Appendix

Additional Implementation Details

Model Input and Architecture. We adopt a unified architecture for both teacher and student models. Each traffic scene contains up to $A = 32$ agents, $M_l = 256$ map polylines (each with $M_p = 30$ waypoints), and $M_t = 16$ traffic lights. The model predicts $T = 80$ future control steps at 0.1s resolution, conditioned on the current agent state and optionally a 1-second history window ($T_h = 11$). The scene encoder comprises $L_E = 6$ query-centric Transformer layers with embedding dimension $D = 256$, while the behavior predictor includes $L_P = 4$ cross-attention layers that output 64 candidate trajectories with probabilities. The diffusion denoiser contains four Transformer layers organized into two decoding blocks. To reduce computation, action sequences are downsampled to $T_f = 40$ steps by repeating each predicted action over two time steps.

Diffusion Process and Optimization. We use a log variance schedule with $K = 50$ diffusion steps, $\bar{\alpha}_{\min} = 10^{-9}$, and $\delta = 0.0031$. The model is trained using a composite loss that includes denoising loss L_D and behavior prediction loss L_P, with balancing weights $\gamma = 0.5$ and $\beta = 0.05$, respectively. Optimization is performed with AdamW (weight decay = 0.01), an initial learning rate of 2e-4, exponential decay factor 0.02 per 1,000 steps, and linear warm-up over the first 1,000 steps. Gradient clipping is applied with a norm threshold of 1.0. All models are trained in bfloat16 mixed precision on 4 NVIDIA A800 GPUs, using a batch size of 4 per device (effective batch size = 16). Phase I and Phase II training are conducted for 16 and 6 epochs, respectively.

Student Training and Distillation Scheduling. The student model follows the same architecture and optimization settings as the teacher but incorporates additional hybrid self-distillation losses: feature-level (L_F), response-level (L_R), and contrastive-level (L_C), as introduced in Section 5. During Phase II continual adaptation, the student is fine-tuned using synthetic scenarios generated from a small batch of target domain samples, while the teacher remains frozen. To balance task performance and distillation supervision over time, we apply a simulated annealing schedule to λ_{SA}, linearly decaying it from 0.8 to 0.2 over 40,000 steps. This ensures strong guidance early in training and increasing autonomy as the student adapts.

Implementation of RlSD Baseline

RlSD captures the structural dependencies among agents by transferring the teacher model's inter-agent relational knowledge to the student. Specifically, we model multi-agent interaction by computing pairwise differences in normalized denoised control sequences, forming a latent interaction graph where each edge encodes directional influence between agents. For all $G = \binom{N}{2}$ agent pairs in a scenario, we compute the action difference vectors as follows:

$$\Delta \mathbf{a}_s^{i,j} = \hat{\mathbf{a}}_{k,S}^i - \hat{\mathbf{a}}_{k,S}^j, \quad \Delta \mathbf{a}_t^{i,j} = \hat{\mathbf{a}}_{k,T}^i - \hat{\mathbf{a}}_{k,T}^j$$

where $\hat{\mathbf{a}}_{k,T}^i = \mathcal{D}_{\theta_T}(\tilde{\mathbf{a}}_k, k, \mathbf{m})$ and $\hat{\mathbf{a}}_{k,S}^i = \mathcal{D}_{\theta_S}(\tilde{\mathbf{a}}_k, k, \mathbf{m})$ denote the denoised control sequences for agent i from the teacher and student models, respectively, at diffusion step k.

To enforce relational consistency, we define the RlSD loss as the average cosine similarity loss across all agent pairs:

$$\mathcal{L}_{\text{RlSD}} = \mathbb{E}_{\tilde{\mathbf{a}}_k, \mathbf{m}, k}\left[\frac{1}{G}\sum_{i \neq j}\left(1 - \cos\left(\Delta \mathbf{a}_t^{i,j}, \Delta \mathbf{a}_s^{i,j}\right)\right)\right]$$

This objective encourages the student to match the teacher's representation of inter-agent relational dynamics, such as relative motion, yielding, and lane-changing behaviors. RlSD enhances generalization by preserving relational structure during simulation, especially in novel scenarios with complex multi-agent interactions.

Computation Details of Evaluation Metrics

Open-loop Trajectory Generation Evaluation Metrics and Qualitative analysis Trajectory generation performance is primarily evaluated using two widely adopted metrics that assess the realism and accuracy of predicted trajectories: *minimum Average Displacement Error* (minADE) and *minimum Final Displacement Error* (minFDE). These metrics measure the spatial deviation between predicted and ground-truth trajectories, and are used in both training and validation stages on WOMD.

Minimum Average Displacement Error (minADE): minADE evaluates the average deviation between predicted and ground-truth trajectories over the entire prediction horizon. For each agent, the model generates M multimodal trajectories, and the error is computed based on the sample that is closest to the ground-truth in terms of Euclidean distance. The minADE is defined as:

$$\min \text{ADE} = \frac{1}{N}\sum_{n=1}^{N}\min_{m \in \{1,\ldots,M\}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\tau}_m^n(t) - \tau^n(t)\right\|_2\right),$$

where $\hat{\tau}_m^n(t)$ denotes the position vector of the n-th agent for the m-th predicted trajectory at time step t, $\tau^n(t)$ represents the corresponding ground-truth position vector, $\|\cdot\|_2$ is the Euclidean norm, T is the number of predicted time steps, and N is the total number of agents. This metric quantifies the average spatial discrepancy across the entire trajectory, emphasizing the model's ability to produce predictions that closely align with the ground-truth over time.

Minimum Final Displacement Error (minFDE): minFDE focuses on the deviation at the final time step of each predicted trajectory, measuring the accuracy of trajectory endpoints. This metric is especially important for long-horizon prediction tasks where endpoint accuracy critically impacts downstream decision-making. The minFDE is computed as:

$$\min \text{FDE} = \frac{1}{N}\sum_{n=1}^{N}\min_{m \in \{1,\ldots,M\}}\left\|\hat{\tau}_m^n(T) - \tau^n(T)\right\|_2,$$

where $\hat{\tau}_m^n(T)$ and $\tau^n(T)$ denote the position vectors of the n-th agent for the m-th predicted trajectory and the ground-truth trajectory, respectively, at the final time step T. The metric evaluates the model's precision in predicting the final position of an agent, which is crucial for ensuring safe and effective navigation in dynamic environments.

Qualitative analysis To further illustrate the generative capability of the proposed CDPT model in multi-agent traffic simulation, Figure 1 presents qualitative trajectory generation results on the Waymo Open Motion Dataset (WOMD). The generated trajectories exhibit smoothness and strong adherence to traffic rules. In particular, CDPT accurately captures road topology constraints (e.g., lane boundaries) and dynamic agent interactions (e.g., yielding behavior) in densely interactive regions such as intersections and lane-changing zones. These results suggest the model's strong ability to reflect both geometric and social constraints in complex urban settings.

Closed-Loop Simulation Evaluation Metrics and Qualitative analysis To comprehensively evaluate the closed-loop simulation quality of our model and baselines, we adopt a suite of standard metrics that quantify safety, rule compliance, physical plausibility, and behavioral realism. These metrics are evaluated over all agents and simulation rollouts, and align with recent autonomous driving benchmarks such as Waymax and INTERACTION.

Off-road: A vehicle is considered off-road if its position crosses to the left of the directed road edge. This binary metric is computed as the percentage of vehicles that transition from on-road to off-road across all simulated agents and scenarios:

$$O_{\text{road}}(\%) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s T}\sum_{n=1}^{N_s}\sum_{t=1}^{T}\mathbb{I}\left(\tau^n(t) \notin \mathcal{R}_s\right) \times 100,$$

where $\tau^n(t)$ is the position of the n-th agent at time step t, $\mathcal{R}_s$ represents the drivable road region for scenario s, $\mathbb{I}(\cdot)$ is the indicator function returning 1 if the condition is true and 0 otherwise, N_s is the number of agents in scenario s, T is the number of time steps, and S is the total number of scenarios. This metric assesses the model's ability to keep vehicles within designated road boundaries, a critical aspect of safe autonomous driving.

Collision: Collisions are detected when the 2D bounding boxes of two agents overlap at the same time step. The collision rate is the proportion of agents involved in any collision across all simulations:

$$C_{\text{coll}}(\%) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s T}\sum_{n=1}^{N_s}\sum_{t=1}^{T} \mathbb{I}\left(\exists j \neq n \text{ or env}, \left\|\tau^n(t) - \tau^j(t)\right\|_2 < \epsilon\right) \times 100,$$

where $\tau^n(t)$ and $\tau^j(t)$ are the positions of agents n and j (or the environment) at time step t, ϵ is a threshold defining the minimum distance for a collision, and other terms are as previously defined. This metric evaluates the model's capability to avoid collisions, a fundamental requirement for ensuring safety in multi-agent environments.

Wrong-way: An agent is flagged as driving in the wrong direction if its heading deviates more than 90 degrees from the closest lane direction for over 1 second. This metric reflects violations of traffic rules:

$$W_{\text{way}}(\%) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s T}\sum_{n=1}^{N_s}\sum_{t=1}^{T} \mathbb{I}\left(\angle(\mathbf{v}^n(t), \mathbf{d}_s(t)) > \theta\right) \times 100,$$

where $\mathbf{v}^n(t)$ is the velocity vector of the n-th agent at time step t, $\mathbf{d}_s(t)$ is the direction vector of the closest lane in scenario s, $\angle(\cdot,\cdot)$ denotes the angle between two vectors, and θ is the threshold angle (typically 90 degrees). This metric measures adherence to traffic flow rules, ensuring that agents navigate in alignment with road regulations.

Kinematic Infeasibility: A trajectory is considered physically infeasible if it exceeds the acceleration threshold ($6\,\text{m/s}^2$) or curvature limit ($0.3\,\text{m}^{-1}$). The rate is calculated as the percentage of agents violating these constraints:

$$K_{\text{inf}}(\%) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s T}\sum_{n=1}^{N_s}\sum_{t=1}^{T} \mathbb{I}\left(\left\|\mathbf{a}^n(t)\right\|_2 > a_{\max} \text{ or } \kappa^n(t) > \kappa_{\max}\right) \times 100,$$

where $\mathbf{a}^n(t)$ is the acceleration of the n-th agent at time step t, $\kappa^n(t)$ is the curvature of the trajectory, $a_{\max} = 6\,\text{m/s}^2$, $\kappa_{\max} = 0.3\,\text{m}^{-1}$, and other terms are as defined previously. This metric ensures that generated trajectories are physically realistic and executable by real-world vehicles.

Log Divergences: Log Divergences measure the average discrepancy between predicted and ground-truth trajectories across all time steps in a closed-loop simulation, providing insight into the overall realism of the simulated behavior. This metric is particularly useful for evaluating the cumulative error in multi-agent interactions over extended simulation horizons. The Log Divergences metric is defined as:

$$L_{\text{div}}(m) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{n=1}^{N_s}\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\tau}^n(t) - \tau^n(t)\right\|_2,$$

where $\hat{\tau}^n(t)$ represents the predicted position of the n-th agent at time step t for the m-th simulation rollout, $\tau^n(t)$ is the corresponding ground-truth position, and other terms are as defined previously. This metric quantifies the average Euclidean distance between predicted and actual trajectories, offering a comprehensive assessment of the model's ability to replicate realistic agent behavior in closed-loop settings.

Qualitative analysis To further assess the realism and safety of multi-agent behaviors generated by CDPT, we conduct a qualitative comparison with the teacher model across several representative closed-loop scenarios in the Waymax simulation environment. These cases cover common but

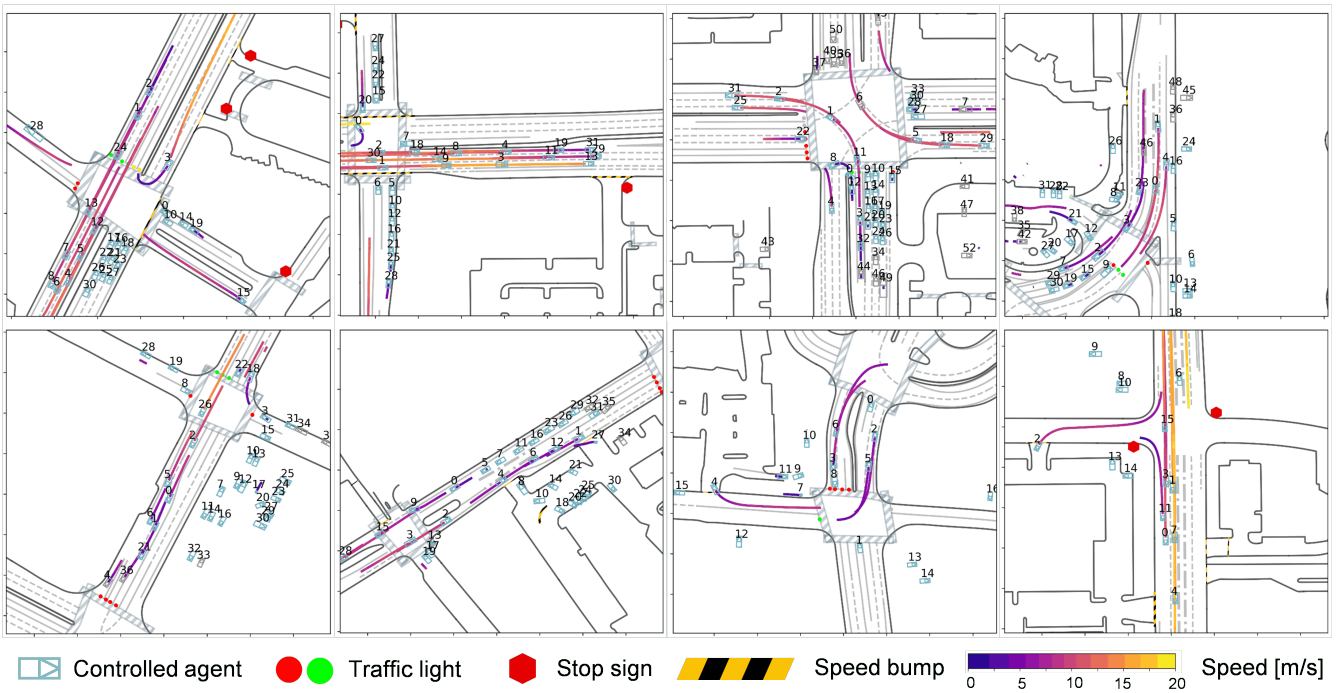

Figure 1: Qualitative analysis and visualization of the CDPT model's open-loop trajectory generation in WOMD.

challenging driving contexts, including intersections, highway merges, and unsignalized junctions, and serve to illustrate the advantages of CDPT in generating human-like, regulation-compliant, and interaction-aware trajectories.

In the highway merging case (Figure 2), vehicle 11 merges onto a multi-lane road. The teacher model causes collisions due to abrupt lane changes and generates erratic behavior from vehicle 2, which crashes into a building. CDPT produces safer behavior: vehicle 11 merges smoothly without collisions, and vehicle 2 follows a valid, stable trajectory. This demonstrates CDPT's advantage in multi-agent coordination and spatial consistency.

In the T-junction scenario (Figure 3), vehicle 5 fails to act under the teacher model despite safe gaps, while vehicle 2 exhibits severe anomalies such as wrong-way driving. CDPT generates more realistic behavior: vehicle 5 executes a timely left turn, and vehicle 2 maintains proper lane discipline. This reflects stronger decision-making and interaction modeling under ambiguity.

In the intersection case (Figure 4), the teacher model frequently violates traffic rules, entering intersections without stopping or yielding. CDPT corrects these issues, producing signal-compliant, smooth trajectories with realistic stop-and-go behavior. These results confirm CDPT's improved rule adherence and environment awareness.

OnSite Benchmark for AV testing

To evaluate the deployability of scenarios generated by CDPT in autonomous vehicle (AV) testing, we adopt the standardized OnSite benchmark (Open Natural Driving Intelligence Automotive Simulation Test Environment). Each task instance provides a static high-definition (HD) map and the initial 31 frames of agent states and trajectories. Without any domain-specific fine-tuning, CDPT is deployed in a zero-shot manner and successfully generates over 800 distinct multi-agent traffic scenarios. These scenarios span a wide range of complex and realistic driving contexts, including highway mainlines, roundabouts, mixed traffic involving vulnerable road users, urban intersections, and highway merge areas.

Evaluation Setup: AV Planners Under Test To assess closed-loop realism and interaction complexity, we integrate five mature trajectory planners as the autonomous vehicle under test (AVUT). These planners vary in architecture and control logic, providing a representative performance envelope across scenario types:

- **IDM (Intelligent Driver Model):** The IDM planner is a classical model widely used for simulating human-like driving behaviors, particularly in car-following and lane-changing contexts. It computes acceleration based on inter-vehicle distance and relative speed, making it simple yet effective for generating realistic driving responses. Its interpretability and robustness make it a strong baseline for AV testing.
- **HMAPlanner (Hierarchical Multi-Agent Planner):** HMAPlanner employs a hierarchical architecture tailored for multi-agent interaction across distinct traffic scenarios. It features a modular design with dedicated components for merges, roundabouts, and intersections, each leveraging scenario-specific logic. The system prioritizes efficient coordination among agents while main-

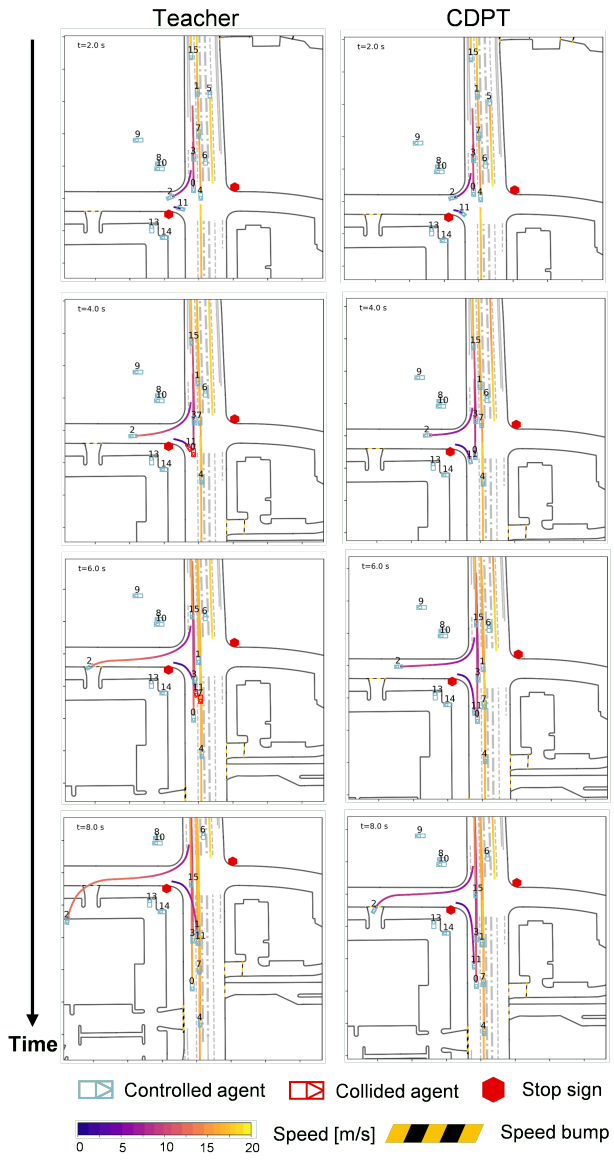

Figure 2: Closed-loop simulation qualitative analysis example 1 in WOMD.

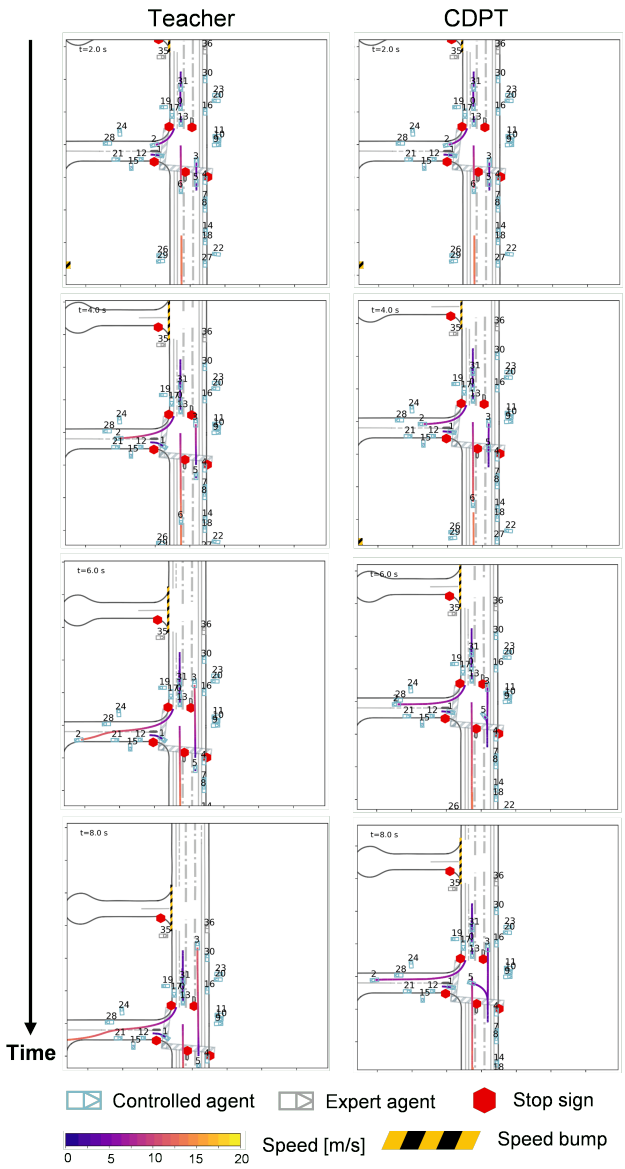

Figure 3: Closed-loop simulation qualitative analysis example 2 in WOMD.

taining computational performance through compiled implementation.

- **CBDES (Comprehensive Behavior Decision and Execution System):** CBDES integrates behavior prediction, global and local planning, and safety-critical control into a cohesive four-layer framework. It leverages rule-based prediction for surrounding agents, map-aware route planning, and adaptive local decision-making. The inclusion of control barrier functions ensures that the executed trajectories comply with safety constraints in dynamic environments.
- **WJHPP (WJ High-Performance Planner):** This high-performance planner is built for scalability and speed, featuring a modular structure optimized through compilation. It supports a wide range of traffic scenarios via parameterized configurations and implements a rule-based longitudinal behavior model. Its flexibility and efficiency make it suitable for real-time testing in complex traffic environments.
- **FQPPlanner (Frenet-QP Planner):** FQPPlanner employs a decoupled path-speed planning strategy within the Frenet coordinate framework. It performs global path planning using A* search and optimizes the speed profile through quadratic programming, resulting in dynamically feasible and smooth trajectories. Its modular design and high precision make it well-suited for structured environments such as highways and arterial roads.

Figure 4: Closed-loop simulation qualitative analysis example 3 in WOMD.

Evaluation Metric for AV Testing To rigorously assess the realism and testing utility of scenarios generated by CDPT, we adopt a quantitative scoring framework aligned with the OnSite benchmark. The evaluation comprises two main categories—**Scene Realism** and **Effectiveness of AV Testing**—with a total score of 100 points. Each subcomponent captures a specific aspect of the simulation's quality or its capacity to test the autonomous vehicle under test (VUT) under challenging conditions.

A. Scene Realism. This category quantifies the physical plausibility and behavioral fidelity of the generated agents using three criteria: safety, distributional consistency, and comfort.

- **Safety Score (30 points).** Let N_{s} be the number of collision-free scenes and N_{t} the total number of test scenes. The safety score is:

$$SS = \frac{N_{\mathrm{s}}}{N_{\mathrm{t}}} \times 30,$$

- **Kinetic distribution consistency (20 points).** Using Jensen–Shannon (JS) divergence between generated and real-world distributions of velocity (v), acceleration (a), and angular velocity (ω), define:

$$KC_v = (1 - \mathrm{JS}(v)) \times \frac{20}{3},$$
$$KC_a = (1 - \mathrm{JS}(a)) \times \frac{20}{3},$$
$$KC_\omega = (1 - \mathrm{JS}(\omega)) \times \frac{20}{3},$$

  The total distribution consistency score:

$$KC = KC_v + KC_a + KC_\omega,$$

- **Comfort Score (Deduction up to 10 points).** Let T_{e} be the total time agents exceed comfort thresholds on any of six dynamic quantities: longitudinal acceleration a_x, lateral acceleration a_y, longitudinal jerk j_x, lateral jerk j_y, yaw rate r, and yaw jerk. Let T_{to} be the total trajectory duration. The comfort penalty is:

$$CS = (1 - \frac{T_{\mathrm{e}}}{T_{\mathrm{to}}}) \times 10,$$

B. Effectiveness of AV Testing. This category evaluates how well generated scenarios differentiate AV planner performance compared to real data.

- **Test Effectiveness (40 points)**
  Let S_{GT} and S_{Gen} be normalized VUT performance scores on real and generated scenarios respectively (e.g., safety, comfort violations). Define the performance gap:

$$\Delta S = S_{\mathrm{GT}} - S_{\mathrm{Gen}},$$

  The bounded scoring function $f(\cdot) : \mathbb{R} \to [0, 1]$ normalizes this gap as:

$$f(\Delta S) = \begin{cases} 0, & \Delta S \leq 0, \\ \frac{\Delta S}{\Delta S_{\max}}, & 0 < \Delta S < \Delta S_{\max}, \\ 1, & \Delta S \geq \Delta S_{\max}, \end{cases}$$

  where $\Delta S_{\max}$ is a predefined maximum threshold.
  Then, the test effectiveness score is computed by scaling the normalized value to the full point range:

$$TE = f(\Delta S) \times 40,$$

C. Total Score. The overall evaluation score TS is:

$$TS = SS + KC + CS + TE,$$

where CS is a penalty (negative) and others are positive contributions.

This composite metric balances physical realism and the practical challenge posed to AV planners, providing a comprehensive measure of scenario generation quality.

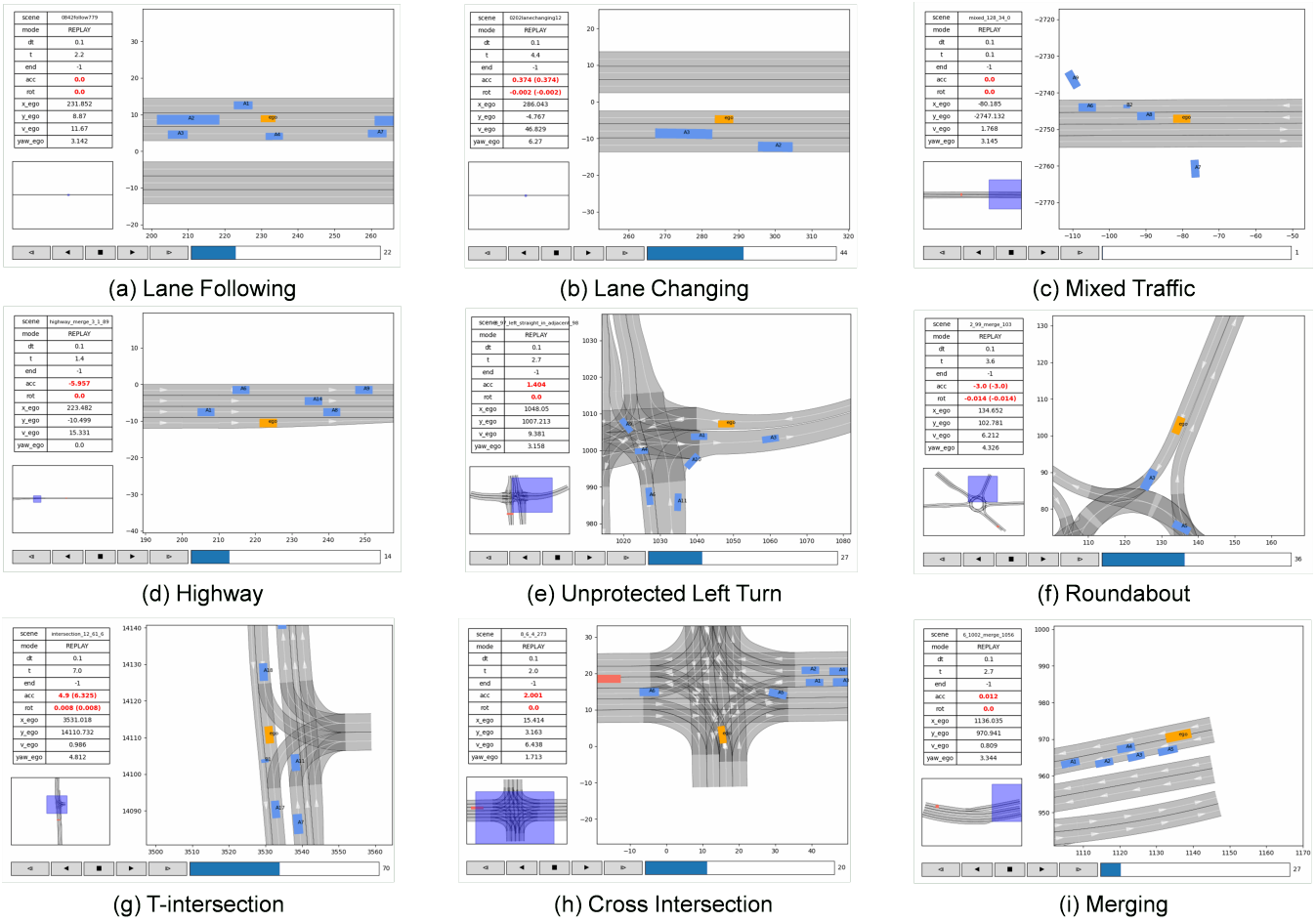

Figure 5: Representative scenarios generated by CDPT in OnSite benchmark (Orange vehicle is the tested AV, and the other vehicles are controlled by CDPT).

Qualitative Analysis of CDPT on the OnSite Benchmark Figure 5 showcases closed-loop replay results across diverse and challenging traffic scenarios, demonstrating that CDPT can generate high-fidelity, multi-agent scenes with realistic interactions in varied environments such as highways, urban intersections, merges, and roundabouts. The ego agents (in orange) exhibit plausible and coordinated behaviors with surrounding traffic (in blue), maintaining trajectory smoothness and respecting lane geometry even in dense and complex settings. These results confirm that CDPT produces diverse, controllable, and predictable simulations, offering a robust foundation for evaluating autonomous driving systems under realistic and safety-critical conditions.