU to the Driven Pose Deviation 19

B.12 Deviation Threshold 19

C Actor Policy Details **21**

C.1 Scripted Baselines 21

C.2 Naturalness Metrics 21

C.3 Adversarial Traffic Behavior 21

C.4 Normal Traffic Behavior 22

C.5 Training Overhead 23

C.6 Tokenizer Robustness to Driven Pose Deviation 23

C.7 Adoption of Occupancy Tokenizer 23

D Benchmark and Leaderboard Details **23**

D.1 Scenario, Termination, and Scoring 23

D.2 Planner Suite 24

D.3 Planner Reproduction 24

D.4 WorldEngine and RLFT Reproduction 25

D.5 Multi-Step RLFT 25
D.6 Per-Planner Analysis 25
D.7 Traffic-Intensity Scaling 26

A Operator Transition Details

We empirically and theoretically study why our proposed rigid-operator transition is lossless. The transition T of Section 3.2 reduces to two operators, Ego Warp-Recenter and Actor Erase-Print, justified by two observations. Adjacent frames share most of their content, and the benchmarked planners stay close to the recorded route, with an open-loop ADE of only 0.32–0.63 m on average (Table 16). We bound the induced operator error theoretically below.

A.1 Spatial-Temporal Overlap of Scene

We verify the first observation empirically. On the whole validation set, we warp each frame-t occupancy into the frame-$(t{+}1)$ ego frame with the recorded ego motion and score the warped frame against the GT next-frame occupancy. Table 5 reports the per-class IoU with no warp, with Ego Warp-Recenter alone, and with Actor Erase-Print added.

Table 5: Spatial-temporal overlap of consecutive frames. mIoU averages the 17 semantic classes, excluding free. IoU is the binary non-free occupancy IoU.

	others	barrier	bicycle	bus	car	constr.	motor.	pedestr.	cone	trailer	truck	drivable	o.flat	sidewalk	terrain	manmade	veget.	free	**mIoU** ↑	**IoU** ↑
No warp (%)	13.3	25.2	8.9	31.8	21.5	17.5	8.4	7.7	9.3	19.0	21.0	49.4	33.7	32.7	30.5	19.0	18.8	94.5	**21.6**	**30.5**
Warp (%)	59.6	62.5	40.5	46.0	39.3	60.6	32.0	30.0	52.1	52.8	49.0	59.1	51.3	52.7	49.6	59.3	60.6	97.4	**50.4**	**59.1**
Warp + Print (%)	59.6	62.5	43.9	65.2	57.1	58.0	48.9	38.6	52.1	55.4	63.2	59.2	51.3	52.7	49.6	59.3	60.6	97.5	**55.1**	**60.0**

Warping nearly doubles the non-empty IoU, from 30.5% to 59.1%, with free space at 97.4%. The residual concentrates on dynamic objects such as car 39.3%, motorcycle 32.0% and pedestrian 30.0%, and grows with ego displacement. 82% non-empty IoU below 1 m of motion vs 53% above 4 m. Actor Erase-Print on the warp repositions each actor at its next-frame pose and recovers most of this residual, lifting mIoU from 50.4% to 55.1%. Car 39.3%→57.1%, bus 46.0%→65.2%, truck 49.0%→63.2%, motorcycle 32.0%→48.9%, and pedestrian 30.0%→38.6%. Static structure is largely shared. Buildings 59%, roads 59%, vegetation 61%. Thus Ego Warp-Recenter alone more than doubles overlap (30.5% to 59.1% IoU) and Actor Erase-Print recovers the residual, to 60.0%. In closed-loop GT replay the occupancy state is re-centred from the recorded next frame, reaching 75.3% IoU. Appendix B reports the per-operator timings.

Algorithm 1 Ego Warp-Recenter transition, with X'_{t+1} the reference occupancy, P the driven pose, and ℓ the cell size

$X_{t+1} \leftarrow \textsc{Free}(H \times W \times Z)$
for each cell centre c' of the reference occupancy X'_{t+1} **do**
  $c \leftarrow \mathrm{Rot}(-\bar{\varphi})(c' - \bar{p}_{t+1})$ ▷ reference frame → ego frame
  $k \leftarrow \lfloor \frac{1}{\ell} c + \frac{1}{2} \rfloor$ ▷ nearest re-centred cell
  $X_{t+1}[k] \leftarrow X'_{t+1}[c']$ if k is in bounds, else discarded
end for

A.2 Bounded Trajectory Deviation

We assume that the driven ego position re-expressed in the reference-state frame (Eq. (5)) stays within ε of the recorded route, i.e., $\|\bar{p}_{t+1}\|_2 \le \varepsilon$ for all valid t. Two arguments justify this assumption. First, the benchmarked planners are trained by imitation learning, so by construction their plans stay close to the recorded route. The open-loop ADE measures this directly at 0.32–0.63 m on average (Table 16). Second, when the deviation is larger than a threshold, the ego can be considered off-route,

Algorithm 2 Actor Erase-Print transition, with X'_{t+1} the occupancy, i the actor, P the ego pose, $\bar{G}^{(i)}_{t+1}$ the actor pose in the ego frame, and $(w^{(i)}, l^{(i)})$ the recorded box size

Setup at selection frame t_0
  $pattern \leftarrow$ CAPTURE(actor body inside its recorded box at the GT pose)
  ERASE(actor's own-class voxels inside the recorded box)
  FILLGROUND(columns freed by the erase, from GT) ▷ restore bottom ground
Each transition $t \rightarrow t{+}1$ while the actor is alive
for fine step in the actor's rollout under its behavior **do**
  if OFFROAD or OBSTACLE or EGOHIT **then**
    stop at the previous safe state ▷ validation-before-paint
  end if
end for
STAMP($pattern$, rigidly rotated to the driven pose)

a condition that closed-loop benchmarks treat as termination (Jia et al., 2024). Therefore, bounded deviation is a proper assumption rather than a restriction we impose on the simulator.

A.3 BOUNDED OPERATOR ERRORS

We bound the error of the two operators. Under bounded deviation of poses, let the grid have cell size ℓ and range R, and let the driven ego pose deviate by at most ε in position and $\Delta\theta$ in heading. A voxel centre at range $r \le R$ moves by at most

$$d_{\max} = \varepsilon + R \sin \Delta\theta \le \varepsilon + R\,\Delta\theta, \tag{17}$$

between the two views. A cell's class in the produced occupancy can differ from the GT view X'_{t+1} only if its voxel centre lies within $d_{\max}$ of a class boundary. Let C_c be the cells of class c in the GT view X'_{t+1}, with cardinality $|C_c|$, and S_c the number of class-c boundary cells (with a neighbour of a different class). Ego Warp-Recenter, which re-centers the whole grid, changes class only within the swept boundary band, at most $S_c\,(1 + 2\lceil d_{\max}/\ell \rceil)$ cells. Actor Erase-Print adds an independent term. A re-driven actor i of class c_i, displaced by at most ε_a and with B_i boundary cells on its rigid body, is unchanged in its interior and changes class only within the band around its body boundary, at most $B_i\,(1 + 2\lceil \varepsilon_a/\ell \rceil)$ cells. Summing over the controlled actors, the total changed cells N_c of class c are bounded by the two terms together,

$$N_c \le S_c \left(1 + 2\lceil d_{\max}/\ell \rceil\right) + \sum_{\substack{i:\ c_i = c \\ \text{controlled}}} B_i \left(1 + 2\lceil \varepsilon_a/\ell \rceil\right), \tag{18}$$

Writing C'_c for the cells labelled c in the produced occupancy X_{t+1}, the per-class IoU $\mathrm{IoU}_c = |C_c \cap C'_c|/|C_c \cup C'_c|$ changes only when a cell's class does. Since at most N_c cells change, $\mathrm{IoU}_c \ge (|C_c| - N_c)/(|C_c| + N_c) \ge 1 - 2N_c/|C_c|$. The IoU error is therefore bounded by

$$1 - \mathrm{IoU}_c \le \frac{2S_c}{|C_c|}\left(1 + 2\lceil d_{\max}/\ell \rceil\right) + \frac{2}{|C_c|} \sum_{\substack{i:\ c_i = c \\ \text{controlled}}} B_i \left(1 + 2\lceil \varepsilon_a/\ell \rceil\right). \tag{19}$$

Under the bounded-deviation assumption, ε is bounded by the benchmark's deviation threshold and ε_a by the per-class per-frame displacement cap (Section B), so $d_{\max}$ and ε_a are finite, with $d_{\max}$ spanning $\lceil d_{\max}/\ell \rceil$ cells. For example, a 5 m threshold at $\ell = 0.4$ m gives at most 13 cells, and a 6 m car step at most 15. The bound is therefore finite but not uniformly tight. The boundary-to-volume ratio $S_c/|C_c|$ is small for large, voluminous classes such as roads, buildings, and vegetation, but can approach $O(1)$ for thin or small classes such as pedestrian and traffic cone, where a small absolute change is a large fraction of the class. The consequence we rely on is not a small per-step error but that the error does *not accumulate*. Each step re-derives the occupancy state from the recorded reference state rather than from the previous prediction, so the per-step bound does not compound with the horizon. The empirical per-step errors are reported in Table 5. This error-bounded, non-accumulating property, together with the use of the reference state, contributes to the long-horizon and high-fidelity simulation in the occupancy modality. □

A.4 Properties of Actor Erase-Print

Stamping is order-independent. Every controlled actor's fine-step trajectory is resolved against the occupancy before any body is stamped, so the result does not depend on the stamping order, and classes are never relabelled, as erasure removes only the actor's own-class voxels and stamping writes only the actor's class. A terminated actor freezes at its last safe pose and is re-stamped there each frame, so it remains a static obstacle instead of vanishing. Φ captures the actor's original voxel pattern once at selection and re-prints it at the driven pose, preserving its shape and class across frames, which keeps the actor in-distribution for the tokenizer. A pose is validated obstacle-free before printing, so Actor Erase-Print never overwrites static or non-traversable structure (Algorithm 2).

B Simulator Details

This appendix details the simulator, including the ego and actor kinematic models, the controllers that execute plans at the substep cadence, the range-aware traffic manager that selects which actors to re-drive, and the fidelity, baselines, and tokenizer studies.

B.1 Datasets

Table 6 summarizes the occupancy datasets used in our experiments. All are ego-centred and metric. Occ3D-nuScenes is recorded natively at 2 Hz while Occ3D-Waymo is 10 Hz. Occ3D-Waymo shares the 0.4 m, $200\times200\times16$ ego-centred grid of Occ3D-nuScenes.

Table 6: Dataset statistics of the occupancy benchmarks used, both from Occ3D (Tian et al., 2023).

Property	**Occ3D-nuScenes**	**Occ3D-Waymo**
Source	nuScenes, validation	Waymo Open Dataset, validation
Scenes / frames	150 scenes, 6,019 frames	202 scenes, ∼39,987 frames
Frames per clip	∼40 (20 s)	∼198 (20 s)
Native frequency	2 Hz (0.5 s)	10 Hz (0.1 s)
Voxel resolution	0.4 m	0.4 m
Point-cloud range	$[-40,-40,-1]$–$[40,40,5.4]$ m	$[-40,-40,-1]$–$[40,40,5.4]$ m
Voxel grid	$200\times200\times16$	$200\times200\times16$
Semantic classes	18	18

B.2 Substep Calibration

The control executes K integration substeps per occupancy frame ($\tau' = \tau/K$). We fix the execution cadence at 20 Hz, giving $K{=}20/\mathrm{fps}$, so $K{=}10$ at 2 Hz under the nuScenes/I^2-World protocol and $K{=}2$ at the native 10 Hz. Table 7 sweeps K over scene-start episodes at the 16 s horizon under GT replay. Accuracy is flat for $K \in [1,5]$ and collapses at $K{=}8$, where the coarse interpolation of the six-waypoint reference becomes unstable; $K{=}2$ is best or within noise of best on Occ3D-Waymo, so we adopt the 20 Hz rule.

Table 7: Substep calibration at 10 Hz over the 16 s horizon under GT replay. Values are IoU/mIoU. $K{=}2$ follows the 20 Hz execution rule.

Dataset	$K{=}1$	$K{=}2$	$K{=}3$	$K{=}5$	$K{=}8$
Occ3D-Waymo	0.829/0.804	**0.842/0.803**	0.841/0.800	0.838/0.799	0.611/0.477

B.3 Teacher-Forcing Protocols

We compare the learned baselines at their strongest, one-frame-lookahead setting, which isolates the effect of the transition operators. At each step the baseline is fed the *recorded* reference frame, so no error compounds and the reference frame is kept largely unaffected; any remaining gap is therefore

attributable to the fidelity of the learned transition rather than to error accumulation. For I^2-World (Liao et al., 2025b), we run the released checkpoint under per-step teacher-forcing, feeding back the *recorded* frame token and ego motion at each step, with the released evaluation settings (same Occ3D-nuScenes windows, ego conditioning, frozen tokenizer, and official *forecasting_miou*). The I^2-World-TF rows therefore measure reconstruction-level fidelity given perfect per-step context, not a multi-step forecast. For DOME (Gu et al., 2024), the DOME-TF row applies the same protocol with the released diffusion model, which generates a block of future occupancy from a recorded context window. To match the one-frame-lookahead comparison, we condition on the *recorded* window of up to four context frames ending at the current frame, generate the block, and score only its *first* frame against the recorded next frame. The rest is discarded and no generated frame is fed back, so DOME enters at its strongest in-distribution point. The long-horizon curve aggregates this lookahead over the same 150 val scenes as WorldOcc. Table 8 reports both. Even with this oracle context, both patched baselines still fall short of WorldOcc, showing the limited fidelity of their learned models.

Table 8: Long-horizon occupancy fidelity in IoU/mIoU ↑(%) under per-step teacher-forcing.

Method	**Year**	1 s	2 s	3 s	8 s	16 s	**Avg**
DOME-TF (Gu et al., 2024)	2024	47.6/41.8	49.6/43.0	49.0/42.4	48.7/42.3	49.1/41.2	48.8/42.1
I^2-World-TF (Liao et al., 2025b)	2025	56.6/53.8	57.2/56.1	57.2/55.3	58.3/56.5	57.4/54.7	57.4/55.3
WorldOcc (ours)	2026	73.4/72.4	75.0/72.7	77.0/77.0	76.3/72.7	73.9/70.5	**75.1/73.1**

B.4 Perceptual Fidelity

We additionally compare the realism of the generated occupancy itself with the generative-metric protocol of SemCity (Lee et al., 2024). Following SemCity's renderer, every frame is rendered to a 2D image with `pyrender` using the front camera, and scored with default `torch-fidelity` InceptionV3 to report the Inception Score (IS) on the generated frames, and the Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) between the generated and reference distributions. All methods are evaluated on the same Occ3D-nuScenes val windows at the matched $1/2/3$ s horizon ($2/4/6$ future frames at 2 Hz), with the recorded GT target frames of those windows, rendered identically, as the reference distribution. DynamicCity (Bian et al., 2025) is omitted as its conditional model weight on this dataset is not released. Table 9 reports that WorldOcc attains the best FID and KID. Its rigid-operator transition preserves the recorded scene by construction, so its re-centred frames stay close to the reference, whereas the learned generators trade fidelity for diversity.

Table 9: Perceptual metrics on Occ3D-nuScenes at the matched $1/2/3$ s horizon. DynamicCity (Bian et al., 2025) is omitted as its conditional model weight on this dataset is not released.

Method	**Year**	IS ↑	FID ↓	KID ↓
DOME (Gu et al., 2024)	2024	2.046	13.13	0.0081
OccWorld (Zheng et al., 2024)	2024	2.173	24.47	0.0187
I^2-World (Liao et al., 2025b)	2025	1.925	36.95	0.0287
WorldOcc (ours)	2026	**2.210**	**1.06**	**0.00004**

B.5 Agent States

The ego state is $E_t = (x_t, y_t, \theta_t, v_t, \delta_t, \alpha_t, g_t)$ (pose (x, y, θ), speed v, steering δ, acceleration α, and a one-hot navigation command $g_t \in \{0,1\}^3$). Each actor's state is $G_t^{(i)} = (x_t^{(i)}, y_t^{(i)}, \theta_t^{(i)}, v_t^{(i)}, c^{(i)}, w^{(i)}, l^{(i)})$, carrying pose and speed with class $c^{(i)}$ and box size $(w^{(i)}, l^{(i)})$. Only the driven poses E_{t+1}, G_{t+1} enter the transition T (Eq. (4)). The command and each actor's class and size are static attributes consumed by the planners, operators, and traffic manager. In particular, the recorded box $(G_{t_0}^{\prime(i)}, w^{(i)}, l^{(i)})$ defines the footprint of Actor Erase-Print, and capture, erasure, and ground restoration are all restricted to it.

B.6 BICYCLE EGO KINEMATIC

The ego is a kinematic bicycle (wheelbase L, frame period τ, state $E = (x, y, \theta, v, \delta, \alpha)$). Its time derivative $\dot{E} = (\dot{x}, \dot{y}, \dot{\theta}, \dot{v}, \dot{\delta}, \dot{\alpha})$, i.e. $\dot{E} = \mathrm{d}E/\mathrm{d}t$, is

$$\dot{E} = \Big(v\cos\theta,\ v\sin\theta,\ \tfrac{v\tan\delta}{L},\ \alpha,\ \dot{\delta},\ \dot{\alpha} \Big), \tag{20}$$

which, discretized with the fine Euler step $\tau' = \tau/K$ at K substeps per frame, gives the update

$$v' = v + \alpha\,\tau', \quad \delta' = \delta + \dot{\delta}\,\tau', \quad \theta' = \theta + \frac{v'\tan\delta'}{L}\,\tau', \quad p' = p + v' \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} \tau', \tag{21}$$

where α and $\dot{\delta}$ are the applied acceleration and steering rate. A single coarse control hold per frame makes the driven pose oscillate on a curved route, so we separate the occupancy frame rate from a fixed 20 Hz execution rate by adjusting the substeps K. This keeps the driven pose close to the recording. Raising K from 1 to 10 cuts the driven-vs-GT error from $\sim$0.9 m/6.5° to $\sim$0.1 m/0.4°, and the trend plateaus beyond K=10. It refines only the low-level execution of the emitted plan and leaves the route unchanged.

At each frame the planner emits six waypoints 0.5 s apart. iLQR cubic-splines them into $6K$ fine waypoints and solves one full-horizon problem at the fine step $\tau' = \tau/K$, producing K control samples that are applied one per substep. It therefore solves once per frame, not per substep, and adds no waypoints beyond the spline. The final substep expresses the driven pose in the next GT frame, as Ego Warp-Recenter expects. Prior simulators (e.g. HUGSIM (Zhou et al., 2026)) apply a single coarse iLQR hold per frame, which under-suppresses the high-frequency curvature present in the noisy recorded trajectory.

B.7 UNICYCLE ACTOR KINEMATIC

Each actor is a unicycle with state $G^{(i)} = (x^i, y^i, \theta^i, v^i)$, where $p^i = (x^i, y^i)$ is the actor's position, and control (α^i, ω^i). Its time derivative $\dot{G}^{(i)} = (\dot{x}^i, \dot{y}^i, \dot{\theta}^i, \dot{v}^i)$, i.e. $\dot{G}^{(i)} = \mathrm{d}G^{(i)}/\mathrm{d}t$, is

$$\dot{G}^{(i)} = \Big(v^i\cos\theta^i,\ v^i\sin\theta^i,\ \omega^i,\ \alpha^i \Big), \tag{22}$$

with per-class limits on speed, acceleration, deceleration and yaw rate, and discretized as

$$v' = \max(0,\ v + \alpha\,\tau'), \quad \theta' = \theta + \omega\,\tau', \quad p' = p + v' \begin{bmatrix}\cos\theta' \\ \sin\theta'\end{bmatrix} \tau'. \tag{23}$$

The action $a_t^{(i)} \in \mathbb{R}^2$ is an *aim point* specifying a target offset from the actor. The controller steers toward it with a pure-pursuit law,

$$e_\theta = \operatorname{atan2}\left(w^i_{\text{aim}} - p\right) - \theta, \qquad \omega = \operatorname{clamp}\left(k\,e_\theta, \pm\omega_{\max}\right), \tag{24}$$

where θ is the actor's heading, e_θ the bearing to the aim point, k the proportional gain, and $\omega_{\max}$ the per-class yaw-rate limit. The desired speed v^* follows a braking-distance profile that leaves the actor able to stop before the aim point, clamped by the per-class maximum speed $v_{\max}(c)$,

$$v^* = \min\left(v_{\max}, \sqrt{2\,a_{\text{dec}}\,\|w^i_{\text{aim}} - p\|}\right)\ \max\left(0, \cos e_\theta\right), \tag{25}$$

where a_{dec} is the per-class deceleration limit. The per-class speed cap $v_{\max}(c)$ decodes the per-frame displacement cap $d_c(c)$ at the control period Δt,

$$v_{\max}(c) = \frac{d_c(c)}{\Delta t}, \tag{26}$$

A per-substep safety brake stops the actor before overlapping a non-traversable cell or the ego, holding one substep short of contact.

B.8 The Controllers

Following Karnchanachari et al. (2024), the ego's iterative Linear Quadratic Regulator (iLQR) controller interpolates the six emitted waypoints with a cubic spline into $6K$ fine waypoints, then solves a receding-horizon tracking problem whose control sequence is re-solved at the frame cadence and consumed at the fine substep rate. Each selected actor is driven by an IDM-inspired waypoint-following controller that tracks the actor's emitted aim point with the per-class kinematic clamps of Eq. (26) and an obstacle safety brake. It follows IDM's spirit of smooth acceleration toward a desired speed and safe braking, but not IDM's car-following formula, since it follows an arbitrary waypoint rather than a lead vehicle (Treiber et al., 2000).

Algorithm 3 Substep iLQR ego control, executed once per frame. E_t is the ego state, $\{w_j\}_{j=1}^{6}$ the six emitted waypoints, K the substeps per frame, and $\tau' = \tau/K$ the fine step.

$r \leftarrow \text{cubic-spline}\big(\{w_j\}\big)$ sampled at τ' ▷ $6K$ fine waypoints
$u \leftarrow \text{iLQR}(E_t, r)$ ▷ standard iLQR (Karnchanachari et al., 2024), with a warm start followed by forward rollout, backward Riccati pass, and forward pass to convergence
for $j = 0, \ldots, K-1$ **do** ▷ one fine step $\tau' = \tau/K$ per substep
  $\alpha_j, \dot{\delta}_j \leftarrow u_j$ ▷ applied acceleration and steering rate
  $v \leftarrow v + \alpha_j \, \tau'$ ▷ Eq. (21)
  $\delta \leftarrow \delta + \dot{\delta}_j \, \tau'$
  $\theta \leftarrow \theta + \frac{v \tan \delta}{L} \, \tau'$
  $p \leftarrow p + v \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tau'$
end for

B.9 Traffic Manager

Each actor class c has a per-frame displacement cap $d_c(c)$, so an actor within the reachable radius $F\, d_c(c)$ can reach the ego within $F = 6$ frames. For temporal diversity we draw the start frame $t_0 \sim \text{Uniform}(\{0, \ldots, \min(F, M)\})$, with M the number of frames remaining in the clip, walk the frames $t = t_0, \ldots, \min(F, M)$, and at each construct the candidate set $\mathcal{C}_t$ of actors that are not yet selected, are valid at t, and lie within their class reach. With $\mathcal{P}$ the participant set, $\mathcal{R}$ the re-driven set, $p_i(t)$ participant i's ego-centric pose, and $m_i(t)$ its validity mask,

$$\mathcal{C}_t = \big\{\, i \in \mathcal{P} \setminus \mathcal{R} : m_i(t) = 1, \;\; \|p_i(t)\|_2 \le F\, d_c\big(c(i)\big) \,\big\}, \tag{27}$$

At the first frame with a non-empty candidate set, we draw one actor uniformly, $i^\star \sim \text{Uniform}(\mathcal{C}_t)$, update $\mathcal{R} \leftarrow \mathcal{R} \cup \{i^\star\}$, and continue until k actors are drawn or the frames run out.

B.10 Operator Timing

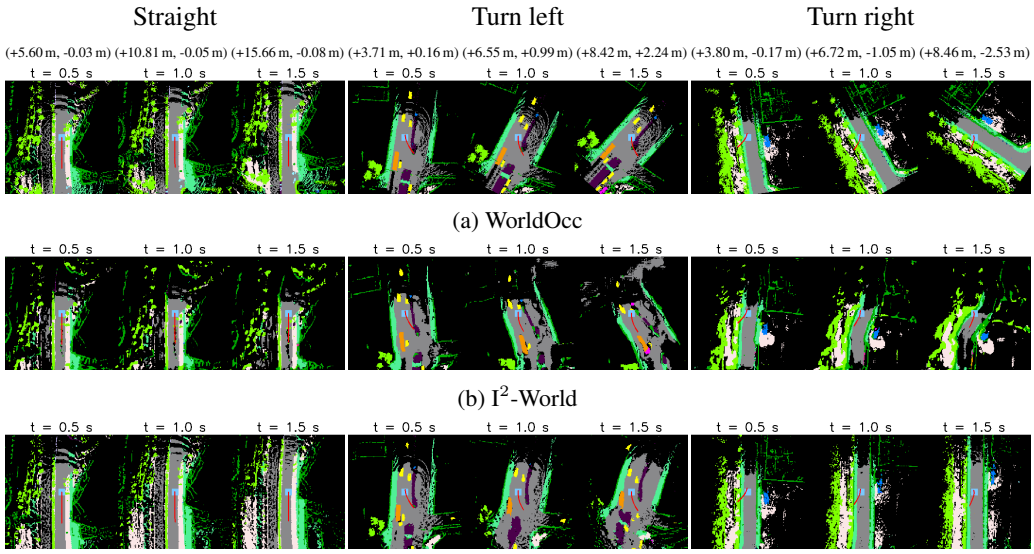
The main-text throughput is the operator-only rate on an RTX 4090, excluding the controller and perception. Ego Warp-Recenter takes 0.79 ms, about 1270 FPS, and Actor Erase-Print 3.58 ms, about 280 FPS. On CPU with numba they take 2.3 ms, about 436 FPS, and 12.4 ms, about 81 FPS.

B.11 Sensitivity of Occupancy IoU to the Driven Pose Deviation

Under a controlled pose deviation, IoU is highly sensitive to heading error ($0.5^\circ \to 68\%$, $1^\circ \to 50\%$) and far less to translation, as sub-15 cm offsets are absorbed by the 0.4 m voxels and it falls only from 20 cm (Table 10). Since the GT-Following ego's heading error is small ($\sim$0.6° median), most of the residual IoU loss is the metric's rotation sensitivity at the controller's tracking error, not scene degradation. At this heading floor the reported 70–80% IoU is near the metric's ceiling, which WorldOcc holds flat while the learned baselines decay.

B.12 Deviation Threshold

The deviation threshold d_{dev} is applied *after* Ego Warp-Recenter, so it also bounds how much of the finite 80×80 m scene can leave the window. Both frames are 80×80 m, so after re-centring at a



(a) WorldOcc

(b) I^2-World

(c) DOME

Figure 6: Controllability under a counterfactual natural turning command.

Table 10: Occupancy IoU / mIoU under a controlled driven-vs-GT deviation, using pure translation (cm) or pure heading rotation (deg).

(a) Translation (cm)

dev (cm)	0	5	10	15	20	30	40	60	80
IoU	1.000	1.000	1.000	1.000	0.628	0.546	0.546	0.448	0.415
mIoU	1.000	1.000	1.000	1.000	0.597	0.517	0.517	0.392	0.350

(b) Heading (deg)

dev (deg)	0.0	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	8.0
IoU	1.000	0.681	0.495	0.416	0.366	0.304	0.266	0.240	0.220	0.193
mIoU	1.000	0.694	0.486	0.376	0.304	0.221	0.173	0.140	0.118	0.089

pose displaced by d the retained scene is the overlap of two co-sized squares. The worst placement is on the diagonal, offset $(d/\sqrt{2}, d/\sqrt{2})$, giving

$$\rho_{\text{ret}}(d) = \left(\frac{80 - d/\sqrt{2}}{80}\right)^2, \tag{28}$$

a corner band of width $d/\sqrt{2}$. For example, 5 m keeps at least 91% of the scene and 10 m only 83% (Table 11), so we adopt $d_{\text{dev}} = 5$ m, which leaves planners enough range to react to varying traffic while retaining a valid fraction of the in-bound scene, whereas 10 m would let them drift too far.

Table 11: Retained scene fraction $\rho_{\text{ret}}(d) = ((80 - d/\sqrt{2})/80)^2$ under a driven-pose displacement d on the worst-case diagonal.

d (m)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R (%)	98	97	95	93	91	90	88	86	85	83	82	80	78	77	75

C Actor Policy Details

C.1 Scripted Baselines

A re-driven actor is controlled from its start frame until the end of the clip and follows one of two scripted behaviors. *Normal scripted* replays the actor's recorded trajectory by aiming at its next GT waypoint, and *aggressive scripted* aims at the ego's current position, the origin of the current ego-centric frame, with no future or ground-truth information. Each behavior defines an aim target

$$w^i_{\text{aim}} = \begin{cases} \text{next GT waypoint}, & \text{normal scripted}, \\ \mathbf{0}, & \text{aggressive scripted}, \end{cases} \tag{29}$$

and the controller of Appendix B steers the actor's heading toward this target while following the braking-distance speed profile of Eq. (25). Actors that are not re-driven keep their recorded motion. The learned policies below share this aim-target and controller interface, emitting a normalized action scaled by $d_c(c)$ into an aim offset.

C.2 Naturalness Metrics

To complement the ego collision rate (Fig. 4), we report two naturalness metrics for the controlled actors. *Off-road density* is the fraction of controlled-actor frames whose footprint is majority off the drivable area, and *non-ego conflict density* the fraction on which the safety brake fires against a non-ego object (Fig. 7).

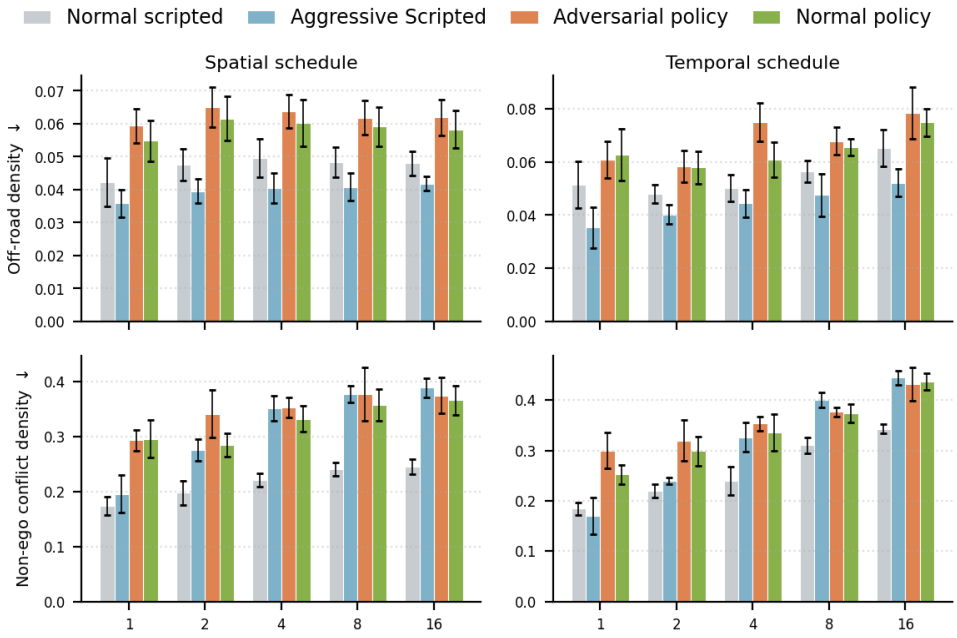


Figure 7: Off-road and non-ego conflict densities over the spatial and temporal k-sweeps, averaged over five seeds with std error bars. Ego collision rate is in Fig. 4.

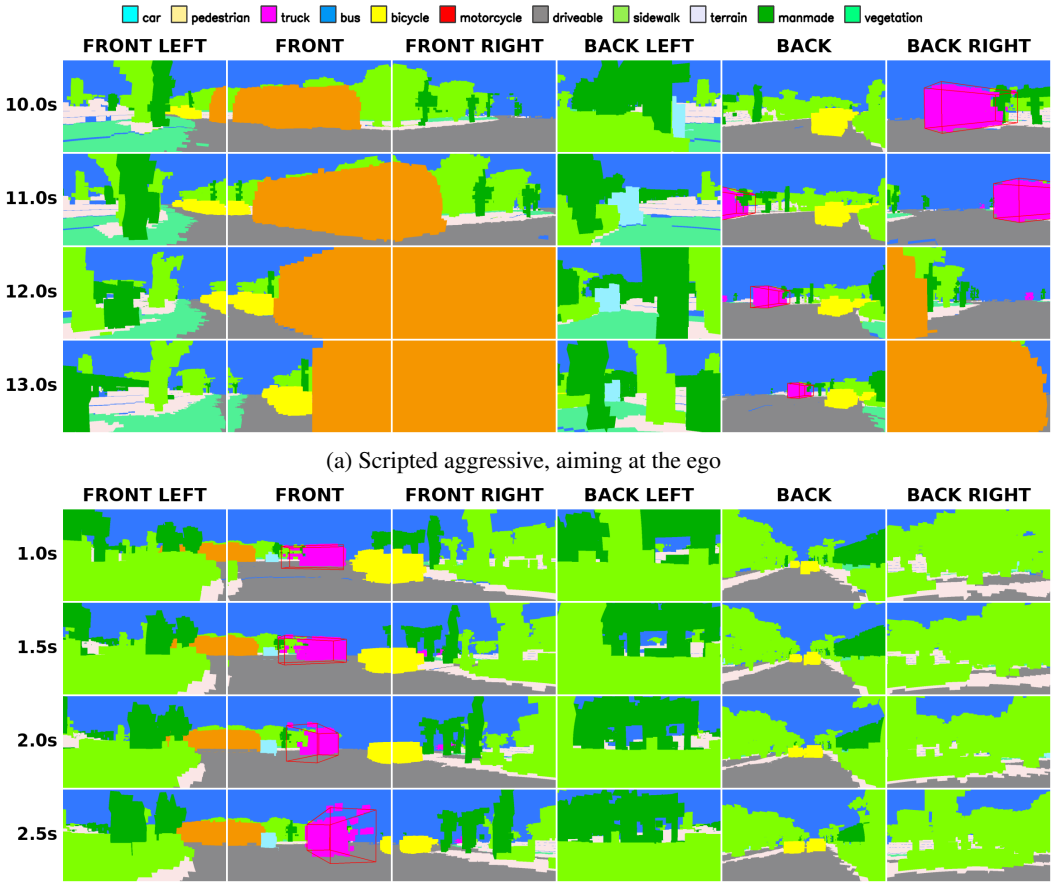
C.3 Adversarial Traffic Behavior

The adversarial actor is a shared-weight policy trained with single-agent PPO in an independent-learning (IPPO-style) setup (Schroeder de Witt et al., 2020) and deployed independently across actors. One network reads each controlled actor's localized scene context and drives the whole fleet. The objective is an expectation over single-actor episodes, each actor maximizing the standard PPO clipped surrogate over its own reward $r_t^{(i)}$ (Eq. (14)),

$$\mathcal{L}^{\text{CLIP}}(\theta) = \mathbb{E}_{i,t}\Big[\min\big(\rho_t^{(i)}(\theta)\,\hat{A}_t^{(i)},\ \text{clip}\big(\rho_t^{(i)}(\theta),\,1-\epsilon,\,1+\epsilon\big)\,\hat{A}_t^{(i)}\big)\Big], \tag{30}$$

where $\rho_t^{(i)}$ is the importance-sampling ratio, $\hat{A}_t^{(i)}$ the advantage from the value head, clip the clipping operator, and ϵ the PPO clipping range. No joint reward or critic is needed. Only the shared weights tie the fleet. Training samples one actor per episode, so the policy learns to drive different actors, and at inference the manager drives all selected actors with it.

The policy is trained with PPO (Schulman et al., 2017) on a single RTX 4090, with rollouts over eight subprocess environments. We use Stable-Baselines3 (Raffin et al., 2021) defaults except a 64-step buffer, 3 update epochs, target KL 0.15, entropy coefficient 0.01, and $\gamma = 0.99$. The extractor applies single-level deformable attention over the 50×50 BEV latent with 8 heads and 4 sampling points per head over a 128-dim feature, and the policy and value heads are two-layer width-64 MLPs over it. The action is a normalized $[-1, 1]^2$ aim point scaled by the per-class displacement cap and tracked by the unicycle controller. The reward uses $r_{\text{coll}} = +20$, $r_{\text{out}} = -10$, $\lambda = 2$, $\kappa = 1$. The actor trains on 20 s episodes that re-drive one uniformly sampled reachable actor per episode. The checkpoint is selected by the mean return over 30 deterministic evaluation episodes. Training is stable, with bounded trust-region, entropy and explained-variance curves. Across both schedules and all k, the adversarial policy's off-road and non-ego conflict densities (Fig. 7) stay near the scripted aggressive level, so its ego-threat gain is not bought by reckless driving (Fig. 8).



(a) Scripted aggressive, aiming at the ego

(b) Learned actor policy (Section 3.3)

Figure 8: The same adversarial encounter, a re-driven truck, in the ego's 2D semantic cameras, driven by the scripted aggressive heuristic (top) and by the learned actor policy (bottom). The re-driven actor is highlighted in red with time along rows and the six cameras along columns.

C.4 Normal Traffic Behavior

The normal actor shares the architecture, observation and action space of the adversarial policy. Its label is the displacement toward the recorded GT pose at the next frame, in the current ego frame and

scaled by the per-class displacement cap d_c, which is 1.5–6 m per frame,

$$a_t^{\star(i)} = \mathrm{clip}\Big(\tfrac{1}{d_c(c)}\big(T_t^{-1} p_{t+1}^{\mathrm{gt}(i)} - p_t^{(i)}\big)\Big), \tag{31}$$

where $p_t^{(i)}$ is the driven actor position, $p_{t+1}^{\mathrm{gt}(i)}$ the recorded next position, and T_t the frame-$t \to t{+}1$ ego transform. We train the behavior-cloning objective of Eq. (12) on the Occ3D *train* split, for 300 episodes, 30 epochs, Adam lr 10^{-3}, and batch 128. The held-out action error is $\|a - \hat{a}\|_1$=0.18. In the multi-actor closed-loop protocol, with the GT-Following ego, both schedules, five seeds, and 150 scenes, the clone tracks the recording to ADE 0.24 m against the scripted re-drive's 0.23 m. Its ego collision rate (Fig. 4) matches the scripted re-drive (0.13 vs. 0.11, both run with the shared low-level ego-obstacle brake) and stays well below the adversarial policy, while its off-road (0.062 vs. 0.051) and non-ego conflict (0.334 vs. 0.238) densities stay close to the scripted re-drive (Fig. 7). It is a learned imitation of recorded traffic, not a perfect tracker, and far less ego-aggressive than the reward-tuned adversarial actor.

C.5 Training Overhead

Table 12 reports the trained component and training compute of each learned world model, excluding the occupancy tokenizer.

Table 12: Training overhead of the learned component excluding the tokenizer.

Method	Trained component	Params	Training compute
DOME (Gu et al., 2024)	spatial-temporal diffusion	248.2 M	32×RTX 4090, 2000 epochs
I^2-World (Liao et al., 2025b)	intra-inter scene encoder	22.7 M	8×RTX 4090, 48 epochs
WorldOcc (ours, RL)	wp-deform. atten. policy (+ critic)	91 K	1×RTX 4090, 0.93 M steps
WorldOcc (ours, BC)	wp-deform. atten. policy	78 K	1×RTX 4090, 300 episodes

WorldOcc trains only a light-weight actor policy, three orders of magnitude smaller than the learned world models (78–91 K parameters against 22.7–248.2 M) and on a single RTX 4090 rather than 8–32 GPUs. The scene transition carries no trained parameters at all, since it is realized by the rigid operators, so adding a new traffic behavior costs one policy training instead of retraining a heavy-weight world model. This is the concrete cost behind the plug-and-play claim of Section 3.3.

C.6 Tokenizer Robustness to Driven Pose Deviation

Adapted planners consume the frozen I^2-tokenizer, and since its encoder is trained on recorded frames, a degraded token on driven frames could masquerade as a planner failure. We rule this out by measuring its *self*-reconstruction fidelity, $\mathrm{decode}(\mathrm{encode}(x))$ against its own input x, on the driven occupancy of the closed loop and on controlled warps of a GT frame, over 20 scenes (Table 13). Self-IoU/mIoU is 65.7%/80.2% in closed loop, 66.2%/80.8% under translations, and 65.5%/79.9% under heading rotations, so it does not fall below the closed-loop value over the actual driven deviation. The tokenizer does not decay out of distribution, so the closed-loop degradation of tokenizer-based planners is not token corruption.

C.7 Adoption of Occupancy Tokenizer

Among open-source tokenizers (Table 14), both the I^2-tokenizer and DOME's reconstruct strongly, at 81.2/68.3 and 83.1/77.3 mIoU/IoU. The I^2-tokenizer is also much lighter (20.2 M vs 97.8 M) and is a VQ-VAE giving deterministic observations, whereas DOME's OCC-VAE samples its latent stochastically. To promote reproducibility, we adopt the I^2-tokenizer as the occupancy tokenizer.

D Benchmark and Leaderboard Details

D.1 Scenario, Termination, and Scoring

A *scenario* is one nuScenes scene under one traffic setting. At each step the planner observes $o_t = (B_t, \eta_t)$ of Section 3.1 and emits an ego trajectory. The scenario ends on a collision, on a

Table 13: Self-reconstruction fidelity of the frozen I^2-tokenizer under controlled non-GT warping.

(a) Translation (cm)

dev (cm)	0	5	10	15	20	30	40	60	80
IoU	66.2	66.2	66.2	66.2	66.3	66.1	66.1	66.3	66.2
mIoU	80.8	80.8	80.8	80.8	81.1	80.9	80.9	81.2	81.0

(b) Heading (deg)

dev (deg)	0.0	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	8.0
IoU	66.2	66.1	66.0	66.0	66.0	65.9	65.8	65.7	65.6	65.5
mIoU	80.8	80.8	80.7	80.7	80.6	80.5	80.4	80.4	80.3	79.9

Table 14: Reported occupancy reconstruction fidelity and parameter count of open-source occupancy tokenizers. Compression is the BEV-plane downsampling factor.

Tokenizer	**Year**	**Params**	**Architecture**	**Compression**	**mIoU** ↑	**IoU** ↑
DOME (OCC-VAE) (Gu et al., 2024)	2024	97.8 M	stochastic VAE	64×	83.1	77.3
OccWorld (Zheng et al., 2024)	2024	14.2 M	VQ-VAE	16×	71.63	78.12
I^2-World (Liao et al., 2025b)	2025	20.2 M	VQ-VAE	16×	81.22	68.30
OccSora (Wang et al., 2026a)	2026	137.4 M	VQ-VAE	512×	27.4	37.0

driven-ego deviation beyond d_{dev} from the reference next pose, or when the clip horizon is reached. A collision is an ego-footprint overlap with non-traversable occupancy above 10% of the footprint, and sub-threshold contacts do not terminate. DAC returns 1.0, 0.5 or 0.0 at drivable-ground footprint fractions ≥ 0.5, ≥ 0.3, or below. TTC extrapolates the ego at constant velocity at $+0.5$ and $+1.0$ s against actor boxes. Comfort checks a rolling 6-step (3 s) path against HUGSIM limits (Zhou et al., 2026). Eq. (15) uses $w_{\text{TTC}} = 5$, $w_{\text{C}} = 2$. Table 4 uses three traffic settings. The *recorded* setting keeps actors on GT, with no re-driving and one deterministic cell per planner, the *normal* setting re-drives actors with the *learned normal actor policy* trained by behavior cloning, detailed in Appendix C, and the *adversarial* setting re-drives them with the *learned adversarial policy* of Section 3.3. For the normal and adversarial settings each planner runs over 150 whole-clip scenes under the two schedules of Section 3.4, *spatial* ($k \in \{1, 2, 4\}$) and *temporal* ($k \in \{1, 2, 4, 8, 16\}$), each over seeds $s \in \{1, \ldots, 5\}$. The values in Table 4 are the mean over the five seeds, with an std below 0.009.

D.2 Planner Suite

GT-Following replays the recorded GT trajectory as its plan. **OccWorld** (Zheng et al., 2024) jointly forecasts future occupancy and plans autoregressively. **Ego-MLP** (Li et al., 2024) is a perception-free planner from the ego status alone, and **Go-Straight** a no-perception baseline returning straight-ahead waypoints. **BEV-Planner** (Li et al., 2024) is a simple end-to-end planner introduced to isolate the effect of ego status. Its **+** and **++** variants add the ego status to the BEV and additionally to the plan decoder over a history of frames. **DiffusionDrive** (Liao et al., 2025a) is a truncated-diffusion end-to-end planner that denoises a compact set of clustered trajectory anchors conditioned on the scene. We retain its truncated-diffusion planning head and feed it the occupancy-token latent in place of its sparse camera features. **ResWorld** (Zhang et al., 2025) is the temporal-residual world model (TR-World) for end-to-end planning, with a future-guided trajectory refinement module. **SSR** (Li & Cui, 2025) is the navigation-guided sparse scene representation for end-to-end planning, with a future feature predictor. **WorldEngine (BC)** is the trajectory-vocabulary scorer of WorldEngine (Li et al., 2026), trained by behavior cloning. **WorldEngine (RLFT)** applies one-step REINFORCE over the trajectory vocabulary with the imitation policy as a frozen reference (Li et al., 2026).

D.3 Planner Reproduction

OccWorld (Zheng et al., 2024) runs from its released checkpoint on the same 12-frame occupancy window, scoring its 6-frame plan under the identical protocol. The BEV-Planner (Li et al., 2024)

family decodes per-mode trajectories from the frozen I^2-tokenizer (Liao et al., 2025b) latent and is trained with AdamW at lr 10^{-4} and weight decay 0.01, following the original setup. Ego-MLP is the official BEV-Planner head, a three-layer MLP over the 12-dim ego status trained by command-masked L1 regression. Go-Straight is a simple baseline that returns six waypoints pointing straight ahead.

D.4 WorldEngine and RLFT Reproduction

WorldEngine (BC) is our occupancy port of the WorldEngine trajectory-vocabulary scorer (Li et al., 2026). Instead of regressing a trajectory directly, it scores a fixed vocabulary of candidate trajectories against the occupancy and selects the highest-scoring one, trained by behavior cloning. WorldEngine (RLFT) adds a one-step REINFORCE over the vocabulary on offline train-split rollouts. The vocabulary holds $L{=}4096$ plans from k-means over the recorded train futures, and only the scoring head is fine-tuned. The objective combines imitation on normal frames, a score-weighted policy gradient with importance sampling to the frozen imitation reference, and an entropy term,

$$\mathcal{L} = w_{\text{il}}\,\overline{\text{CE}(y, \log \pi)\,\mathbf{1}[\text{normal}]} + w_{\text{pg}}\,\overline{-\left(\pi_\theta/\pi_{\text{orig}}\right)_{\text{detach}} \cdot s \cdot \log \pi} + w_{\text{ent}}\,\overline{\textstyle\sum_k \pi \log \pi}, \tag{32}$$

where the overbar is the mean over frames, π is the plan-selection distribution over the vocabulary, y the imitation label, s the score of the selected plan, and $\text{CE}(\cdot,\cdot)$ the cross-entropy; the importance-sampling weight $(\pi_\theta/\pi_{\text{orig}})_{\text{detach}}$ is clipped at 10 and set to 1 on failure frames, and $w_{\text{il}}{=}0.5$, $w_{\text{pg}}{=}1.0$, $w_{\text{ent}}{=}0.2$, LR $5{\times}10^{-5}$, and 8 epochs over 300 train-scene rollouts. Since WorldEngine is evaluated on nuPlan with the NAVSIM PDM Score, not nuScenes/Occ3D, its published scores are not comparable to our WorldOcc-Score. Its method is ported but not reproduced, so we report our port's WorldOcc-Bench scores (Table 4) and leave its open-loop cells empty.

D.5 Multi-Step RLFT

The RLFT in (Li et al., 2026) uses a one-step return. We also vary the TD(n) horizon of the policy-gradient reward, $R_t = \sum_{i=0}^{n-1} \gamma^i r_{t+i}$ with $\gamma{=}1$, undiscounted to match the fixed-horizon driving score, computed within each logged episode. Here n counts 0.5 s frames, so $n{=}6$ is 3 s. All other settings are unchanged. Table 15 reports the recorded-setting score for $n \in \{1, \ldots, 6\}$. The improvement peaks at $n{=}2$ (0.309), then degrades for $n \geq 3$ as the unnormalized return grows with n while the entropy weight is fixed. The degradation is a scale effect rather than a discounting effect, since with $\gamma{=}1$ the n-step return is an undiscounted sum of nonnegative per-frame rewards whose magnitude grows with n and progressively outweighs the fixed entropy and imitation weights. Discounting ($\gamma{<}1$) would shrink that magnitude but truncate the effective horizon and bias the return away from the fixed-horizon driving score, so it is not the faithful remedy. Repeating the sweep with a discounted return ($\gamma{=}0.99$) shows the same peak at $n{=}2$ and decline beyond, confirming that the discount factor is not the cause. We hypothesize that return normalization is the next step to improve multi-step RLFT, towards the long-horizon driving capacity.

Table 15: WorldOcc-Score$_{\text{recorded}}$ versus the policy-gradient horizon n.

Model	NC ↑	DAC ↑	TTC ↑	Comf. ↑	RC ↑	WorldOcc-Score$_{\text{recorded}}$ ↑
IL base	0.941	0.874	0.930	0.791	0.312	0.273
RLFT $n{=}1$	0.976	0.970	0.969	0.960	0.331	0.293
RLFT $n{=}2$	**0.982**	**0.975**	**0.975**	0.951	**0.343**	**0.309**
RLFT $n{=}3$	0.971	0.966	0.955	0.956	0.293	0.255
RLFT $n{=}4$	0.975	0.965	0.961	0.955	0.293	0.259
RLFT $n{=}5$	0.969	0.966	0.958	0.935	0.302	0.262
RLFT $n{=}6$	0.976	0.969	0.960	0.921	0.293	0.251

D.6 Per-Planner Analysis

The perception-based planners share the same frozen I^2-tokenizer latent, and our open-loop reproductions match the reported metrics within 0.01 m for OccWorld and Ego-MLP and under the VAD (Jiang et al., 2023) rule for ResWorld and SSR (Table 16). ResWorld reproduces both reported open-loop variants (Table 16) yet collapses in closed loop, OccWorld degrades the same way despite being occupancy-native, and Ego-MLP, near-best open-loop, degrades most under re-driven traffic.

Table 16: Simulator-free open-loop planning metrics, L2 in m and collision in %. †OccWorld's ST-P3 protocol uses single-waypoint L2 on a 0.5 m grid.

	L2 (m) ↓				Collision (%) ↓				
Method	1 s	2 s	3 s	Avg	1 s	2 s	3 s	Avg	**Ckpt. source**
OccWorld-O (Zheng et al., 2024)†	0.43	1.08	1.99	1.17	0.07	0.38	1.35	0.60	reported
OccWorld-O†	0.43	1.08	1.99	1.17	0.07	0.40	1.35	0.61	reproduced
OccWorld-O	0.31	0.60	0.98	0.63	1.28	3.09	5.17	3.18	reproduced
BEV-Planner+ (Li et al., 2024)	0.28	0.42	0.68	0.46	0.04	0.37	1.07	0.49	reported
BEV-Planner+	0.16	0.31	0.57	0.34	0.04	0.39	1.07	0.50	reproduced
BEV-Planner++ (Li et al., 2024)	0.16	0.32	0.57	0.35	0.00	0.29	0.73	0.34	reported
BEV-Planner++	0.17	0.34	0.61	0.37	0.06	0.61	1.68	0.78	reproduced
Go-Straight (Li et al., 2024)	0.38	0.79	1.33	0.83	0.15	0.60	2.50	1.08	reported
Go-Straight	0.17	0.50	0.97	0.55	0.16	0.72	2.29	1.05	reproduced
Ego-MLP (Li et al., 2024)	0.15	0.32	0.59	0.35	0.00	0.27	0.85	0.37	reported
Ego-MLP	0.15	0.32	0.59	0.36	0.02	0.35	1.31	0.56	reproduced
ResWorld (Zhang et al., 2025)	0.17	0.32	0.55	0.35	0.01	0.02	0.16	0.07	reported
ResWorld	0.16	0.32	0.57	0.35	0.02	0.25	0.94	0.40	reproduced
ResWorld♢ (Zhang et al., 2025)	0.14	0.27	0.49	0.30	0.01	0.03	0.14	0.06	reported
ResWorld♢	0.14	0.29	0.53	0.32	0.02	0.21	0.92	0.38	reproduced
SSR (Li & Cui, 2025)	0.19	0.36	0.62	0.39	0.10	0.10	0.24	0.15	reported
SSR	0.15	0.31	0.56	0.34	0.02	0.39	1.31	0.57	reproduced

Table 16 reports open-loop metrics from the papers and our runs. our reproductions agree (OccWorld/Ego-MLP within 0.01 m with matching collision, ResWorld/SSR matching their reported L2). OccWorld's ST-P3 protocol uses single-waypoint L2 on a 0.5 m grid, which differs. Under the shared per-step ADE it scores 0.63 m / 3.18% instead of 1.17 m, so the main table compares it on the same measure as the rest. Fig. 5 shows a representative episode in the closed loop under recorded and adversarial re-driven traffic.

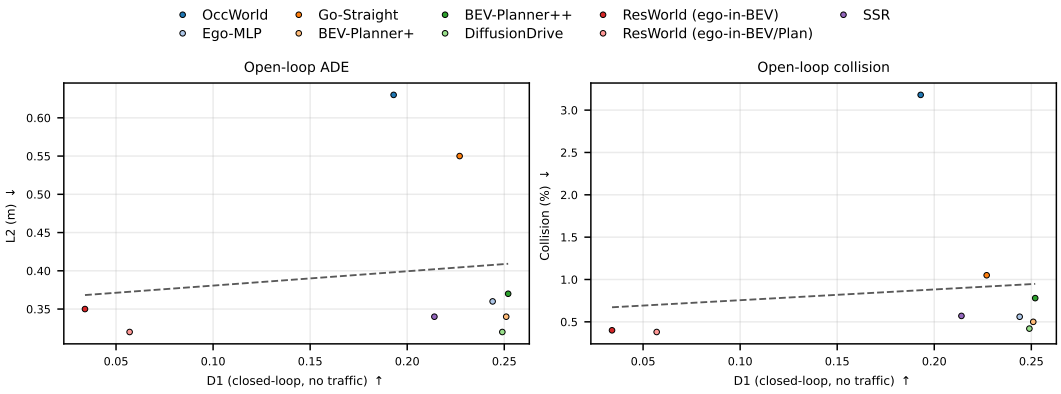


Figure 9: Open-loop and closed-loop correlation.

D.7 Traffic-Intensity Scaling

Table 17 reports how each planner's WorldOcc-Score degrades as the number of re-driven actors k grows, under both the normal and the adversarial settings. Two trends stand out. First, the temporal schedule degrades planners more than the spatial one at the same k, since the spatial candidate pool saturates near $k \approx 4$ whereas temporal encounters remain independent and keep growing to $k = 16$. Even the GT-Following oracle falls from 0.81 to 0.63 under adversarial temporal traffic, so the degradation is not a planner artifact. Second, planners differ widely in robustness. Ego-MLP is stable under normal temporal traffic (0.24 throughout) yet collapses under adversarial temporal

traffic ($0.24 \rightarrow 0.06$ at k=16), while BEV-Planner++ and WorldEngine (RLFT) degrade only mildly under either, and ResWorld stays near zero throughout because it already fails at k=1. The intensity sweep thus exposes the actor load at which each planner breaks.

Table 17: WorldOcc-Score vs the number of re-driven actors k under the normal (top) and adversarial (bottom) settings. Spatial runs $k \in \{1, 2, 4\}$ and temporal $k \in \{1, 2, 4, 8, 16\}$. †GT-Following is a human reference rather than a learned planner.

	Spatial			**Temporal**				
Planner	k=1	2	4	k=1	2	4	8	16
Normal								
GT-Following†	0.842	0.827	0.815	0.832	0.831	0.798	0.770	0.717
BEV-Planner++ (Li et al., 2024)	0.250	0.251	0.248	0.252	0.251	0.248	0.249	0.246
Ego-MLP (Li et al., 2024)	0.243	0.242	0.242	0.244	0.242	0.242	0.239	0.235
OccWorld (Zheng et al., 2024)	0.191	0.190	0.190	0.194	0.191	0.190	0.190	0.189
ResWorld (Zhang et al., 2025)	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035
WorldEngine (RLFT) (Li et al., 2026)	0.291	0.291	0.289	0.291	0.290	0.289	0.280	0.279
Adversarial								
GT-Following†	0.814	0.792	0.767	0.810	0.791	0.743	0.679	0.627
BEV-Planner++ (Li et al., 2024)	0.250	0.245	0.243	0.249	0.244	0.243	0.236	0.225
Ego-MLP (Li et al., 2024)	0.065	0.065	0.065	0.065	0.065	0.065	0.064	0.064
OccWorld (Zheng et al., 2024)	0.188	0.188	0.187	0.191	0.188	0.182	0.177	0.168
ResWorld (Zhang et al., 2025)	0.035	0.035	0.035	0.035	0.035	0.035	0.034	0.034
WorldEngine (RLFT) (Li et al., 2026)	0.284	0.279	0.276	0.284	0.271	0.256	0.239	0.212