

Article

PriorAug: Reliability-Aware Traffic Scenario Initialization Under Sensor-Perceived Structural and Interaction Priors

Kaixi Yang * and Shiru Qu

School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

* Correspondence: kaixiyang@mail.nwpu.edu.cn

Abstract

Traffic scenario initialization is a key component of autonomous driving simulation. The generated initial traffic states directly affect downstream motion prediction, planning evaluation, closed-loop simulation, and safety assessment. Existing traffic scenario generators usually rely on structural and interaction priors, but these priors are commonly treated as deterministic and fully reliable during both training and deployment. In practical autonomous driving systems, however, such priors are reconstructed from heterogeneous perception pipelines involving sensors, localization, tracking, online mapping, map alignment, and behavior estimation. Their reliability may therefore vary continuously because of sensing noise, localization drift, incomplete observations, occlusion, tracking instability, and imperfect interaction inference. To address this issue, this paper proposes PriorAug, a reliability-aware prior augmentation framework for robust traffic scenario initialization under sensor-perceived prior uncertainty. PriorAug formulates traffic scenario initialization as conditional generation over sensor-perceived priors and introduces a two-dimensional reliability space to characterize structural reliability and interaction reliability separately. Based on this space, reliability-conditioned prior operators are designed to construct diverse prior configurations during training, enabling the generator to learn from continuously varying perception conditions without modifying the underlying generation architecture or introducing additional learnable inference modules or iterative decoding stages. Experiments on the Waymo Open Motion Dataset show that PriorAug maintains competitive full-set distribution fidelity, improves robustness under shifted-prior conditions and traffic density variations, and improves the physical plausibility of generated initial states. Reliability landscape analysis further demonstrates that coverage-oriented prior augmentation yields smoother performance transitions across the reliability space. These results indicate that explicitly modeling sensor-perceived prior reliability provides an effective and practical route to robust traffic scenario initialization for autonomous driving simulation.

Keywords: autonomous driving simulation; traffic scenario initialization; sensor-perceived priors; structural reliability; interaction reliability; prior augmentation; Waymo Open Motion Dataset; intelligent transportation systems



Academic Editor: Linheng Li

Received: 1 July 2026

Revised: 18 July 2026

Accepted: 27 July 2026

Published: 2 August 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

High-fidelity traffic scenario initialization is a fundamental component of autonomous driving simulation. The initialized traffic state determines the spatial layout, motion attributes, and interaction context of surrounding traffic participants, and therefore directly affects downstream motion prediction, planning evaluation, closed-loop simulation, and safety assessment. Compared with manually designed scenarios, data-driven traffic

scenario generation can learn from large-scale autonomous driving datasets and provide diverse traffic states with stronger statistical realism. Recent studies have developed various learning-based generation frameworks, including traffic scene generation, unified initial-state-and-trajectory modeling, controllable diffusion-based generation, and language-conditioned traffic generation [1–5]. These methods have improved the realism, diversity, and controllability of simulation scenarios.

Most existing traffic scenario initialization methods rely on structural and interaction priors to constrain generation. Structural priors, such as lane geometry, road topology, road boundaries, and driving directions, provide spatial constraints for vehicle placement and motion attributes. Interaction priors, such as following, yielding, merging, overtaking, and collision-avoidance relationships, describe behavioral dependencies among agents. By conditioning generation on these priors, traffic generators can produce more realistic initial traffic states than unconstrained generation. However, existing methods often treat these priors as deterministic, complete, and fully reliable during both training and deployment. It should be emphasized that this work does not question the necessity of structural and interaction information for autonomous driving. Instead, it focuses on the reliability variation of these priors after they are reconstructed by upstream perception and scene-construction modules. The deterministic-prior assumption is difficult to satisfy in practical autonomous driving systems. In deployment, the priors available to a traffic generation module are not perfectly annotated ground-truth inputs. Instead, they are reconstructed from upstream perception and scene-construction processes that may involve cameras, LiDARs, radars, localization, object detection, multi-object tracking, online mapping, map alignment, and behavior estimation. High-definition maps and online map representations may suffer from incompleteness, localization drift, road changes, registration errors, and update latency [6–9]. Vectorized scene representations and lane-graph encodings used by downstream models also depend on the quality of perceived road structures and agent states [10–12]. Therefore, the priors used by traffic scenario initialization are more appropriately regarded as *sensor-perceived priors* rather than deterministic ground-truth priors.

The uncertainty of sensor-perceived priors is not uniform. Structural information and interaction information originate from different components of the perception and inference process and may vary independently. Structural priors are mainly affected by sensing noise, localization drift, map alignment errors, incomplete observations, and road structure reconstruction quality. Interaction priors are inferred from dynamic object states and observed trajectories, making them sensitive to occlusion, tracking instability, traffic density, behavioral ambiguity, and imperfect interaction estimation. Thus, prior uncertainty should not be reduced to a single binary condition such as map available versus map unavailable. A more expressive formulation is needed to characterize the continuous and heterogeneous reliability changes of perceived structural and interaction information.

Such reliability variation creates an important source of robustness degradation. A traffic generator trained under one fixed prior configuration may perform well near its native operating point, but its generation quality can deteriorate when the reliability of structural or interaction priors shifts. This issue is particularly important for traffic scenario initialization, because initialization errors may propagate to downstream simulation and lead to unrealistic vehicle placement, distorted motion attributes, or physically invalid overlapping states. Although previous studies have improved generation fidelity under fixed or idealized input conditions, the effect of continuously changing sensor-perceived prior reliability remains underexplored.

This perspective is related to domain randomization and robust learning. Domain randomization improves deployment robustness by exposing a model to randomized

environmental or simulation conditions during training. Robust reinforcement learning similarly studies how policies can remain stable under uncertain dynamics, perturbed observations, or distributional changes. PriorAug follows the same broad principle of training under diversified conditions, but adapts it to traffic scenario initialization rather than policy learning. The randomized factors in PriorAug are not simulator physics or control dynamics; they are the reliabilities of sensor-perceived structural and interaction priors. This distinction motivates the reliability space formulation used in this paper.

To address this problem, this paper proposes PriorAug, a reliability-aware prior augmentation framework for robust traffic scenario initialization. Instead of assuming deterministic structural and interaction priors, PriorAug explicitly models structural reliability and interaction reliability using two continuous variables, denoted by r_s and r_i , respectively. These variables define a two-dimensional reliability space in which each point corresponds to a specific sensor-perceived prior condition. Based on this formulation, reliability-conditioned prior operators are designed to adjust structural and interaction information independently during training. By exposing the generator to diverse reliability configurations, PriorAug improves robustness under shifted-prior conditions without modifying the underlying generation architecture or introducing additional learnable inference modules or iterative decoding stages.

The main contributions of this paper are summarized as follows:

- We formulate traffic scenario initialization as a reliability-aware conditional generation problem over sensor-perceived priors, addressing the mismatch between deterministic-prior training assumptions and uncertain perception conditions in practical autonomous driving systems.
- We introduce a two-dimensional reliability space that separately characterizes structural reliability and interaction reliability, enabling different sources of prior uncertainty to be modeled explicitly.
- We propose PriorAug, a reliability-conditioned prior augmentation framework that constructs diverse prior configurations through structural and interaction prior operators, improving robustness under continuously varying perceived-prior conditions.
- We evaluate the proposed framework on the Waymo Open Motion Dataset under a unified protocol. Experimental results show that PriorAug maintains competitive full-set distribution fidelity, improves robustness under shifted-prior and traffic density variations, and improves the physical plausibility of generated initial states.

2. Related Work

2.1. Traffic Scenario Generation with Structured Conditions

Traffic scenario generation aims to construct realistic and diverse traffic states for autonomous driving simulation. Learning-based approaches have been widely studied because they can capture statistical regularities from large-scale driving datasets. SceneGen learns realistic traffic scene layouts from recorded driving data [1], while TrafficGen provides a representative framework for structured traffic scenario generation [2]. More recent methods extend traffic generation toward joint initialization and trajectory modeling, controllable generation, and semantic control. For example, UniGen jointly models initial agent states and future trajectories [3], SceneControl introduces diffusion-based controllable traffic scene generation [4], and language-conditioned generation explores textual or semantic guidance for traffic generation [5].

These studies show that structured conditioning information is essential for high-quality traffic scenario generation. However, the conditioning inputs used by existing generators are usually treated as reliable once they are provided to the model. In practical autonomous driving systems, such inputs are produced by upstream perception, localiza-

tion, tracking, map-processing, and behavior estimation modules. Their reliability may vary with sensing quality, localization accuracy, observation completeness, and traffic complexity. Therefore, traffic scenario generation should not only consider what prior information is available, but also how reliable this information is under deployment conditions.

2.2. Sensor-Perceived Structural and Interaction Information

Structural information provides road-level constraints for autonomous driving perception, prediction, planning, and simulation. High-definition maps and online map representations provide lane geometry, road topology, road boundaries, and driving directions, but they may be affected by incompleteness, update latency, localization drift, registration errors, and maintenance cost [6–8]. Road change detection and HD map registration studies further indicate that aligning perceived road structures with prior maps remains challenging under real-world conditions [9]. In learning-based autonomous driving models, vectorized scene representations and lane-graph encodings are also widely used to represent structured road information [10–12]. These representations demonstrate the importance of structural priors, but their quality still depends on perception and scene-construction reliability.

Interaction information provides another important source of traffic context. Multi-agent behavior is shaped not only by road structure but also by dynamic relationships among surrounding participants. Social GAN models socially acceptable trajectories through generative adversarial learning [13], Trajectron++ represents heterogeneous multi-agent interactions in a dynamically feasible prediction framework [14], and AgentFormer captures socio-temporal dependencies using agent-aware transformers [15]. Although these studies demonstrate the value of interaction modeling, interaction cues in practical systems are inferred from detected objects, tracked trajectories, and behavior estimation modules rather than directly observed as ground-truth. Therefore, interaction information may have reduced reliability under occlusion, dense traffic, tracking instability, and behavioral ambiguity.

These observations motivate separating structural reliability from interaction reliability. Structural priors and interaction priors originate from different upstream processes and may fail in different ways. A traffic generator that assumes both types of information to be fully reliable may become sensitive to reliability shifts during deployment. This paper explicitly models structural and interaction reliability as two independent dimensions of sensor-perceived prior uncertainty.

2.3. Robust Generation Under Prior Reliability Shift

Existing traffic generation methods commonly improve robustness through larger datasets, stronger conditional encoders, or richer contextual representations. However, their structural and interaction conditions are generally treated as reliable once provided to the generator, and few studies explicitly characterize how generation quality changes when the reliability of these conditions shifts. This limitation is important for traffic scenario initialization, because errors in perceived structure or interaction may directly affect generated positions, headings, speeds, sizes, and geometric feasibility. PriorAug addresses this gap by independently controlling structural and interaction reliability and training the generator over a two-dimensional reliability space rather than at one deterministic operating point.

2.4. Domain Randomization and Robust Learning

The training strategy of PriorAug is related to the broader methodological paradigms of domain randomization and robust learning. Domain randomization has been widely used to improve sim-to-real transfer by randomizing visual, environmental, or physical

parameters during training, so that the learned model is less likely to over-specialize to a single source-domain condition [16–18]. Related ideas have also been explored in robust reinforcement learning, where policies are trained or optimized to remain effective under uncertain dynamics, model perturbations, observation noise, adversarial disturbances, or distribution shifts [19,20]. The common principle behind these methods is to improve deployment robustness by exposing the model to a family of possible operating conditions during training rather than to one fixed condition.

PriorAug follows this general principle, but adapts it to the specific problem of traffic scenario initialization under sensor-perceived prior uncertainty. Unlike conventional domain randomization methods that randomize simulator textures, physical parameters, or transition dynamics, PriorAug randomizes the reliability of the structural and interaction priors used by a traffic generator. Unlike robust reinforcement learning, PriorAug does not learn a control policy or optimize closed-loop behavior under uncertain dynamics. Instead, it trains a conditional traffic initialization generator to remain robust when the reliability of its sensor-perceived prior inputs changes.

The specific novelty of PriorAug lies in its reliability space formulation. Rather than treating prior perturbation as an unstructured input dropout or a single degraded condition, PriorAug defines a two-dimensional reliability space that separately characterizes structural reliability and interaction reliability. This formulation allows road structure information and agent-interaction information to be controlled independently during training. Therefore, PriorAug can be viewed as a reliability-aware adaptation of the domain-randomization principle to traffic scenario initialization, where the randomized factors are not generic simulator conditions but the reliability of sensor-perceived priors.

3. Proposed Method

This section presents the proposed PriorAug framework. We first define the sensor-perceived scene graph used by the traffic generator. We then introduce a two-dimensional reliability space to characterize structural and interaction reliability. Based on this space, reliability-conditioned prior operators are designed to construct diverse perceived-prior configurations during training. Finally, we describe the reliability-aware generator and the PriorAug training strategy.

Figure 1 summarizes the overall reliability-aware training framework. The upstream scene-construction process provides a sensor-perceived scene graph. The reliability-conditioned prior operators then adjust structural and interaction information according to sampled reliability variables. The resulting scene graph is fed into the traffic generator, enabling robust traffic scenario initialization under continuously varying sensor-perceived prior conditions.

3.1. Sensor-Perceived Scene Graph

Traffic scenario initialization relies on structured priors that describe road geometry and traffic participant relationships. In practical autonomous driving systems, such priors are not directly obtained as perfect ground-truth inputs. Instead, they are produced by upstream perception and scene-construction processes involving sensor observation, localization, object detection, tracking, map alignment, online map construction, and behavior estimation. This paper does not explicitly model the full perception network itself. Rather, we abstract its downstream output as a sensor-perceived prior and study how reliability variations in this prior affect traffic scenario initialization.

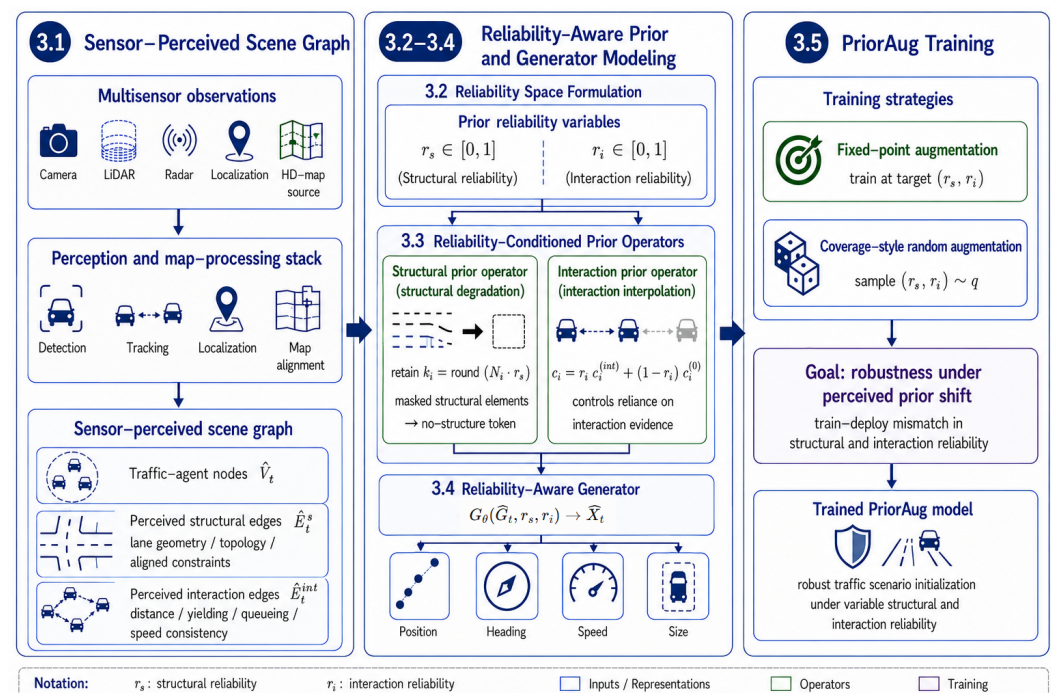


Figure 1. Overall framework of the proposed PriorAug. The framework consists of sensor-perceived scene-graph construction (Section 3.1), reliability space formulation and reliability-conditioned structural and interaction prior modeling (Sections 3.2–3.4), and fixed-point or coverage-oriented reliability augmentation during training (Section 3.5).

Let

$$P_t = \Phi(\mathcal{O}_t), \quad \mathcal{O}_t = \{I_t, L_t, R_t, \ell_t\}, \quad (1)$$

where I_t , L_t , R_t , and ℓ_t denote camera images, LiDAR measurements, radar measurements, and localization information, respectively. Here, $\Phi(\cdot)$ abstracts the upstream perception and scene-construction process; the present work studies its structured output rather than implementing the complete perception pipeline.

The abstract sensor-perceived prior P_t is represented in the subsequent formulation as the scene graph G_t .

We represent the sensor-perceived prior as a scene graph:

$$G_t = (V_t, E_t^s, E_t^i), \quad (2)$$

where V_t denotes traffic-agent nodes, E_t^s denotes structural edges, and E_t^i denotes interaction edges. Each node $v_k \in V_t$ contains the state attributes of one traffic participant, including position, heading, speed, size, and semantic category. Structural edges encode perceived road-level constraints such as lane geometry, road topology, driving direction, and local road connectivity. Interaction edges encode inferred behavioral relationships among agents, such as following, yielding, merging, overtaking, and collision avoidance.

Because both E_t^s and E_t^i are derived from perceived scene information, they may be incomplete, noisy, or partially unreliable under practical sensing conditions. This motivates modeling their reliability explicitly rather than treating the scene graph as a deterministic input.

3.2. Reliability Space Formulation

The uncertainty of sensor-perceived priors does not affect all information equally. Structural information and interaction information originate from different upstream pro-

cesses and therefore exhibit different reliability characteristics. To model this difference, we introduce two continuous reliability variables.

Structural reliability. Structural reliability measures the completeness and correctness of perceived road structure information. It is denoted by

$$r_s \in [0, 1], \quad (3)$$

where $r_s = 1$ indicates fully reliable structural information, and $r_s = 0$ indicates that structural information is unavailable.

Interaction reliability. Interaction reliability measures the confidence of inferred behavioral relationships among traffic participants. It is denoted by

$$r_i \in [0, 1], \quad (4)$$

where larger values indicate more reliable interaction information.

Together, the two variables define a two-dimensional reliability space:

$$\mathcal{R} = [0, 1] \times [0, 1]. \quad (5)$$

Each point $(r_s, r_i) \in \mathcal{R}$ corresponds to one sensor-perceived prior condition. Methods relying mainly on structural priors can be interpreted as operating near high structural reliability and low interaction reliability, while methods using both structural and interaction priors operate near the upper-right region of the reliability space. In practical deployment, however, sensing quality, localization accuracy, tracking stability, and behavior estimation confidence may change over time, causing the operating condition to move continuously within this space.

Figure 2 illustrates the proposed reliability space. Instead of optimizing traffic generation for a single deterministic operating point, the objective of this work is to learn a generator that remains robust across different structural and interaction reliability conditions.

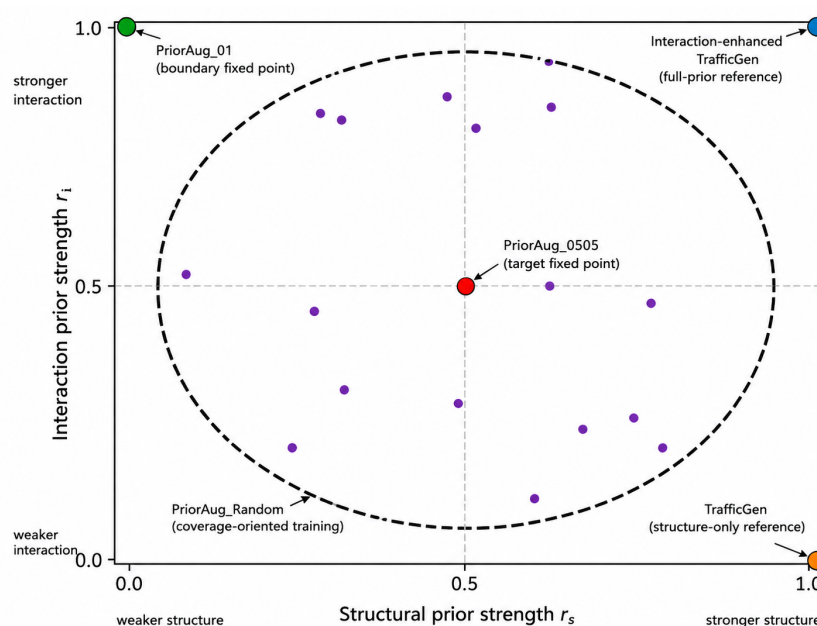


Figure 2. Two-dimensional reliability space of sensor-perceived priors. The horizontal and vertical axes denote structural reliability r_s and interaction reliability r_i , respectively. The expressions “structural prior strength” and “interaction prior strength” used in the schematic are visual descriptions

of structural and interaction reliability, respectively. The corner and central points illustrate representative deterministic operating conditions. The dashed ellipse schematically indicates the coverage region, and the purple circles denote sampled interior reliability conditions used by PriorAug_Rand. The label PriorAug_Random in the schematic corresponds to PriorAug_Rand in the text and tables. TrafficGen is shown conceptually near (1,0) as a structure-only reference; in implementation tables, its interaction reliability is marked as N/A because it contains no explicit interaction-control branch.

3.3. Reliability-Conditioned Prior Operators

To instantiate different reliability conditions during training, we introduce two reliability-conditioned prior operators. The structural operator adjusts the availability of perceived structural information, while the interaction operator adjusts the strength of inferred interaction information.

3.3.1. Structural Prior Operator

Given structural edges E_t^s and structural reliability r_s , we generate a binary mask for structural elements:

$$M_t^s(e) \sim \text{Bernoulli}(r_s), \quad e \in E_t^s. \quad (6)$$

The reliability-adjusted structural edges are then obtained as

$$\tilde{E}_t^s = M_t^s \odot E_t^s, \quad (7)$$

where $\odot$ denotes element-wise masking. When $r_s = 1$, all structural information is preserved. When $r_s = 0$, structural information is removed. Intermediate values correspond to partially perceived structural conditions.

Here, $M_t^s(e)$ is a Bernoulli variable indicating whether structural element e is retained. Each element is preserved with probability r_s and masked with probability $1 - r_s$; therefore, r_s controls the expected retained proportion of structural information. The mask is applied before the structural representation enters the generator and does not modify the backbone architecture or decoding procedure.

3.3.2. Interaction Prior Operator

Interaction information is inferred from dynamic traffic context and may degrade gradually rather than disappear abruptly. In this work, interaction reliability is controlled at the context-representation level after the interaction information has been encoded and aggregated by the interaction module. This design avoids directly deleting raw agent features and allows the generator to smoothly adjust the additional contribution introduced by explicit agent-agent interaction aggregation.

Let $\mathbf{c}_i^{(0)}$ denote the sample-specific base context obtained by bypassing the explicit agent-agent interaction aggregation module, and let $\mathbf{c}_i^{(\text{int})}$ denote the interaction-enhanced context produced when the interaction module is fully enabled. The interaction-neutral endpoint used in this work is therefore $\mathbf{c}_i^{(0)}$. It is not a zero vector, a dataset-wide average representation, or a learnable neutral embedding.

The interaction-induced context increment is defined as

$$\Delta \mathbf{c}_i^{(\text{int})} = \mathbf{c}_i^{(\text{int})} - \mathbf{c}_i^{(0)}. \quad (8)$$

Given the interaction reliability $r_i \in [0, 1]$, the reliability-adjusted interaction context is constructed as

$$\tilde{\mathbf{c}}_i(r_i) = \mathbf{c}_i^{(0)} + r_i \Delta \mathbf{c}_i^{(\text{int})}. \quad (9)$$

Equivalently, this can be written as

$$\tilde{\mathbf{c}}_i(r_i) = r_i \mathbf{c}_i^{(\text{int})} + (1 - r_i) \mathbf{c}_i^{(0)}. \quad (10)$$

When $r_i = 0$, the explicit interaction aggregation contribution is removed while the base context is retained. When $r_i = 1$, the complete interaction-enhanced context is used. Intermediate values provide a smooth transition over the interaction reliability dimension by scaling the interaction-induced context increment.

The base context $\mathbf{c}_i^{(0)}$ is used as the interaction-neutral endpoint because it preserves sample-specific non-interaction information and the decoder input dimensionality. In contrast, a zero vector would remove both interaction and non-interaction context, a dataset mean would lose sample specificity, and a learnable embedding could implicitly encode interaction patterns. Thus, r_i controls only the additional context introduced by explicit interaction aggregation.

3.3.3. Unified Reliability-Conditioned Prior

After applying the structural and interaction prior operators, the generator receives a reliability-conditioned prior representation. The structural operator produces the reliability-adjusted structural representation $\tilde{E}_t^s$, while the interaction operator produces the reliability-adjusted interaction context $\{\tilde{\mathbf{c}}_i(r_i)\}_{i=1}^N$. We denote the resulting reliability-conditioned prior representation as

$$\tilde{G}_t^{(r_s, r_i)} = \left(V_t, \tilde{E}_t^s, \{\tilde{\mathbf{c}}_i(r_i)\}_{i=1}^N \right). \quad (11)$$

Here, V_t denotes the traffic-agent nodes, $\tilde{E}_t^s$ denotes the structural information adjusted by the structural reliability r_s , and $\tilde{\mathbf{c}}_i(r_i)$ denotes the interaction context adjusted by the interaction reliability r_i . This notation is used to represent the complete reliability-conditioned input to the traffic generator. The interaction component should be understood as the context representation obtained after interaction encoding and aggregation, rather than as a raw edge-level interpolation. N denotes the number of traffic-agent nodes in the scene.

Each pair (r_s, r_i) therefore defines one realization of the sensor-perceived prior condition. By varying r_s and r_i , the training process exposes the generator to a broad range of structural and interaction reliability configurations. This enables the model to learn traffic scenario initialization under continuously varying perceived-prior conditions rather than under a single deterministic operating point.

3.4. Reliability-Aware Generator

Given the reliability-conditioned prior representation $\tilde{G}_t^{(r_s, r_i)}$, traffic scenario initialization aims to generate the initial states of traffic participants:

$$X_t = \{x_1, x_2, \dots, x_N\}, \quad (12)$$

where each x_k contains the generated position, heading, speed, size, and category of one traffic participant.

Instead of learning a deterministic mapping from a fixed scene graph

$$\hat{X}_t = F_\theta(G_t), \quad (13)$$

Here, θ denotes the parameters of the traffic generator.

PriorAug learns a reliability-aware conditional generator:

$$\hat{X}_t = F_\theta(\tilde{G}_t^{(r_s, r_i)}, r_s, r_i). \quad (14)$$

Equivalently, the model learns the conditional distribution

$$p_\theta(X_t | \tilde{G}_t^{(r_s, r_i)}, r_s, r_i), \quad (15)$$

which allows generation quality to adapt to different perceived-prior reliability conditions.

3.5. PriorAug Training

PriorAug modifies the perceived prior during training rather than changing the backbone generator architecture. For each training iteration, a reliability pair is sampled from a predefined reliability sampling distribution:

$$(r_s, r_i) \sim q, \quad (16)$$

where q denotes the training distribution over the two-dimensional reliability space. Two strategies are considered.

Fixed-point PriorAug. The reliability pair is fixed at a predefined operating point, such as $(0, 1)$ or $(0.5, 0.5)$. This strategy can be regarded as sampling from a degenerate distribution concentrated at the target reliability point:

$$q(r_s, r_i) = \delta(r_s - r_s^*)\delta(r_i - r_i^*). \quad (17)$$

Here, $\delta(\cdot)$ denotes the Dirac delta function, and (r_s^*, r_i^*) denotes the prescribed target reliability point.

Random PriorAug. The reliability pair is randomly sampled from the reliability space:

$$q(r_s, r_i) = U(0, 1) \times U(0, 1), \quad (18)$$

or equivalently,

$$r_s, r_i \sim U(0, 1). \quad (19)$$

Here, $U(0, 1)$ denotes the uniform distribution on $[0, 1]$. Random sampling provides broader coverage over the reliability space and reduces over-specialization to a single operating point.

The training objective follows the underlying traffic generation model:

$$\mathcal{L} = \mathcal{L}_{\text{gen}} + \lambda \mathcal{L}_{\text{reg}}, \quad (20)$$

where $\mathcal{L}_{\text{gen}}$ is the original generation loss and $\mathcal{L}_{\text{reg}}$ denotes the regularization term inherited from the backbone architecture, while λ is the weighting coefficient of the regularization term.

Since PriorAug modifies perceived prior conditions during offline training and does not introduce an additional learnable inference module or iterative decoding stage, the trained generator follows the standard inference procedure of the underlying backbone.

4. Experiments

This section evaluates the proposed PriorAug framework on the Waymo Open Motion Dataset. We assess generation quality from three perspectives: distribution fidelity, robustness across shifted-prior and traffic density conditions, and physical plausibility. We further analyze the reliability landscape over the proposed two-dimensional reliability space.

4.1. Experimental Settings

Dataset and task. The proposed framework is instantiated on the Waymo Open Motion Dataset (WOMD), a large-scale autonomous driving motion dataset collected using heterogeneous sensing systems, including cameras, LiDARs, localization modules, and high-definition mapping pipelines [21]. The dataset is publicly released through the official Waymo Open Dataset platform [22]. Rather than treating WOMD as a generic trajectory dataset, we interpret it as a structured scene representation reconstructed from multi-sensor perception, which naturally matches the sensor-perceived prior formulation introduced in Section 3.

Following the traffic scenario initialization setting, the model receives partial scene context and generates the initial states of the remaining traffic participants, including position, heading angle, speed, and physical size.

Reliability evaluation protocol. To validate robustness under continuously varying perception conditions, all evaluations are conducted within the proposed two-dimensional reliability space (r_s, r_i) , where r_s and r_i denote structural reliability and interaction reliability, respectively.

Structural reliability is controlled through the proposed structural prior operator by masking perceived structural information, while interaction reliability is adjusted through the interaction prior operator using reliability-aware interaction interpolation. Besides evaluating representative operating points, we further investigate coverage-oriented training over the entire reliability space to examine robustness under continuously changing sensor-perceived priors.

Compared methods. We compare the proposed framework with deterministic-prior baselines and additional training-strategy baselines. TrafficGen serves as the structural prior baseline operating at $(r_s, r_i) = (1, N/A)$, where the model uses structural information but does not contain an explicit interaction reliability branch. Interaction-enhanced TrafficGen serves as the full-prior baseline operating at $(1, 1)$, where both structural and interaction priors are fully preserved.

To further isolate the effect of reliability-aware training, we introduce three additional controlled baselines. TrafficGen-MapRand randomizes structural reliability during training on the original TrafficGen backbone. Since TrafficGen has no explicit interaction branch, r_i is not applicable to this baseline. IE-TrafficGen-MapRand randomizes only structural reliability while keeping interaction reliability fixed at $r_i = 1$. IE-TrafficGen-InterRand randomizes only interaction reliability while keeping structural reliability fixed at $r_s = 1$. These baselines are designed to distinguish joint reliability space training from structural-only randomization and interaction-only randomization.

For the proposed framework, PriorAug_01 represents the weak-structure fixed operating point $(0, 1)$, PriorAug_0505 corresponds to the target shifted operating point $(0.5, 0.5)$, and PriorAug_Rand denotes the coverage-oriented variant trained by jointly sampling structural and interaction reliability conditions from the proposed two-dimensional reliability space.

Recent traffic generation methods such as UniGen and SceneControl are also discussed in Section 2.1. However, their original formulations target joint initial-state-and-trajectory generation or controllable diffusion-based scene generation and use different output targets, conditioning interfaces, decoding procedures, and evaluation protocols. Their originally reported results are therefore not directly comparable with the initialization-only reliability-shift protocol adopted in this work. A direct numerical comparison would require architecture-specific reimplementations and retraining rather than applying the original methods unchanged.

To isolate the contribution of reliability-aware training under a unified architecture and evaluation protocol, we therefore introduce TrafficGen-MapRand, IE-TrafficGen-MapRand, and IE-TrafficGen-InterRand as controlled baselines. These variants share the same backbone, data split, decoding procedure, and inference budget, and differ only in the dimensions of reliability randomized during training.

Evaluation metrics. To comprehensively evaluate traffic scenario initialization, we consider three complementary aspects: distribution fidelity, robustness across scene complexity, and physical plausibility.

MMD-based distribution fidelity. Distribution fidelity is evaluated using Maximum Mean Discrepancy (MMD) between the generated and recorded traffic-attribute distributions. The metric is computed independently for each scene and each attribute $a \in \{\text{speed, position, heading, size}\}$, and is subsequently aggregated over all valid evaluation scenes.

The recorded samples are drawn from the held-out WOMD evaluation split and serve only as the empirical reference for offline benchmarking. No additional real-world data are collected for the generated scenes, and these reference samples are not required by the generator during deployment.

For scene s and attribute a , let

$$\mathcal{X}_s^a = \{x_{s,i}^a\}_{i=1}^{n_s}, \quad \mathcal{Y}_s^a = \{y_{s,j}^a\}_{j=1}^{n_s}, \quad \mathcal{Z}_s^a = \mathcal{X}_s^a \cup \mathcal{Y}_s^a, \quad (21)$$

where $\mathcal{X}_s^a$ and $\mathcal{Y}_s^a$ are the recorded and generated samples, respectively, and $|\mathcal{Z}_s^a| = N_s = 2n_s$.

The adaptive kernel scale is computed from the average off-diagonal squared Euclidean distance between the pooled samples:

$$b_{s,a} = \frac{1}{N_s(N_s - 1)} \sum_{p \neq q} \|z_{s,p}^a - z_{s,q}^a\|_2^2. \quad (22)$$

A single Gaussian kernel is then defined as

$$k_{s,a}(u, v) = \exp\left(-\frac{\|u - v\|_2^2}{b_{s,a}}\right). \quad (23)$$

Under the conventional parameterization $k(u, v) = \exp[-\|u - v\|_2^2 / (2\sigma_{s,a}^2)]$, this is equivalent to $\sigma_{s,a} = \sqrt{b_{s,a}/2}$.

The scene-level biased empirical squared MMD is calculated as

$$\begin{aligned} \widehat{\text{MMD}}_{s,a}^2 &= \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{i'=1}^{n_s} k_{s,a}(x_{s,i}^a, x_{s,i'}^a) \\ &\quad + \frac{1}{n_s^2} \sum_{j=1}^{n_s} \sum_{j'=1}^{n_s} k_{s,a}(y_{s,j}^a, y_{s,j'}^a) \\ &\quad - \frac{2}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} k_{s,a}(x_{s,i}^a, y_{s,j}^a). \end{aligned} \quad (24)$$

Let $\mathcal{S}_a$ denote the set of valid evaluation scenes for attribute a . The squared discrepancy is first averaged over the evaluation scenes, after which the square root is taken:

$$\text{MMD}_a = \sqrt{\max\left(\frac{1}{|\mathcal{S}_a|} \sum_{s \in \mathcal{S}_a} \widehat{\text{MMD}}_{s,a}^2, 0\right)}. \quad (25)$$

The same adaptive bandwidth estimation rule and kernel configuration are applied independently to every scene, attribute, and compared method. Lower MMD values indicate closer agreement between generated and recorded traffic-attribute distributions.

For the reliability-landscape analysis, the average MMD over the four attributes is further reported as

$$\text{AvgMMD} = \frac{1}{4} \sum_{a \in \{\text{speed, position, heading, size}\}} \text{MMD}_a. \quad (26)$$

Scene-complexity robustness. The fixed subset is divided into four buckets according to the number of traffic participants: 8–10, 11–15, 16–20, and 21+.

Overlap-based physical plausibility. Physical plausibility is evaluated using SceneOverlap and PairOverlap. For a generated scene s , let B_i^s denote the oriented bounding box of agent i , and define the pairwise overlap indicator as

$$o_{ij}^s = \mathbb{I}(\text{Area}(B_i^s \cap B_j^s) > \epsilon), \quad i < j, \quad (27)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, and $\epsilon = 10^{-6} \text{ m}^2$ is used to avoid numerical artifacts. Let $\mathcal{S}$ denote the set of evaluation scenes. SceneOverlap measures the percentage of scenes containing at least one overlapping vehicle pair:

$$\text{SceneOverlap} = \frac{100}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \mathbb{I}\left(\sum_{i < j} o_{ij}^s > 0\right). \quad (28)$$

Let A_s denote the number of agents in scene s . PairOverlap measures the average percentage of overlapping vehicle pairs within each scene:

$$\text{PairOverlap} = \frac{100}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \frac{1}{\binom{A_s}{2}} \sum_{i < j} o_{ij}^s. \quad (29)$$

Lower SceneOverlap and PairOverlap values indicate more physically feasible traffic initialization results.

Unless otherwise specified, full-set MMD is computed on all 10,000 evaluation scenarios, containing more than 60,000 valid samples after preprocessing. For computationally intensive shifted-prior, bucket-wise, and overlap analyses, 1000 scenarios were selected by scenario-level simple random sampling with a fixed seed and then processed using the same sample-construction pipeline, yielding more than 6000 valid samples. No stratification by sample count or density bucket was used. The same fixed subset was applied to every method, and subset results are used only to compare relative method behavior under identical conditions, while the main distribution fidelity results remain based on the complete set.

Implementation details. All reproduced baselines are trained using the same training-validation split and evaluated under an identical inference protocol. Repeat sampling, collision retry, and all decoding hyperparameters are kept unchanged across different methods so that performance differences primarily reflect modeling capability rather than inference budget.

The models are trained for 100 epochs using a batch size of 32 with the Adam optimizer. The initial learning rate is 3×10^{-4} , the weight decay is 4×10^{-4} , and ten data-loading workers are employed throughout training. All experiments are conducted on a single NVIDIA GeForce RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA) using PyTorch 1.10.0

with CUDA 11.3 and PyTorch Lightning 1.5.10. For the principal PriorAug models, three random seeds are adopted and the average performance is reported whenever applicable.

PriorAug is designed as an offline training-time augmentation strategy rather than an end-to-end real-time autonomous driving framework. During training, its reliability operators perform element-wise structural masking and interaction-context interpolation. These operations introduce no additional learnable inference modules or iterative decoding stages. After training, the generator follows the standard inference procedure of the underlying backbone. An end-to-end real-time evaluation covering upstream sensing, tracking, mapping, and closed-loop simulation is outside the scope of this study.

We first evaluate whether PriorAug improves traffic scenario initialization under standard and shifted-prior conditions. The analysis covers full-set distribution fidelity, density-stratified robustness, and overlap-based physical plausibility.

4.2. Distribution Fidelity

Table 1 reports the overall distribution fidelity evaluated by Maximum Mean Discrepancy (MMD) on four fundamental traffic attributes, including speed, position, heading, and size. Lower values indicate better agreement with the real traffic distribution.

Table 1 shows that Interaction-enhanced TrafficGen achieves the strongest performance at its native full-prior operating point, while PriorAug_Rand retains competitive full-set fidelity despite being trained over a broader reliability space. Differences in Size MMD are smaller than those in Speed, Position, and Heading, suggesting that object dimensions are less sensitive to prior-reliability variation. These matched-condition results establish that robustness-oriented training does not substantially sacrifice native distribution fidelity; shifted-condition robustness is examined next.

Table 1. Full-set MMD comparison under matched training–test reliability settings. All results are computed on the complete evaluation set. Lower values are better.

Model	Training	Test	Speed	Position	Heading	Size
TrafficGen	(1, N/A)	(1, N/A)	0.1602	0.1192	0.1189	0.0932
Interaction-enhanced TrafficGen	(1, 1)	(1, 1)	0.0634	0.0740	0.0697	0.0665
PriorAug_01	(0, 1)	(0, 1)	0.1143 ± 0.0002	0.1144 ± 0.0004	0.1141 ± 0.0007	0.0734 ± 0.0002
PriorAug_0505	(0.5, 0.5)	(0.5, 0.5)	0.0732	0.0766	0.0747	0.0791
PriorAug_Rand	(Rand, Rand)	(Rand, Rand)	0.0734	0.0753	0.0741	0.0747

Note: For fixed-point models, the test condition matches the corresponding training reliability point. “Rand” denotes reliability randomization during training and evaluation. For PriorAug_Rand, r_s and r_i are independently sampled from $\mathcal{U}(0, 1)$. For the original TrafficGen backbone, r_i is marked as N/A because this backbone does not include an explicit interaction reliability branch. Results under the common shifted condition $(r_s, r_i) = (0.5, 0.5)$ are reported in Table 2.

Table 2. Subset MMD comparison under the common shifted operating point $(r_s, r_i) = (0.5, 0.5)$. All methods are evaluated on the same fixed scenario-level subset. Lower values are better.

Model	Training	Test	Speed	Position	Heading	Size
TrafficGen	(1, N/A)	(0.5, N/A)	0.3615	0.3786	0.1933	0.7750
TrafficGen-MapRand	(Rand, N/A)	(0.5, N/A)	0.2756	0.2369	0.1736	0.1889
Interaction-enhanced TrafficGen	(1, 1)	(0.5, 0.5)	0.7264	0.3263	0.3586	0.1679
IE-TrafficGen-MapRand	(Rand, 1)	(0.5, 0.5)	0.4936	0.2376	0.2876	0.1267
IE-TrafficGen-InterRand	(1, Rand)	(0.5, 0.5)	0.0898	0.0962	0.0879	0.0802
PriorAug_01	(0, 1)	(0.5, 0.5)	0.1102	0.1117	0.1112	0.0739
PriorAug_0505	(0.5, 0.5)	(0.5, 0.5)	0.0706	0.0744	0.0745	0.0791
PriorAug_Rand	(Rand, Rand)	(0.5, 0.5)	0.0706	0.0753	0.0696	0.0771

Note: “Rand” denotes reliability randomization during training. For the original TrafficGen backbone, r_i is marked as N/A because this backbone does not include an explicit interaction reliability branch. All methods are evaluated on the same fixed scenario-level subset under the target shifted condition.

4.3. Robustness Under Prior Shift and Scene Complexity

Overall distribution fidelity may conceal different behaviors under traffic scenes with varying complexity. In practical autonomous driving systems, dense scenes usually introduce stronger occlusions, denser interactions, more difficult object association, and higher behavior estimation uncertainty. Therefore, we further evaluate robustness across different traffic density levels.

Before bucket-wise analysis, Table 2 reports the overall MMD results on the fixed one-tenth evaluation subset under the common shifted operating point $(r_s, r_i) = (0.5, 0.5)$. This table includes additional controlled baselines that isolate the effects of structural-only and interaction-only reliability randomization.

The results first show that deterministic-prior baselines degrade clearly once the perceived-prior condition shifts away from their native operating points. TrafficGen, which is trained with full structural information and no explicit interaction branch, suffers from large errors under partial structural reliability. TrafficGen-MapRand reduces the MMD values across all four attributes, indicating that structural-reliability randomization improves robustness on the structure-only backbone. However, its performance remains clearly worse than the interaction-aware PriorAug variants, suggesting that structural randomization alone is insufficient for robust initialization under the representative shifted condition.

For the interaction-enhanced backbone, IE-TrafficGen-MapRand improves over Interaction-enhanced TrafficGen, showing that randomizing structural reliability can partially alleviate structural prior shift. Nevertheless, it still produces relatively large errors because the interaction branch remains trained under full interaction reliability and is therefore mismatched when r_i decreases at test time. In contrast, IE-TrafficGen-InterRand substantially reduces the MMD values by exposing the model to varying interaction reliability during training. This indicates that interaction-pathway calibration is an important source of robustness at the target shifted point.

Most importantly, PriorAug_Rand achieves the lowest average MMD among the compared training strategies. Compared with IE-TrafficGen-InterRand, joint structural–interaction randomization further reduces the MMD values across all four attributes. This confirms that single-dimension randomization cannot fully replace joint coverage over the two-dimensional reliability space. The proposed PriorAug training strategy therefore improves robustness not merely by adding generic random perturbations, but by explicitly covering the coupled variation of structural and interaction reliability.

To further examine how this degradation varies with traffic complexity, we divide the fixed evaluation subset into four density buckets, namely 8–10, 11–15, 16–20, and 21+ traffic participants. The bucket-wise results are shown in Figure 3.

Across all density buckets, the PriorAug variants retain substantially lower MMD than the deterministic-prior baselines, particularly for Speed, Position, and Heading, demonstrating that reliability-aware training remains effective as scene complexity increases.

4.4. Physical Plausibility Analysis

While distribution fidelity measures statistical consistency with real traffic, physically feasible traffic scenario initialization requires that generated traffic participants also satisfy fundamental geometric constraints. In practical autonomous driving applications, unrealistic overlaps or collisions among initialized vehicles may lead to physically invalid simulation states, thereby reducing the reliability of downstream planning and simulation. Therefore, besides distribution matching, physical plausibility serves as another important criterion for evaluating traffic scenario initialization.

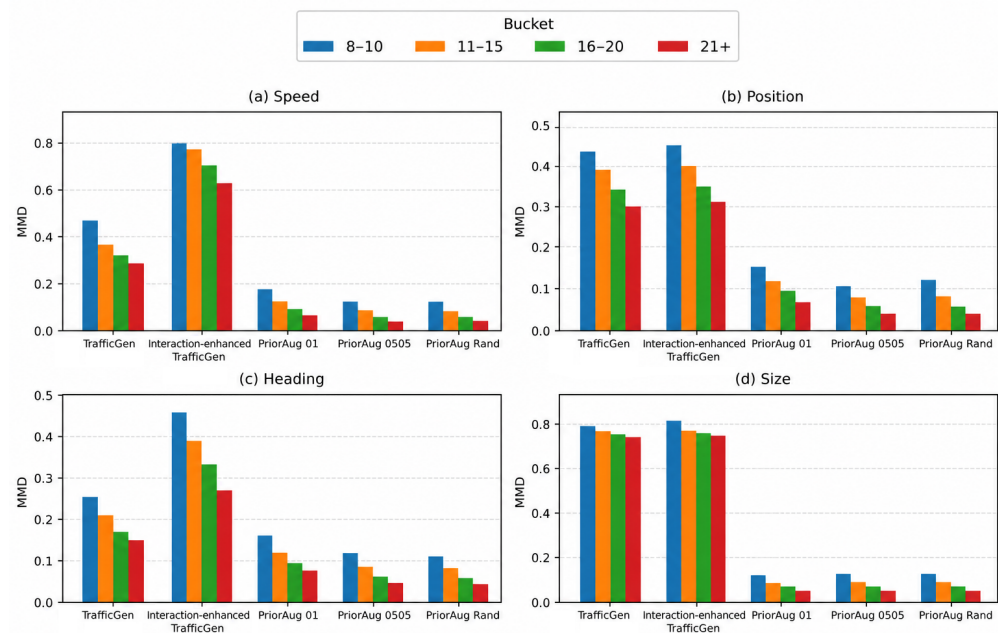


Figure 3. Bucket-wise MMD comparison under different traffic densities at the shifted operating point $(r_s, r_i) = (0.5, 0.5)$.

Table 3 reports the overlap evaluation using SceneOverlap and PairOverlap. Lower values indicate fewer geometric conflicts and therefore better physical feasibility.

Table 3. Physical plausibility evaluation under the common shifted test condition $(r_s, r_i) = (0.5, 0.5)$ on the same fixed scenario-level subset used in Table 2. Lower values indicate fewer geometric conflicts and better physical plausibility.

Model	Training Setting	Test Setting	SceneOverlap (%)		PairOverlap (%)	
			Pred.	GT	Pred.	GT
TrafficGen	(1, N/A)	(0.5, N/A)	9.616	1.821	0.095025	0.014166
Interaction-enhanced TrafficGen	(1, 1)	(0.5, 0.5)	30.053	1.821	0.348252	0.014166
PriorAug_01	(0, 1)	(0.5, 0.5)	1.008	1.821	0.008375	0.014166
PriorAug_0505	(0.5, 0.5)	(0.5, 0.5)	1.761	1.821	0.014993	0.014166
PriorAug_Rand	(Rand, Rand)	(0.5, 0.5)	1.550	1.821	0.012511	0.014166

Note: All methods use the same fixed subset and the test condition of Table 2. For TrafficGen, r_i is N/A. Generation uses repeat = 10 and max_retry = 5; the ground-truth values are computed on the same subset.

Table 3 shows that all PriorAug variants substantially reduce geometric conflicts relative to the deterministic-prior baselines. PriorAug_01 produces the lowest overlap rates, PriorAug_0505 is closest to the ground-truth reference, and PriorAug_Rand also remains near the ground-truth level.

Because no collision-related objective is used during training, these results suggest that exposure to varying structural and interaction reliability improves the learned spatial relationships among traffic participants. The overlap metrics therefore complement MMD by confirming that the robustness improvement also extends to physical plausibility.

4.5. Reliability Landscape Analysis

The shifted-point and bucket-wise results provide evidence at representative reliability conditions.

To further examine global robustness, we evaluate the trained models over the two-dimensional reliability space and visualize the resulting AvgMMD landscapes in Figure 4.

Interaction-enhanced TrafficGen achieves its best performance near its native full-prior configuration $(r_s, r_i) = (1, 1)$, but its performance rapidly deteriorates when either structural or interaction reliability decreases. This forms a steep and localized low-error region, indicating strong dependence on one deterministic-prior configuration. In contrast, PriorAug_Rand produces a much flatter landscape, with low-error regions extending over a broader range of reliability combinations. This demonstrates that coverage-oriented prior augmentation encourages smoother conditional distribution learning under varying perception conditions.

To quantify local robustness, we compute the Local Reliability Sensitivity (LRS) around representative operating points. Let $f(r_s, r_i)$ denote the average MMD over the evaluated attributes. The local sensitivities along the structural and interaction dimensions are defined as

$$\text{LRS}_s = \left| \frac{\partial f(r_s, r_i)}{\partial r_s} \right|, \quad \text{LRS}_i = \left| \frac{\partial f(r_s, r_i)}{\partial r_i} \right|. \quad (30)$$

Because the reliability grid interval is $\Delta = 0.5$, central differences are used for interior points and forward differences for boundary points.

For Interaction-enhanced TrafficGen, evaluated around its native operating point $(1, 1)$, the estimated sensitivities are $\text{LRS}_s = 0.0390$ and $\text{LRS}_i = 0.0338$. For PriorAug_Rand, evaluated around $(0.5, 0.5)$, the sensitivities decrease to $\text{LRS}_s = 0.0070$ and $\text{LRS}_i = 0.0075$. Thus, PriorAug_Rand reduces local sensitivity by approximately $5.6\times$ in the structural direction and $4.5\times$ in the interaction direction.

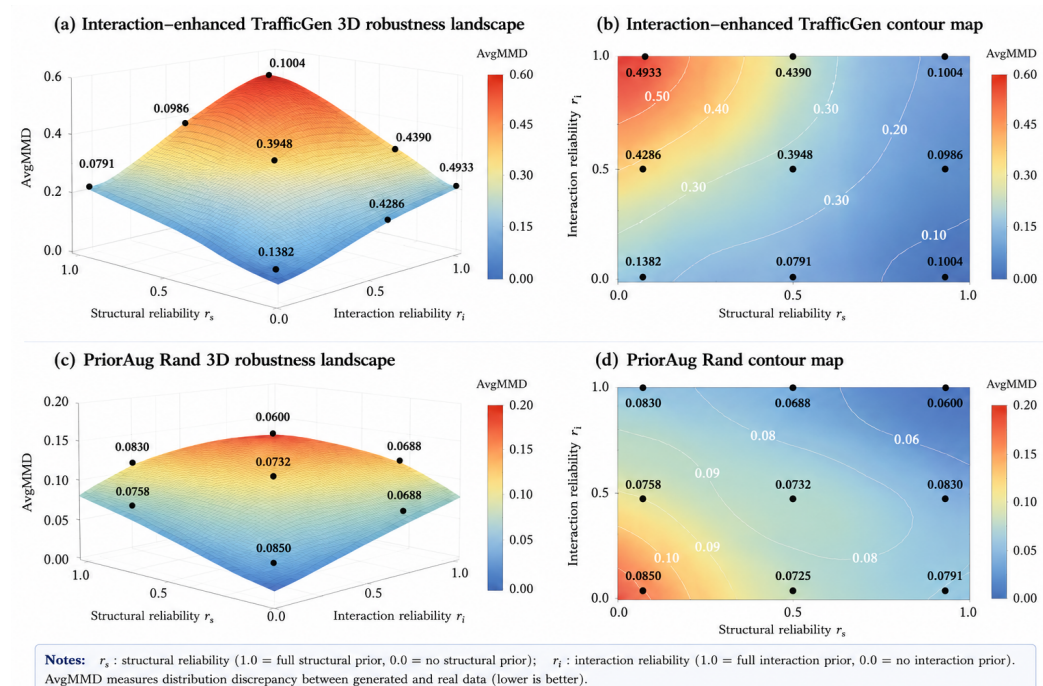


Figure 4. AvgMMD landscapes over structural and interaction reliability. AvgMMD denotes the average MMD across speed, position, heading, and size. The reliability axes are structural reliability r_s and interaction reliability r_i . Panels (a,b) show the three-dimensional surface and contour map, respectively, for Interaction-enhanced TrafficGen, whereas panels (c,d) show the corresponding results for PriorAug_Rand. Lower values indicate better point-wise fidelity, while flatter surfaces and smoother contours indicate stronger robustness to reliability shifts.

The flatter surface and substantially lower LRS values of PriorAug_Rand show that its improvement extends beyond isolated test points and reflects a smoother response throughout the reliability space.

4.6. Sensitivity of the Interaction Reliability Operator

To test whether the generator responds to the effective interaction contribution defined by the interaction reliability operator, we compare $r_i = 1.0$ and $r_i = 0.5$ at $r_s = 0.5$. As shown in Table 4, reducing r_i consistently lowers Speed MMD across all density buckets, while Position MMD remains similar. Thus, r_i affects generation behavior rather than acting as a purely notational interpolation coefficient. This experiment is a sensitivity analysis of interaction strength, not a full ablation of alternative neutral endpoints, which would require retraining separate zero-vector, dataset-mean, learnable-embedding, and base-context models.

Table 4. Sensitivity to interaction strength under the representative shifted structural prior condition $r_s = 0.5$. Lower values indicate better distribution fidelity.

Bucket	Speed MMD		Position MMD	
	$r_i = 1.0$	$r_i = 0.5$	$r_i = 1.0$	$r_i = 0.5$
8–10	0.1677 ± 0.0225	0.1483 ± 0.0175	0.1567 ± 0.0081	0.1606 ± 0.0092
11–15	0.1320 ± 0.0321	0.1070 ± 0.0146	0.1178 ± 0.0058	0.1203 ± 0.0108
16–20	0.1087 ± 0.0299	0.0797 ± 0.0110	0.0910 ± 0.0031	0.0935 ± 0.0071
21+	0.0818 ± 0.0305	0.0536 ± 0.0083	0.0706 ± 0.0049	0.0701 ± 0.0045

5. Discussion

The results show that deterministic-prior generators can perform well near their native operating conditions but become sensitive when structural or interaction reliability shifts. In contrast, PriorAug maintains competitive matched-condition fidelity and improves shifted, density-stratified, landscape, and overlap-based performance. The controlled baselines further indicate that joint structural–interaction randomization provides broader robustness than randomizing either dimension alone.

This work has two main limitations. First, reliability is controlled by predefined degradation operators rather than estimated directly from a real perception system. Second, the shifted-prior, bucket-wise, and overlap analyses use a fixed one-tenth subset because of computational cost. Future work will connect r_s and r_i to perception confidence, localization uncertainty, online-map quality, and tracking confidence, and extend the evaluation to larger-scale, long-horizon, and closed-loop simulation.

6. Conclusions

This paper introduced PriorAug, which models traffic scenario initialization over a two-dimensional space of structural and interaction reliability. Reliability-conditioned training preserves competitive full-set distribution fidelity while improving robustness under shifted-prior conditions and varying traffic densities, as well as the physical plausibility of generated initial states. The smoother reliability landscape and lower local sensitivity of PriorAug_Rand further show that coverage-oriented training reduces over-specialization to isolated deterministic operating points. These results support reliability-aware prior augmentation as a practical approach to robust traffic scenario initialization under uncertain sensor-perceived conditions.

Author Contributions: Conceptualization, K.Y. and S.Q.; methodology, K.Y.; software, K.Y.; validation, K.Y. and S.Q.; formal analysis, K.Y.; investigation, K.Y.; resources, K.Y. and S.Q.; data curation, K.Y.; writing—original draft preparation, K.Y.; writing—review and editing, K.Y. and S.Q.; visualization, K.Y.; supervision, S.Q.; project administration, S.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Waymo Open Motion Dataset analyzed in this study is publicly available from the Waymo Open Dataset platform subject to its data license and access requirements. No new public dataset was created in this study.

Acknowledgments: The authors acknowledge the Waymo Open Dataset for providing large-scale autonomous driving motion data for academic research.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AV	Autonomous vehicle
HD map	High-definition map
ITS	Intelligent transportation system
MMD	Maximum mean discrepancy
WOMD	Waymo Open Motion Dataset

References

1. Tan, S.; Wong, K.; Wang, S.; Manivasagam, S.; Ren, M.; Urtasun, R. SceneGen: Learning to Generate Realistic Traffic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 20–25 June 2021; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2021; pp. 892–901.
2. Feng, L.; Li, Q.; Peng, Z.; Tan, S.; Zhou, B. TrafficGen: Learning to Generate Diverse and Realistic Traffic Scenarios. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, London, UK, 29 May–2 June 2023.
3. Mahjourian, R.; Mu, R.; Likhoshesterov, V.; Mougine, P.; Huang, X.; Messias, J.; Whiteson, S. UniGen: Unified Modeling of Initial Agent States and Trajectories for Generating Autonomous Driving Scenarios. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, 13–17 May 2024.
4. Lu, J.; Wong, K.; Zhang, C.; Suo, S.; Urtasun, R. SceneControl: Diffusion for Controllable Traffic Scene Generation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, 13–17 May 2024; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2024; pp. 16908–16914.
5. Tan, S.; Ivanovic, B.; Weng, X.; Pavone, M.; Krahenbuhl, P. Language Conditioned Traffic Generation. In *Proceedings of the Conference on Robot Learning (CoRL)*; PMLR: Atlanta, GA, USA, 2023; Volume 229, pp. 2714–2752.
6. Elghazaly, G.; Frank, R.; Harvey, S.; Safko, S. High-Definition Maps: Comprehensive Survey, Challenges and Future Perspectives. *IEEE Open J. Intell. Transp. Syst.* **2023**, *4*, 527–550. [\[CrossRef\]](#)
7. Asrat, K.T.; Cho, H.-J. A Comprehensive Survey on High-Definition Map Generation and Maintenance. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 232. [\[CrossRef\]](#)
8. Zang, A.; Xu, R.; Zhou, F. Data Issues in High-Definition Maps: A Survey. *ACM Trans. Spat. Algorithms Syst.* **2024**, *10*, 1–37. [\[CrossRef\]](#)
9. Hui, N.; Jiang, Z.; Cai, Z.; Ying, S. Vision-HD: Road Change Detection and Registration Using Images and High-Definition Maps. *Int. J. Geogr. Inf. Sci.* **2024**, *38*, 454–477. [\[CrossRef\]](#)
10. Zhang, G.; Lin, J.; Wu, S.; Song, Y.; Luo, Z.; Xue, Y.; Lu, S.; Wang, Z. Online Map Vectorization for Autonomous Driving: A Rasterization Perspective. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2023; Volume 36.

11. Gao, J.; Ma, C.; Li, M.; Dai, S.; Zhu, Y.; Wang, Y.; Yang, B. VectorNet: Encoding HD Maps and Agent Dynamics from Vectorized Representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
12. Liang, M.; Yang, B.; Hu, R.; Wang, Y.; Urtasun, R. Learning Lane Graph Representations for Motion Forecasting. In *Computer Vision–ECCV 2020*; Springer: Cham, Switzerland, 2020.
13. Gupta, A.; Johnson, J.; Fei-Fei, L.; Savarese, S.; Alahi, A. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
14. Salzmann, T.; Ivanovic, B.; Chakravarty, P.; Pavone, M. Trajectron++: Dynamically-Feasible Trajectory Forecasting with Heterogeneous Data. In *Computer Vision–ECCV 2020*; Springer: Cham, Switzerland, 2020.
15. Yuan, Y.; Weng, X.; Ou, Y.; Kitani, K. AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
16. Tobin, J.; Fong, R.; Ray, A.; Schneider, J.; Zaremba, W.; Abbeel, P. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada, 24–28 September 2017; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2017; pp. 23–30.
17. Sadeghi, F.; Levine, S. CAD2RL: Real Single-Image Flight without a Single Real Image. In Proceedings of Robotics: Science and Systems (RSS), Cambridge, MA, USA, 12–16 July 2017.
18. Peng, X.B.; Andrychowicz, M.; Zaremba, W.; Abbeel, P. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 21–25 May 2018; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018; pp. 3803–3810.
19. Rajeswaran, A.; Ghotra, S.; Ravindran, B.; Levine, S. EPOpt: Learning Robust Neural Network Policies Using Model Ensembles. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
20. Pinto, L.; Davidson, J.; Sukthankar, R.; Gupta, A. Robust Adversarial Reinforcement Learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, Sydney, Australia, 6–11 August 2017; PMLR: Atlanta, GA, USA, 2017; Volume 70, pp. 2817–2826.
21. Ettinger, S.; Cheng, S.; Caine, B.; Liu, C.; Zhao, H.; Pradhan, S.; Chai, Y.; Sapp, B.; Qi, C.R.; Zhou, Y.; et al. Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 10–17 October 2021; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2021; pp. 9710–9719.
22. Waymo Open Dataset. Available online: <https://waymo.com/open/> (accessed on 18 June 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.