2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Semantic-Level Conflict Traffic Scenario Generation via Spatiotemporal Polygon Anchors

Yunwei Li[1*], Anran Wang[2*], Siyu Wu[1], Shengjie Fu[1], Shuo Feng[1], Hong Wang[1], and Jun Li[1]

***Abstract*— Autonomous Driving Systems (ADS) require rigorous and complex testing under diverse conditions to fulfill various demands and purposes of testing tasks, such as occlusion-triggered events, necessitating semantic-level control in scenario generation. Existing methods, reliant on low-level state controls, struggle to represent high-level semantic intents for task-oriented testing. We propose SPATSG, a novel framework for *event-driven, semantically aligned* Traffic Scenario Generation, leveraging Spatiotemporal Polygon Anchors (SPA) to bridge high-level test requirements and low-level diffusion guidance. SPAs encapsulate critical geometric and temporal patterns of traffic agents, derived from a set of targeted scenarios. During diffusion denoising, SPATSG integrates SPAs via an auxiliary loss to steer sampling toward desired semantics. A dynamic resampling strategy further intensifies guidance and prioritizes promising trajectory candidates progressively to balance exploration and refinement. We evaluate SPATSG on *SinD*, a Chinese intersection benchmark featuring complex interactions and diverse conflicts. Experiments on occlusion-triggered scenario generation show that SPATSG demonstrates superior semantic controllability, effectively reveals risk events across ADS, and maintains diversity and realism compared to baselines. This work offers a principled, interpretable approach for semantically controllable ADS testing and evaluation.**

I. Introduction

With the rapid advancement of autonomous driving systems (ADS), verifying their safety has become a pressing challenge. As emphasized by the Safety of the Intended Functionality (SOTIF) [1], safety assurance must go beyond system malfunction detection and address performance limitations under complex, realistic, and diverse conditions. Scenario-based ADS testing is widely recognized as a foundational strategy—serving as a core methodology in both emerging standards [2], [3] and industrial frameworks such as SAKURA [4] and PEGASUS [5].

Central to the scenario-based safety testing is the capability to generate scenarios that reflect two fundamental properties: 1) **Realism**—the ability to produce behaviorally plausible traffic situations consistent with naturalistic driving patterns; and 2) **Controllability**—the capacity to generate semantically meaningful scenarios that fulfill a wide range of high-level, event-specific testing demands. While realism ensures credibility of simulation outcomes, controllability determines whether the generated scenarios can meaningfully challenge the system under specific, safety-relevant conditions (e.g., occlusions, conflicts, or violations). Together, these two dimensions define the scope and effectiveness of scenario-based ADS testing.

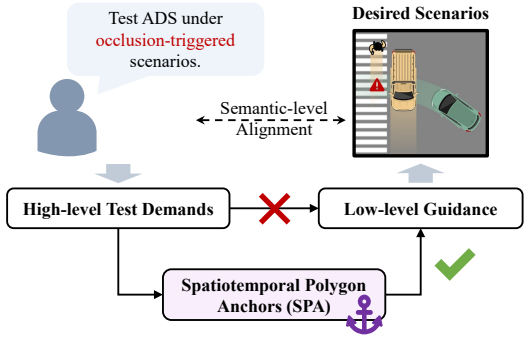

Fig. 1: In scenario-based ADS testing, high-level and semantic-level test demands are difficult to be expressed directly through low-level generative inputs. We introduces Spatiotemporal Polygon Anchor (SPA) as an interpretable intermediate representation, enabling semantic-level alignment and controllable scenario generation.

However, achieving both realism and controllability remains a highly non-trivial challenge in traffic scenario generation. Data-driven methods such as TrafficSim [6] and TrafficBots [7] leverage large-scale driving datasets to replicate realistic traffic agent interactions, yet they falter in generating safety-critical events due to the long-tailed distribution of such rare events in real-world data [8], which severely restricts their controllability for edge-case evaluation tasks. Conversely, optimization- and sampling-based approaches—including importance sampling [9], [10], adversarial generation [11], [12], and rule-based synthesis [13]—enhance the criticality of generated scenarios by identifying hazardous parameter combinations, thereby achieving broader coverage of critical traffic scenarios. Yet these methods often suffer from poor controllability and realism, tending to produce excessive and physically implausible risks and compromise the credibility of test results [14].

Recently, guidable deep generative models like diffusion models [15] have opened new possibilities for achieving both realism and controllability in traffic scenario generation. Prior works such as CTG/CTG++ [16] and DiffScene [17] demonstrate that diffusion models can be conditionally

This work was supported by the National Key R&D Program of China (2022YFB2503005), the National Natural Science Foundation of China (52072215, 52221005), the Beijing Natural Science Foundation (L243025), the State Key Laboratory of Intelligent Green Vehicle and Mobility, and Didi Chuxing.

*The authors contributed equally.
[1]Tsinghua University, Beijing, China
li-yw23@mails.tsinghua.edu.cn,
sy-wu20@tsinghua.org.cn, fu_sj@outlook.com,
{fshuo, hong_wang, lijun1958}@tsinghua.edu.cn
[2]East China University of Science and Technology, Shanghai, China
anran19931072017@163.com

guided to produce diverse, high-fidelity traffic behaviors under compositional constraints. However, existing approaches rely on low-level or direct forms of guidance (e.g., trajectory endpoints), which exhibit two major limitations: 1) **Limited semantic controllability.** Existing methods cannot support the *event-level, semantic-aligned scenario generation* required by SOTIF testing, where evaluation demands are often abstract (Fig. 1). Translating such high-level concepts into effective low-level signals remains a significant challenge. 2) **Lack of generalizable guidance formulation.** These approaches are typically tailored to specific scenario types or handcrafted objectives, which restricts their applicability across diverse semantic events or testing requirements.

To address these challenges, we propose **SPATSG**, a framework for generating semantically controlled, realistic traffic conflict scenarios using diffusion-based methods. Our approach introduces the **Spatiotemporal Polygon Anchor (SPA)**, a formal representation capturing the semantic and spatiotemporal structure of traffic conflicts, bridging high-level semantic constraints with low-level generative guidance. An automated pipeline extracts SPA properties from target datasets and converts them into differentiable guidance functions integrated into the reverse diffusion process. A **dynamic resampling strategy** balances early-stage exploration with late-stage convergence to high-quality trajectories. Experiments on occlusion-based conflict scenarios show that SPATSG generates diverse, physically plausible scenarios that effectively reveal functional weaknesses in ADS.

Key Contributions. This paper makes the following contributions to semantically controllable traffic scenario generation for autonomous driving testing:

- We propose the **Spatiotemporal Polygon Anchor (SPA)**, a structured and formal representation that captures the spatiotemporal semantics of traffic conflicts. SPA serves as a generic, transferable interface bridging high-level semantic constraints with low-level trajectory guidance in scenario generation.
- We develop **SPATSG**, a unified framework that: (a) automatically extracts SPA representations via weak supervision; (b) translates SPAs into differentiable guidance functions for diffusion-based generative models; and (c) incorporates a dynamic resampling strategy to balance early-stage diversity exploration and late-stage convergence to high-quality trajectories. SPATSG achieves a well-balanced trade-off between semantic controllability, scene realism and diversity for test scenario generation.
- We conduct extensive experiments on occlusion-based conflict scenarios, demonstrating the ability of SPATSG to generate semantically aligned, physically plausible, and risk-revealing traffic scenarios. The results confirm its effectiveness in exposing the performance limitations of different autonomous driving systems.

II. Related Work

Safety-critical Scenario Generation for ADS Testing. Adversarial sampling approaches like AdvSim [18] directly perturb trajectories in simulation to expose system failures, while optimization-based frameworks such as STRIVE [19] and Learning to Collide [20] explore parameter spaces to identify risky configurations with higher efficiency. More recent frameworks like CaDRE [21] integrate domain knowledge and black-box optimization to synthesize diverse safety-critical scenarios. However, many of these methods require expert-defined rules or iterative search loops, which may limit scalability and adaptability. Moreover, the generated scenarios may lack semantic-level structure and often do not generalize well across different types of safety-critical events.

Diffusion Models and guidance. Recent advances in guided diffusion models, have enabled controllable and realistic generation of images [15], [22] and videos [23] with test-time control in conditional guidance. Diffuser [24] extends the diffusion model to short-term trajectory prediction in robotics. Further, Scenario Diffusion [25] conditions latent diffusion on map features and scenario tokens to produce diverse agent placements and trajectories aligned with high-level directives. Similarly, Peng et al. [26] propose a guided latent diffusion model with graph-based representation learning to generate safety-critical scenarios that balance physical realism and adversarial exposure. Other approaches, such as CTG++ [27], incorporate Signal Temporal Logic (STL) specifications or rule-based objectives to steer diffusion sampling with formal semantic constraints. Despite such progress, these methods generally rely on low-level guidance signals—such as desired endpoints or trajectory goals—and are limited in expressing high-level event semantics.

III. Methodology

In this chapter, we present SPATSG: a semantic-guided, diffusion-based framework for traffic scenario generation (Fig. 2). We first define a single Spatio-Temporal Polygon Anchor (SPA), then show how multiple SPAs compose a scenario's semantic constraints, and finally detail how these constraints guide the reverse diffusion process.

A. Problem Statement

Formally, a traffic *scenario* is described as

$$\mathcal{S} = \big(\mathcal{TP},\, S_0,\, \tau^a\big) \tag{1}$$

where $\mathcal{TP}$ is the set of N traffic participants (agents), $S_0 = \{s_i(0)\}_{i=1}^{N}$ are their initial states, and $\tau^a = \{\tau_i^a\}_{i=1}^{N}$ with $\tau_i^a = \big(a_i(0), a_i(1), \ldots, a_i(T-1)\big)$ is the sequence of actions for each agent over T time steps.

Each agent i's state and action at time step t are defined as

$$\begin{aligned} s_i(t) &= \big[x_i(t),\, y_i(t),\, \theta_i(t)\big]^{\mathsf{T}} \in \mathbb{R}^3, \\ a_i(t) &= \big[\dot{v}_i(t),\, \dot{\theta}_i(t)\big]^{\mathsf{T}} \in \mathbb{R}^2 \end{aligned} \tag{2}$$

where $x_i(t)$,$y_i(t)$ are the 2D position coordinates, $\theta_i(t)$ is the heading angle. Given a known kinematic model f, the state trajectory τ_i^s of agent i is recovered via:

$$\tau_i^s = \big(s_i(1),\, s_i(2),\, \ldots,\, s_i(T)\big), \tag{3}$$

$$s_i(t+1) = f\big(s_i(t),\, a_i(t)\big) \tag{4}$$

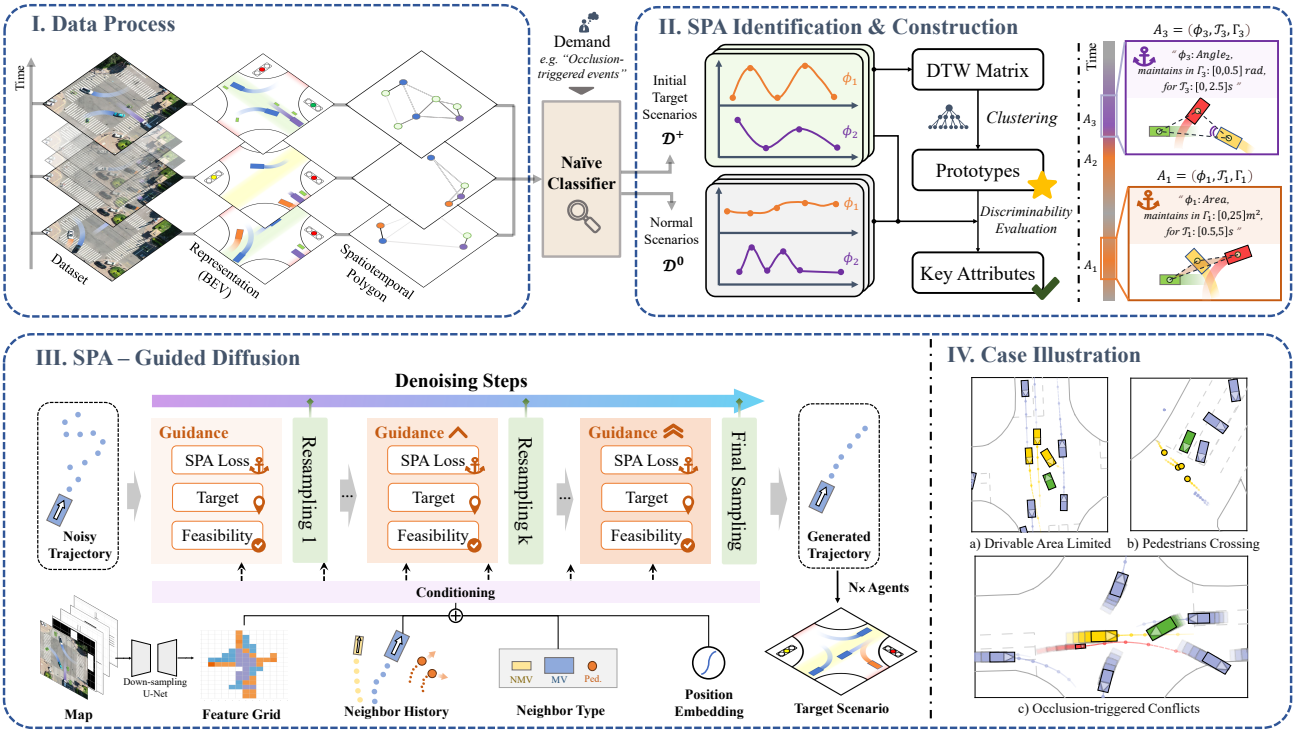

Fig. 2: **Overview of the proposed SPATSG framework.** (I) Represent scenarios as spatiotemporal polygons from dataset. (II) Given a high-level semantic demand, key attributes are extracted via clustering over DTW-based trends, and converted into discriminative SPAs. (III) During diffusion denoising, SPA guidance is applied via auxiliary loss and dynamic resampling. (IV) Example generated scenarios with specific semantic characteristics.

Therefore, a scenario $\mathcal{S}$ induces the joint trajectories

$$\tau^s = \{\tau_i^s\}_{i=1}^N, \quad \tau^a = \{\tau_i^a\}_{i=1}^N. \tag{5}$$

Controllable Scenario Generation. Given a static semantic map $\mathcal{M} \in \mathbb{R}^{H\times W\times C}$ (where H, W are spatial dimensions and C is the number of semantic channels) and initial states S_0, we aim to generate a scenario $\mathcal{S}$ whose trajectories satisfy specified testing objectives:

$$R(\tau^s) \geq R_{\text{th}} \quad \text{and} \quad E(\tau^s) = 1, \tag{6}$$

where

- $R(\tau^s)$ is a *risk metric* for the ego vehicle (e.g., minimum time-to-collision);
- R_{th} is a predefined risk threshold to ensure the generated scenario is safety-critical (e.g., TTC ≤ 2 seconds)
- $E(\tau^s) = 1$ indicates that a *semantic conflict event* (e.g., occlusion) is triggered.

Formally, the scenario generation task is modelled as:

$$\begin{aligned} \tau^{a*} &= \arg\max_{\tau^a} p_\theta(\tau^s, \tau^a \mid \mathcal{M}, S_0) \\ \text{s.t.} \quad & R(\tau^s) \geq R_{\text{th}}, \quad E(\tau^s) = 1, \\ & \tau^s = f(\tau^a, S_0) \end{aligned} \tag{7}$$

where τ^s is uniquely determined by the action sequence τ^a and initial states S_0 via the kinematic model f, and p_θ denotes the generative model that models the joint distribution of state and action trajectories conditioned on the semantic map and initial states.

B. Spatiotemporal Polygon Anchor (SPA)

To achieve semantically controllable traffic conflict scenarios, direct sampling from joint distribution is typically intractable due to the high-dimensionality and sparsity in valid events. Instead, we propose to leverage Spatio-Temporal Polygon Anchors (SPA) as an intermediate abstraction layer to encode key interaction semantics in a structured and expressive form.

The motivation stems from the observation that many traffic conflicts involve specific spatial configurations among agents. These can be naturally captured by geometric polygons formed by agent positions at critical moments. Further advantages include: 1) Spatiotemporal polygons provide an intuitive and efficient way for modeling multi-agent interactions, and 2) key geometric properties such as area, perimeter, and centroid are differentiable with respect to vertex positions in the spatiotemporal domain, enabling effective low-level guidance in diffusion-based generative processes.

A single SPA constraint is defined as:

$$A_j = (\phi_j, \mathcal{T}_j, \Gamma_j) \tag{8}$$

where $\phi_j \in \Phi$ is a geometric attribute function (e.g., polygon area, relative angle, etc.), $\mathcal{T}_j$ is the required maintenance time of the anchor, and $\Gamma_j = [\gamma_j^{\min}, \gamma_j^{\max}]$ is the admissible value range for the attribute function ϕ_j.

Let $\mathcal{P}_j(t)$ denote the polygon composed of agents involved in the SPA constraint A_j at time t. Then, the constraint is

satisfied if and only if there exists a continuous time interval $[t_s, t_e] \subseteq [0, T]$ with $t_e - t_s \geq \mathcal{T}_j$, such that $\phi_j\big(\mathcal{P}_j(t)\big) \in \Gamma_j$ for all $t \in [t_s, t_e]$.

Further, a set of SPAs can be temporally organized into a semantic constraint sequence:

$$\mathcal{C} = \big\{(A_j, \Delta t_j)\big\}_{j=1}^{K} \tag{9}$$

where $\Delta t_j \in [0, T]$ denotes the temporal gap between SPA A_j and A_{j+1}. Then, a generated scenario is considered semantically valid if all constraints in $\mathcal{C}$ are satisfied. This provides a flexible yet rigorous way to enforce high-level semantics in traffic scenario generation.

C. Anchor Attribute Identification

To enable automatic and interpretable scenario construction, we propose a pipeline based on temporal signal analysis to identify salient attributes that characterize specific scenario semantics and should be encoded as SPAs.

The core idea is to identify attributes that exhibit significant temporal pattern differences between target semantic scenarios and general scenarios, and to derive their corresponding estimators. To this end, an initial target scenario dataset $\mathcal{D}^+$ and a baseline scenario dataset $\mathcal{D}^0$ are required, where $\mathcal{D}^+$ can be obtained through rule-based or weakly supervised methods. For each scenario, we first extract time series of candidate attribute functions $\phi_k(t)$, such as polygon area, relative distances, angular spread, etc. These attribute functions are evaluated on agent groups of interest across the entire simulation duration. Then, the following steps are executed to evaluate whether ϕ_k qualifies as an anchor attribute and identify SPAs.

DTW-based Temporal Prototype Extraction. To measure structural similarity between attribute sequences, we compute the pairwise Dynamic Time Warping (DTW) [28] distances among all sequences in $\mathcal{D}^+$. Then we apply agglomerative hierarchical clustering to the DTW distance matrix $D^+ \in \mathbb{R}^{N^+ \times N^+}$ to discover clusters of target scenarios with similar temporal patterns, with a distance threshold τ_{hc}. For each cluster c, we extract a prototype sequence $\hat{\phi}_k^{(c)}(t)$, defined as the medoid (i.e., the most central sequence).

Discriminability Evaluation & SPA Instantiation. To assess the discriminative power of ϕ_k, the *distance-to-prototype* for i-th sample is calculated:

$$d_i = \min_j \mathrm{DTW}(\phi_k^{(i)}(t), \hat{\phi}_k^{(c)}(t)). \tag{10}$$

Let $D^+_{\text{proto}} = \{d_i\}_{\phi_k^{(i)} \in \mathcal{D}^+}$ and $D^0_{\text{proto}} = \{d_i\}_{\phi_k^{(i)} \in \mathcal{D}^0}$ denote the sets of distances to prototypes from high-risk and normal samples, respectively. Then we perform a *Kolmogorov–Smirnov (KS) test* between D^+_{proto} and D^0_{proto} to assess the separability of the two distributions. A statistically significant difference indicates that ϕ_k is a *discriminative anchor attribute*, and the prototype sequence $\hat{\phi}_k(t)$ can be used to instantiate SPAs by selecting:

- a time interval $\mathcal{T}_j$ during which the prototype stays within a bounded fluctuation range;
- a value range $\Gamma_j = [\gamma_j^{\min}, \gamma_j^{\max}]$ derived from the prototype's variation band.

D. Scenario Generation via SPA-Guided Diffusion

To realize the semantically controllable scenario generation, our pipeline consists of three key components: training a base trajectory generator, introducing SPA-driven guidance during sampling, and a dynamic resampling strategy to further refine the generated trajectories.

Base Trajectory Generation via Conditional Diffusion Model. Given scene context, we adopt a conditional diffusion model with a U-Net backbone to generate plausible future motion sequences. Let $\tau \in \mathbb{R}^{T \times d}$ denote a clean future trajectory, and $\mathbf{c}$ represent the conditioning inputs. The model learns to reverse a predefined noise schedule using the denoising diffusion probabilistic model (DDPM) framework [15]:

$$\mathcal{L}_{\text{ddpm}} = \mathbb{E}_{\tau,\epsilon,t}\left[\|\epsilon_\theta(\tau_t, t, \mathbf{c}) - \epsilon\|^2\right], \tag{11}$$

where $\tau_t = \sqrt{\bar{\alpha}_t}\tau + \sqrt{1-\bar{\alpha}_t}\epsilon$, and $\epsilon \sim \mathcal{N}(0, I)$ is standard Gaussian noise.

SPA-Guided Conditional Sampling. To guide trajectory generation with high-level semantics, we define a loss function that quantifies the degree to which a trajectory satisfies a semantic constraint sequence:

$$\mathcal{C} = \{(A_j, \Delta t_j)\}_{j=1}^{K}, \tag{12}$$

where each A_j denotes a Spatial-Pattern Anchor (SPA) requiring continuous satisfaction for at least $\mathcal{T}_j$, and Δt_j specifies the allowed temporal gap between A_j and A_{j+1}.

Given a trajectory $\tau^s = \{\mathbf{x}_t\}_{t=0}^{T_{\text{pre}}}$, we define the satisfaction signal:

$$s_j(t) = \mathbb{I}\left[\phi_j\big(\mathcal{P}_j(t)\big) \in \Gamma_j\right], \tag{13}$$

and identify the longest continuous interval where $s_j(t) = 1$, denoted by $\ell_j^{\max}$. The per-SPA loss is then defined as:

$$\mathcal{L}_{\text{spa}}^{(j)} = \max\left(0, \mathcal{T}_j - \ell_j^{\max}\right). \tag{14}$$

Let t_j^s denote the start time of the satisfying interval for A_j. To ensure temporal consistency across the SPA sequence, we define the temporal gap loss:

$$\mathcal{L}_{\text{gap}}^{(j)} = \max\left(0, \left|t_{j+1}^s - t_j^s\right| - \Delta t_j^{\max}\right). \tag{15}$$

The total SPA-driven semantic loss is:

$$\mathcal{L}_{\mathcal{C}} = \sum_{j=1}^{K} \mathcal{L}_{\text{spa}}^{(j)} + \sum_{j=1}^{K-1} \mathcal{L}_{\text{gap}}^{(j)}. \tag{16}$$

This formulation provides a soft yet interpretable objective to guide semantically-aligned trajectory generation. With commonly used target point guidance $\mathcal{L}_{tar}$ [24] and fasibility guidance $\mathcal{L}_{fea}$ [17] ensuring physical realism, the reverse diffusion transition can be modified as:

$$p_\theta(\tau^{k-1} \mid \tau^k) \approx \mathcal{N}(\tau^{k-1};\ \mu + \Sigma\nabla\mathcal{L}_(\tau),\ \Sigma), \tag{17}$$

where the weighted loss gradient $\nabla\mathcal{L}(\tau) = \nabla(w_1\mathcal{L}_{tar}(\tau) + w_2\mathcal{L}_{fea}(\tau) + w_3\mathcal{L}_{\mathcal{C}}(\tau))$ serves as the final guidance signal.

Dynamic Resampling Strategy. To balance exploration and refinement in the denoising process, we apply dynamic

resampling at selected M timesteps. Specifically, at each key step m:

1) *Trajectory Resampling.* Each trajectory τ_i is assigned a normalized sampling probability:

$$P_i \propto \exp\left(-\lambda_1 \cdot \mathcal{L}_{\text{fea}}(\tau_i) - \lambda_2 \cdot \mathcal{L}_{\mathcal{C}}(\tau^i)\right), \tag{18}$$

where $\mathcal{L}_{\text{fea}}$ evaluates physical realism. Then a new trajectory is resampled with injected noise for continued denoising:

$$\tau^r = \tau_i \sim \text{Multinomial}(P_i) + \mathcal{N}(0, \sigma^2). \tag{19}$$

2) *Guidance Amplification.* We progressively amplify the semantic guidance by increasing coefficient ξ_m, which scales the gradient of the loss during denoising:

$$p_\theta(\tau^{k-1} \mid \tau^k) \approx \mathcal{N}(\tau^{k-1};\ \mu + \Sigma \cdot \xi_m \cdot \nabla\mathcal{L}(\tau),\ \Sigma) \tag{20}$$

IV. Experiments

In this section, we present extensive open-loop and closed-loop experiments on the SinD dataset [29] to evaluate the effectiveness of **SPATSG** in semantic-level controllable scenario generation. Focusing on occlusion-triggered risk events, our results demonstrate that SPATSG significantly outperforms existing baselines in terms of *controllability*, *criticality*, and *realism*. Moreover, the generated scenarios successfully expose latent risks in various ADS. Additional ablation studies confirm the practical benefits brought by SPA-guided conditional sampling and the adaptive resampling strategy.

A. Experimental Setup

Dataset. All experiments are conducted on the SinD dataset [29], a large-scale BEV traffic dataset across multiple urban intersections in China, featuring diverse driving behaviors and complex multi-agent interactions.

ADS Under Test. We evaluate scenario effectiveness on two representative autonomous driving systems:

- **Risk-extended Intelligent Driver Model (RIDM)** [30]: A rule-based controller extending IDM [31] with a 2D risk-aware interaction module for intersection navigation.
- **ASAPRL** [32]: A hierarchical learning-based autonomous driving framework that combines imitation learning and reinforcement learning for robust policy learning in complex scenarios.

Baseline Methods. We compare SPATSG with three state-of-the-art scenario generation methods:

- **MPGA** [11]: A multi-population genetic algorithm for discovering safety-critical scenarios in black-box systems.
- **STRIVE** [19]: An adversarial framework combining learned behavioral priors and optimization-based attacks to generate challenging traffic scenarios.
- **TRACE** [33]: A diffusion-based trajectory generation model supporting low-level state control, aimed at diverse and realistic motion prediction.

Evaluation Metrics. To assess *controllability*, we introduce the **Event Satisfaction Score (ESS)**, which quantifies how well the generated scenario satisfies the intended semantic event. ESS is composed of two components: an event form score and a validity score, combined as $ESS = S_{\text{form}} \times S_{\text{valid}}$. In the case of occlusion-triggered conflicts, we define:

$$S_{\text{form}} = \min\left(1, \max\left(\frac{t_{\text{occ}}}{t_{\text{thr}}}\right)\right), \tag{21}$$

$$S_{\text{valid}} = \frac{1}{k_1 \cdot GL(t^*) + k_2 \cdot t^*_{\text{pov}}} \tag{22}$$

Here, S_{form} evaluates whether the longest occlusion duration t_{occ} exceeds a predefined threshold t_{thr}, while S_{valid} reflects the instant risk emergence at the moment t^* when the occlusion ends. It incorporates both *relative risk*, represented by the Gap Length (GL)—the temporal difference between two agents passing through the conflict zone [34]—and *absolute risk*, measured by the remaining time t^*_{pov} for the principle other vehicle (POV) to reach the conflict zone. We empirically set $k_1 = 4$ and $k_2 = 1$, as relative risk is typically considered more significant in safety assessments.

For *criticality*, we use two metrics: (1) the **minimum gap length (minGL)** observed throughout the scenario simulation, and (2) the **collision rate (CR)** across test runs.

For *feasibility*, we report: (1) the **illegality ratio over k steps (IR@k)**—which quantifies violations of traffic rules during simulation—and (2) **motion consistency (MC)**, computed as the KL divergence of space-vise velocity distributions between generated scenarios and real-world trajectories from *SinD*.

B. SPA Identification & Construction

Occlusion-triggered scenarios typically involve three interacting agents: Ego vehicle, Occluder vehicle, and Target vehicle—i.e., the vehicle that emerges from behind the occluder and eventually conflicts with the Ego. We construct a candidate attribute set Φ encompassing geometric and kinematic features of these agents, including bounding box area, heading angle, centroid position, centroid velocity, and pairwise distances.

To extract the initial positive scenario set $\mathcal{D}^+$, we employ a rule-based simulator that identifies Ego's line-of-sight occlusion conditions and potential critical conflicts, ultimately yielding 565 valid test scenarios. We use both p-value < 0.05 and Kolmogorov-Smirnov (KS) distance > 0.3 as discriminability criterion and identify three key attributes that distinguish occlusion-triggered scenarios: the area of the occlusion polygon (Area), the distance between the Ego and Target vehicle ($d_{\text{ego, tar}}$), and the incident angle of the Occluder vehicle at the occlusion location (α_{occ}) (Table I). Representative trends and top-2 prototypes are visualized in Fig. 3.

Through analysis of shared trends and peak patterns in the attribute distributions, we finally define four SPAs and formalize them into a semantic constraint sequence $\mathcal{C}$ for guiding scenario generation. The instantiation parameters are summarized in Table II.

TABLE I: Statistical Test Results for SPA Attributes.

Attribute	p-value	KS Distance	Significant
Centroid Pos.	8.36×10^{-2}	0.1939	False
Centroid Vel.	5.64×10^{-2}	0.2269	False
Area	8.70×10^{-13}	0.5713	True
$d_{\text{ego,occ}}$	1.92×10^{-2}	0.2820	False
$d_{\text{ego,tar}}$	4.53×10^{-6}	0.3925	True
$d_{\text{occ,tar}}$	7.12×10^{-2}	0.2145	False
α_{tar}	9.88×10^{-2}	0.1782	False
α_{ego}	3.46×10^{-2}	0.2651	False
α_{occ}	5.55×10^{-16}	0.6437	True

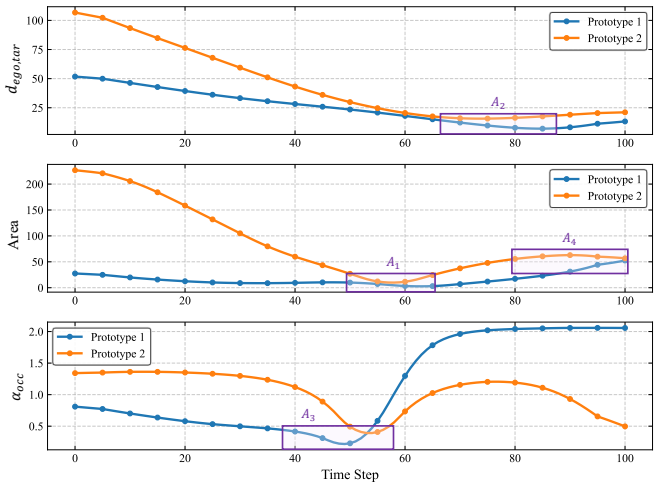

Fig. 3: Top-2 prototype curves of salient attributes and the corresponding prioritized SPAs.

C. Evaluation on Scenario Generation

We evaluate the effectiveness of SPATSG in generating occlusion-triggered risk scenarios under both open-loop setting (ego vehicle follows dataset-recorded trajectories) and closed-loop setting (ego vehicle is controlled by the autonomous driving algorithm under test). As shown in Table III and Table IV, SPATSG consistently achieves the highest Event Satisfaction Score (ESS), indicating superior controllability and semantic alignment with the intended testing objectives. Compared to baselines, SPATSG@Top1 yields over 30% improvement in ESS under open-loop settings and over 50% under closed-loop settings across both SUTs. Representative examples are illustrated in Fig. 4.

In terms of risk severity, SPATSG maintains strong performance with low minGL and elevated CR, effectively balancing controllability and scenario criticality. While MPGA attains the highest CR by directly optimizing for collisions, it does so at the expense of naturalness and motion diversity, as reflected in its extremely low minGL and motion consistency. In contrast, SPATSG retains natural property while still achieve the highest CR over STRIVE and TRACE.

Regarding naturalness and realism, SPATSG achieves the lowest IR throuth all tests, indicating fewer unnatural or overly aggressive behaviors during interaction. Additionally, as shown in Table V, SPATSG yields the lowest space-wise velocity KL divergence in motion distribution. These results

TABLE II: SPA Instantiation Parameters.

SPA	ϕ_i	$\mathcal{T}_i$ (s)	Γ_i	Δt_i (s)
A_1	Area	[0.5, 5]	[0, 25] m^2	[0.5, 5]
A_2	$d_{\text{ego,tar}}$	[0.5, 2]	[0, 20] m	[0, 1.5]
A_3	α_{occ}	[0.5, 2]	[0, 0.5] rad	[0, ∞]
A_4	Area	[1.0, ∞]	[30, ∞] m^2	-

TABLE III: Open-loop test results on occlusion-triggered scenarios.

Method	ESS↑	minGL(s)↓	CR(%)↑	IR@15/30(%)↓
Original	0.287	0.618	–	– / –
MPGA	0.000	**0.001**	**92.7**	30.5 / 32.1
STRIVE	0.000	0.463	22.5	48.8 / 49.2
TRACE	1.090	0.050	21.6	28.1 / 29.8
SPATSG @Top5	1.125	0.055	21.3	28.3 / 33.2
SPATSG @Top3	1.143	0.049	22.5	19.1 / 30.5
SPATSG @Top1	**1.439**	**0.045**	**32.3**	**18.4 / 25.0**

validate that SPATSG can generate behaviorally consistent manner that aligns with real-world motion patterns.

Fig. 5 further analyzes the diversity and rationality of the generated scenario types. It shows that target vehicles more frequently exhibit opposing-direction conflicts, as such situations are more likely to produce significant risks. Meanwhile, occluders are predominantly found in the same-direction cases, which tend to create longer-lasting effective occlusions.

D. Ablation Study

Table VI shows closed-looped ablation test results to evaluate the impact of SPA-guided conditional sampling and dynamic resampling strategy (DRS). Without any guidance, the model fails to generate meaningful controllable scenarios. Introducing only target guidance (TG) improves risk-inducing behavior but leads to unstable and less realistic outcomes. Adding SPA guidance enhances semantic alignment and criticality but sacrifices execution stability without DRS. When both modules are combined does the system achieve a strong balance between controllability, risk, and realism. This confirms that SPA provides meaningful scenario semantics, while DRS ensures stability and consistency during execution.

V. Conclusion

We propose **SPATSG**, a novel framework for generating semantically aligned and event-driven traffic scenarios, addressing the limitations of low-level control in existing ADS testing methods. By introducing Spatiotemporal Polygon Anchors (SPA) to represent key geometric and dynamic semantics of critical agents, SPATSG establishes an effective bridge between high-level testing demands and low-level generative control. Our conditional sampling scheme and dynamic resampling strategy jointly enable controllable,

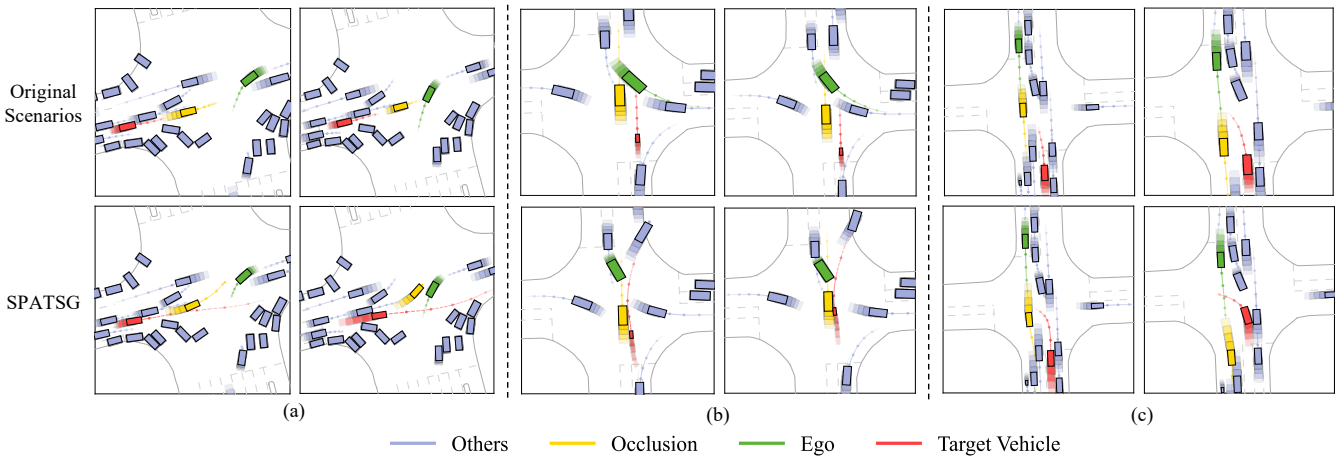

Fig. 4: Representative Generated Cases. SPATSG (bottom) constructs longer occlusion durations and enhances risk at the occlusion termination point compared to original scenarios (top), showcasing improved semantic controllability and alignment with intended testing goals.

TABLE IV: Closed-loop test results on occlusion-triggered scenario generation.

Method	ESS↑	minGL(s)↓	CR(%)↑	IR@15/30(%)↓
SUT1: RIDM				
Original	0.426	0.356	–	– / –
MPGA	0.042	**0.001**	**93.5**	31.5 / 32.7
STRIVE	0.000	0.285	25.2	54.4 / 56.1
TRACE	0.360	0.077	25.6	13.6 / 19.6
SPATSG	**0.546**	**0.068**	**28.4**	**8.8 / 17.2**
SUT2: ASAPRL				
Original	0.235	0.109	–	– / –
MPGA	0.415	**0.001**	**91.5**	31.2 / 32.9
STRIVE	0.025	0.751	27.1	33.2 / 34.9
TRACE	0.358	0.071	25.6	17.2 / 19.6
SPATSG	**0.895**	**0.083**	**32.0**	**11.6 / 16.9**

TABLE V: Motion Consistency (MC) between generated and original scenarios of different methods.

	MPGA	STRIVE	TRACE	SPATSG
KL Divergence ↓	4.424	4.663	0.668	**0.657**

diverse, and realistic scenario generation. Experiments on occlusion-triggered scenarios demonstrate that SPATSG outperforms existing methods in revealing critical events, while maintaining high-quality execution and coverage across diverse testing needs.

In future work, we aim to explore richer and more expressive scene representations, such as incorporating multi-agent intent and context-aware features into SPA. We also plan to investigate learning-based approaches for automatic SPA construction from data, potentially leveraging spatiotemporal attention or geometric encoders. Furthermore, integrating large language models (LLMs) to translate natural-language descriptions or testing requirements into semantic constraints

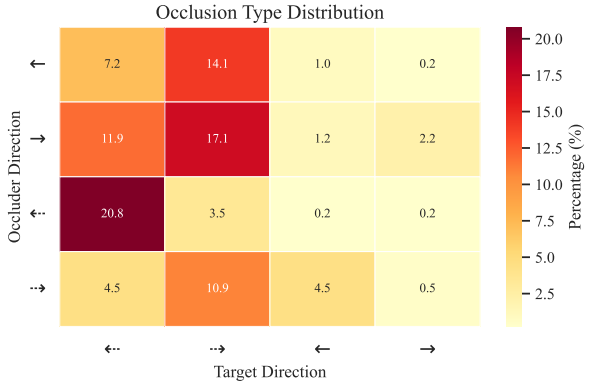

Fig. 5: Distribution of generated occlusion types by relative motion to ego at risk onset: opposite-left, opposite-right, same-left, same-right. The deeper color shows that opposite-direction (dotted arrow) target vehicles lead to higher risk, while same-direction (solid arrow) occluders yield longer occlusions.

TABLE VI: Ablation study results.

Guidance	DRS	ESS↑	minGL(s)↓	CR(%)↑	IR@15/30(%)↓
None	✓	0.180	0.141	15.1	5.2 / 19.3
TG	✓	0.155	0.045	22.1	4.7 / 18.6
TG+SPA	×	0.420	0.245	**34.4**	33.3 / 52.8
TG+SPA	✓	**0.594**	**0.006**	24.8	**2.1 / 17.8**

offers a promising direction for broader accessibility and adaptability in ADS testing.

REFERENCES

[1] ISO/DIS, "ISO 21448:2022, Road vehicles — Safety of the intended functionality," 2022.
[2] ISO, "ISO/DIS 34502, Road vehicles — Scenario-based safety evaluation framework for Automated Driving Systems," 2022.
[3] ISO, "ISO/AWI 34504, Road vehicles — Scenario attributes and categorization," 2024.

[4] A.-M. Jacobo, U. Nobuyuki, Y. Kunio, O. Koichiro, K. Eiichi, and T. Satoshi, "Development of a safety assurance process for autonomous vehicles in Japan," in *Proceedings of ESV Conference*, 2019.

[5] M. Scholtes, L. Westhofen, L. R. Turner, K. Lotto, M. Schuldes, H. Weber, N. Wagener, C. Neurohr, M. H. Bollmann, F. Kortke, J. Hiller, M. Hoss, J. Bock, and L. Eckstein, "6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment," *IEEE Access*, vol. 9, pp. 59131–59147, 2021.

[6] S. Suo, S. Regalado, S. Casas, and R. Urtasun, "TrafficSim: Learning to Simulate Realistic Multi-Agent Behaviors," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Nashville, TN, USA), pp. 10395–10404, IEEE, June 2021.

[7] Z. Zhang, A. Liniger, D. Dai, F. Yu, and L. Van Gool, "TrafficBots: Towards World Models for Autonomous Driving Simulation and Motion Prediction," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, (London, United Kingdom), pp. 1522–1529, IEEE, May 2023.

[8] H. X. Liu and S. Feng, "Curse of rarity for autonomous vehicles," *nature communications*, vol. 15, no. 1, p. 4808, 2024.

[9] D. Zhao, X. Huang, H. Peng, H. Lam, and D. J. LeBlanc, "Accelerated Evaluation of Automated Vehicles in Car-Following Maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, pp. 733–744, Mar. 2018.

[10] S. Feng, X. Yan, H. Sun, Y. Feng, and H. X. Liu, "Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment," *Nature Communications*, vol. 12, p. 748, Feb. 2021.

[11] Y. Li, S. Wu, and H. Wang, "Adaptive Mining of Failure Scenarios for Autonomous Driving Systems Based on Multi-population Genetic Algorithm," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, (Jeju Island, Korea, Republic of), pp. 2458–2464, IEEE, June 2024.

[12] J. Wu, C. Lu, A. Arrieta, and S. Ali, "Multi-Objective Reinforcement Learning for Critical Scenario Generation of Autonomous Vehicles," Feb. 2025.

[13] W. Ding, M. Xu, and D. Zhao, "CMTS: Conditional Multiple Trajectory Synthesizer for Generating Safety-critical Driving Scenarios," Oct. 2019.

[14] W. Ding, C. Xu, M. Arief, H. Lin, B. Li, and D. Zhao, "A Survey on Safety-Critical Driving Scenario Generation – A Methodological Perspective," Feb. 2022.

[15] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[16] Z. Zhong, D. Rempe, D. Xu, Y. Chen, S. Veer, T. Che, B. Ray, and M. Pavone, "Guided Conditional Diffusion for Controllable Traffic Simulation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, (London, United Kingdom), pp. 3560–3566, IEEE, May 2023.

[17] C. Xu, A. Petiushko, D. Zhao, and B. Li, "Diffscene: Diffusion-based safety-critical scenario generation for autonomous vehicles," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 8797–8805, 2025.

[18] J. Wang, A. Pun, J. Tu, S. Manivasagam, A. Sadat, S. Casas, M. Ren, and R. Urtasun, "AdvSim: Generating Safety-Critical Scenarios for Self-Driving Vehicles," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Nashville, TN, USA), pp. 9904–9913, IEEE, June 2021.

[19] D. Rempe, J. Philion, L. J. Guibas, S. Fidler, and O. Litany, "Generating Useful Accident-Prone Driving Scenarios via a Learned Traffic Prior," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (New Orleans, LA, USA), pp. 17284–17294, IEEE, June 2022.

[20] W. Ding, B. Chen, M. Xu, and D. Zhao, "Learning to Collide: An Adaptive Safety-Critical Scenarios Generating Method," July 2020.

[21] P. Huang, W. Ding, B. Stoler, J. Francis, B. Chen, and D. Zhao, "CaDRE: Controllable and Diverse Generation of Safety-Critical Driving Scenarios using Real-World Trajectories," Dec. 2024.

[22] L. Zhang, A. Rao, and M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models," Nov. 2023.

[23] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and D. J. Fleet, "Video diffusion models," *Advances in neural information processing systems*, vol. 35, pp. 8633–8646, 2022.

[24] M. Janner, Y. Du, J. B. Tenenbaum, and S. Levine, "Planning with Diffusion for Flexible Behavior Synthesis," Dec. 2022.

[25] E. Pronovost, M. R. Ganesina, N. Hendy, Z. Wang, A. Morales, K. Wang, and N. Roy, "Scenario diffusion: Controllable driving scenario generation with diffusion," in *Advances in Neural Information Processing Systems* (A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, eds.), vol. 36, pp. 68873–68894, Curran Associates, Inc., 2023.

[26] M. Peng *et al.*, "Safety-critical traffic simulation with guided latent diffusion model," *arXiv preprint arXiv:2505.00515*, May 2025.

[27] Z. Zhong, D. Rempe, Y. Chen, B. Ivanovic, Y. Cao, D. Xu, M. Pavone, and B. Ray, "Language-guided traffic simulation via scene-level diffusion," 2023.

[28] S. Salvador and P. Chan, "Fastdtw: Toward accurate dynamic time warping in linear time and space," in *KDD Workshop on Mining Temporal and Sequential Data*, pp. 70–80, 2004.

[29] Y. Xu, W. Shao, J. Li, K. Yang, W. Wang, H. Huang, C. Lv, and H. Wang, "SIND: A Drone Dataset at Signalized Intersection in China," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2471–2478, Oct. 2022.

[30] Y. Li, S. Wu, A. Wang, L. Yang, H. Wang, J. Li, and C. Huang, "High-dimensional functional boundaries search for deviation-robust testing of autonomous driving system," *Accident Analysis & Prevention*, vol. 221, p. 108156, Oct. 2025.

[31] M. Treiber and A. Kesting, *Traffic Flow Dynamics: Data, Models and Simulation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[32] L. Wang, J. Liu, H. Shao, W. Wang, R. Chen, Y. Liu, and S. Waslander, "Efficient Reinforcement Learning for Autonomous Driving with Parameterized Skills and Priors," in *Robotics: Science and Systems XIX*, Robotics: Science and Systems Foundation, July 2023.

[33] D. Rempe, Z. Luo, X. B. Peng, Y. Yuan, K. Kitani, K. Kreis, S. Fidler, and O. Litany, "Trace and Pace: Controllable Pedestrian Animation via Guided Trajectory Diffusion," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Vancouver, BC, Canada), pp. 13756–13766, IEEE, June 2023.

[34] United States Department of Transportation, "Baseline analysis of driver performance at intersections for the left-turn assist and intersection movement assist applications," technical report, National Transportation Library, Bureau of Transportation Statistics, 2020. Accessed: 2025-07-28.